American Society of Human Genetics 63rd Annual Meeting
October 22–26, 2013 Boston

POSTER ABSTRACTS

The program and abstract/poster board number next to each listing is followed by a **W** (Wednesday), or **T** (Thursday), or **F** (Friday) to indicate the day on which authors must be present at their poster boards. Posters will remain on the boards for all three days (Wednesday through Friday).

Session Topic/Title	*Abstract/Poster Board Numbers Start #*	*End #*
Epigenetics	412	511
Genome Structure, Variation and Function	512	694
Pharmacogenetics	695	743
Complex Traits and Polygenic Disorders	744	1180
Psychiatric Genetics, Neurogenetics and Neurodegeneration	1181	1430
Bioinformatics and Genomic Technology	1431	1693
Statistical Genetics and Genetic Epidemiology	1694	1942
Evolutionary and Population Genetics	1943	2105
Cardiovascular Genetics	2106	2260
Therapy for Genetic Disorders	2261	2312
Metabolic Disorders	2313	2406
Genetics/Genomics Education	2407	2424
Ethical, Legal, Social and Policy Issues in Genetics	2425	2476
Genetic Counseling	2477	2500
Health Services Research	2501	2521
Clinical Genetic Testing	2522	2639
Clinical Genetics and Dysmorphology	2640	2816
Prenatal, Perinatal and Reproductive Genetics	2817	2884
Molecular Basis of Mendelian Disorders	2885	3160
Development	3161	3187
Cytogenetics	3188	3257
Cancer Genetics	3258	3507

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

POSTER AUTHOR SCHEDULE

The program and abstract/poster board number next to each listing is followed by a **W** (Wednesday), **T** (Thursday), or **F** (Friday) to indicate the day on which authors must be present at their poster boards. Refer to the schedule below for presentation times and for the poster mounting/removal schedule. **Posters should remain on the boards for all three days.**

Wednesday

10:00 am–10:30 am	Authors place posters on boards
10:00 am–6:00 pm	Posters open for viewing
10:30 am–12:30 pm	**Poster Session I (W)**
	10:30 am–11:30 am (*odd poster board numbers; author must be present*)
	11:30 am–12:30 pm (*even poster board numbers; author must be present*)

Thursday

10:00 am–4:30 pm	Posters open for viewing
10:30 am–12:30 pm	**Poster Session II (T)**
	10:30 am–11:30 am (*odd poster board numbers; author must be present*)
	11:30 am–12:30 pm (*even poster board numbers; author must be present*)

Friday

10:00 am–2:00 pm	Posters open for viewing
10:30 am–12:30 pm	**Poster Session III (F)**
	10:30 am–11:30 am (*odd poster board numbers; author must be present*)
	11:30 am–12:30 pm (*even poster board numbers; author must be present*)
2:00 pm–2:30 pm	**Authors must remove posters**
2:30 pm	Exhibit Hall and Posters closed

Copyright © 2013 The American Society of Human Genetics. All rights reserved.

412T

Quality Control and Data Normalisation in large Illumina Infinium HumanMethylation450 Datasets. *A. Drong[1], B. Lehne[2], M. Loh[2,3], C. Blancher[1], M.-R. Jarvelin[2,3,4,5], C.M. Lindgren[1], P. Elliott[2,4], M.I. McCarthy[1,6], J.S. Kooner[7], J.C. Chambers[2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 3) 3Institute of Health Sciences, University of Oulu, Oulu, Finland; 4) 4MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom; 5) National Institute of Health and Welfare, University of Oulu, Oulu, Finland; 6) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 7) National Heart and Lung Institute, Imperial College London, London, United Kingdom.

Objectives: We aimed to develop methodology for epigenome-wide association studies (EWAS) of DNA methylation. This includes assessments and optimisation of current approaches for data preprocessing for use on large-scale datasets (N>2,000). Methods: The EpiMigrant DNA methylation data set was generated using the Illumina HM450 platform and includes baseline DNA samples of non-diabetic South Asians (N=2,687), approximately half of whom went on to develop diabetes. Of these samples, 36 were measured in duplicate to constitute the replication subset. We filtered methylation scores by a Bonferroni-corrected detection p-value, cutoff, instead of the default value of p<0.05. As a metric of performance, we used the count of false methylation calls on the Y chromosome in females. To test whether results were valid for autosomal markers, we compared the autosomal per-marker call rate to the rate of outliers (beta > 1.5 inter-quartile range). We then assessed two different methods for normalisation of methylation scores across arrays using the probe- wise correlation between duplicates in the technical replication dataset: QN1, quantile normalisation on Beta values, separated by probe type and colour channel into three categories and QN2, quantile normalisation on intensity values, further separated by methylated or unmethylated targets into six categories. Results:. We find that using the Bonferroni-corrected threshold (p<10-16 in our dataset) reduces the number of false methylation calls on the Y chromosome in females. Remaining intensity signals can be explained by probes cross-hybridizing to autosomal regions. For autosomal markers, using a lower detection p-value cutoff decreases the outlier rate, while only filtering 0.4% of markers. We find that before normalisation, only 9.9% of markers show high correlation (r>=0.8) between the two duplicates. QN1 increases this proportion to 10.6%, while QN 2 leads to 21.8% of probes reproducing (r>=0.8). The remaining high percentage of weakly performing probes can be explained by experimental variation exceeding inter-individual variation in methylation. These invariant markers are also unlikely to give rise to an association signal. Conclusion: We conclude that any analysis of large-scale EWAS data should be based on probes filtered for a Bonferroni-corrected detection p-value. To further increase data quality, we recommend separate quantile normalisation on intensities of the six different probe categories.

413F

Integrating genotype, methylome, chromatin states and disease state in a cohort of 750 individuals. *M.L. Eaton[1], G. Srivastava[2], A. Kundaje[1], L.B. Chibnik[2], B.T. Keenan[2], J. Ernst[3], D. Bennett[4], B. Bernstein[5], P. DeJager[2], M. Kellis[1].* 1) CSAIL, MIT, Cambridge, MA; 2) Neurosciences Institute, Brigham & Women's Hospital, Boston, MA; 3) UCLA, Los Angeles, CA; 4) Rush University Medical Center, Chicago, IL; 5) Massachusetts General Hospital, Boston, MA.

While the methodological framework for relating genotype to disease has been well established through the development of genome-wide association studies, the incorporation of intermediate molecular phenotypes such as DNA methylation and histone modifications into disease studies remains largely unexplored. Using a cohort of 750 individuals, half of whom were diagnosed with Alzheimer's Disease (AD), we sought to address the relationship between genotype, methylation status, chromatin state, and disease. Our first challenge was to nullify DNA methylation variability not attributable to genotype or AD, as several covariates confound this relationship including gender, batch and cell type tissue heterogeneity. We used methods including ICA and non-negative least squares to computationally remove this variation. We were then able to discover ~55K CpG probes with detectable genotype associations in cis (~12% of the array). These CpGs are preferentially located in enhancers and underrepresented in promoters based on a chromatin map generated by chromHMM on 7 histone marks of a matched brain region, suggesting distinct regulatory architectures. In associating DNA methylation changes with Alzheimer's disease, we found a pervasive signal of small effect size, robustly enriched for distal enhancer regions. Moreover, by linking the distal enhancers to the genes they regulate by correlating chromatin with gene expression, we found robust concentrations of associated CpGs in certain neuron- and signaling-specific gene pathways. Interestingly, disease-associated CpGs were enriched for meQTLs of weak effect but depleted for meQTLs of strong effect even after permutation, arguing further for a model of many loci of weak effect contributing overall to a disease onset burden. Overall, our results suggest that global regulatory changes are associated with complex disease, and suggest a general methodology for integration of genetic and epigenetic variation in the context of human disease.

414T

Epigenome-wide DNA methylation study reveals hypermethylated collagen genes and suggests a role for TGFβ in osteoarthritis. *M.A. Jeffries[1,2], J.A. James[2], A.H. Sawalha[3].* 1) Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Arthritis & Immunology, OMRF, Oklahoma City, OK; 3) Rheumatology, University of Michigan, Ann Arbor, MI.

Background: Osteoarthritis (OA) is the leading cause of chronic disability in the U.S., affecting 40% of individuals over the age of 70 and costing $128 billion annually in the US alone. Late-stage OA chondrocytes exhibit a host of gene transcription changes leading to upregulation of enzymes that contribute to cartilage breakdown. Herein, we characterize epigenome-wide DNA methylation changes in osteoarthritic compared to healthy cartilage from the same joints. Methods: Articular cartilage tissue from 12 OA femoral heads was dissected from affected and normal areas, frozen in liquid nitrogen, and DNA extracted. Following sodium bisulfite-treatment, DNA methylation was quantified at >485,000 CpG sites across the genome using Illumina HumanMethylation450 arrays. CpG sites with an absolute methylation difference between OA and normal cartilage ($\Delta\beta$) $\geq$ 15%, and P <0.01 after correction for multiple testing using a false discovery rate of 5%, were considered significant and used for analysis. Results: We found 442 differentially methylated CpG sites: 260 hypo- and 182 hypermethylated. Overrepresented gene sets included 'Connective tissue disorder' (n=53, p=4.73E-6 to 7.36E-3), 'Developmental disorder' (n=44, p=4.73E−6 to 7.36E−3), and 'Skeletal & muscular disorder' (n=59, p=4.73E−6 to 7.36E−3). Interesting DM in OA include hypermethylation of COL11A2, which functions to maintain spacing and diameter of type II collagen and is mutated in patients with Stickler syndrome and OSMED. Additionally, we found hypermethylation of COL6A2, hypomethylation of the fibrillar collagen gene COL1A1, and hypermethylation of COL18A1. Hypermethylation was also noted within the WNT pathway coreceptor LRP5, associated with OA in mice and is a susceptibility gene for OA in humans. Hypomethylation was found in the transcription factor RUNX1. Upstream regulator analysis identified significant association of TGFβ1 with 44 differentially methylated genes. Finally, canonical pathway analysis identified enrichment of several pathways, most prominently the ERK signaling pathway among differentially methylated genes (p=1.51E−4). Conclusions: We detected methylation changes in multiple collagen genes in OA. Our data suggest an epigenetic basis for defective collagen production in OA. Furthermore, we found evidence for epigenetic dysregulation of WNT signaling, and enrichment of genes associated with TGFβ- and ERK-pathway signaling, both are enticing targets for OA.

415F

Genome-wide profiling of DNA from blood reveals regions differently methylated in osteoarthritis patients. *G. Zhai, E. Aref Eshghi, H. Zhang, G. Martin, A. Furey, R. Green, G. Sun, P. Rahman.* Discipline of Genetics, Fac Med, Memorial Univ Newfoundland, St John's, NL, Canada.

Background: Osteoarthritis (OA) represents the most common form of arthritis and has substantial clinical and economic impact. Although recent studies have identified numerous genetic loci associated with OA, many of the risk factors for OA remain unexplained. We hypothesize that epigenetic mechanisms are responsible for the missing heritability in OA and tested our hypothesis using an epigenome-wide association approach. Methods: A total of 17 patients having knee- or hip-joint replacement due to primary OA and 24 healthy controls participated in this study. DNA was extracted from blood. DNA methylation profiling was performed using the Illumina Infinium HumanMethylation 450k chip, which measures about 480,000 different CpG sites covering 96% of RefSeq genes. It provides comprehensive gene region coverage, targeting multiple sites including the promoter, 5' UTR, 1st exon, gene body and 3' UTR. The methylation level at each CpG site was measured by β values varying from 0 (no methylation) to 1 (100% methylation). The significance level was defined as $p < 1.3 \times 10{-}7$, after Bonferroni correction for multiple testing. Results: We found 539 individual CpG sites significantly associated with OA, with a mean difference in methylation level of >10%. Among them, 516 were hypermethylated and 23 were hypomethylated in OA patients. These individual CpG sites represent 322 independent gene regions. Although methylation patterns may be tissue specific, 15 of the CpG regions that we identified were recently found to also be differentially methylated in articular chondrocytes from OA patients (Ann Rheum Dis 2013 Mar 16). Conclusion: We demonstrate that DNA methylation pattern in blood permit differ between OA patients and healthy controls. This may be valuable for identifying those at high risk for OA and the development of effective treatment for OA.

416T

Diurnal Rhythms of Clock Gene DNA Methylation and their Relationship to Rhythms of Clock Gene Expression in the Human Cerebral Cortex.
A.S. Lim[1], G.P. Srivastava[2,3], L. Yu[4], A.S. Buchman[4], J.A. Schneider[4], A.J. Myers[5], D.A. Bennett[4], P.L. De Jager[2,3]. 1) Division of Neurology, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; 2) Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 5) Department of Psychiatry, University of Miami, Miami, FL.

BACKGROUND: The mammalian circadian clock is regulated by a highly conserved series of 'clock' genes participating in a near 24-h transcription-translation negative feedback loop. Recent work in Neurospora has suggested that rhythms of DNA methylation in or near clock genes may play a role in regulating the core circadian clock. However, whether such rhythms are also present in human tissues, and how they relate to clock gene expression, is uncertain. METHODS: We quantified DNA methylation at 128 CpG sites in or near 6 canonical clock genes - PER2, PER3, CRY1, CRY2, ARNTL, and CLOCK - using Illumina Infinium HumanMethylation450 microarray data from dorsolateral prefrontal cortex samples from 753 deceased individuals in 2 cohort studies of older individuals, the Religious Orders Study and the Rush Memory and Aging Project. We quantified transcript abundance for these genes using Illumina Human HT-12 Expression microarray data from a subset of 490 of these individuals. Transcript abundance and methylation level at each CpG site was parameterized as a function of time of death using cosine curves. RESULTS: Significant daily rhythms of methylation were seen in 63/128 CpG sites ($p<0.05$), and in the expression of PER2 ($p=3.1\times10^{-4}$), PER3 ($p=2.5\times10^{-4}$), CRY1 ($p=7.0\times10^{-4}$), ARNTL ($p=5.9\times10^{-5}$), and CLOCK ($p=7.3\times10^{-3}$) with the timings of peak transcript abundance mirroring those seen in other diurnal mammals. The timing of the nadir of methylation was site specific. For rhythmic CpG sites within the gene body, the timing of the nadir of methylation clustered between 16:00 and 22:00, irrespective of the timing of transcript abundance. However, for sites upstream of the promoter region and in the 5'UTR, the timing of the nadir of methylation was roughly in-phase with the timing of peak abundance of the corresponding transcript showing temporal correlation between hypomethylation and expression. Finally, significant sex differences were seen, with transcript abundance peaking later in men than in women for all 5 rhythmic clock genes, and the nadir of methylation occurring later in men than in women for 22 of the 63 rhythmic CpG sites. CONCLUSIONS: There are site-specific diurnal rhythms of DNA methylation in and near canonical clock genes, with characteristic site-specific phase relationships to transcript abundance and to global circadian rhythms. Rhythms of DNA methylation may play a role in the regulation of the circadian clock in human cerebral cortex.

417F

DNA methylation analysis of human neuronal and non-neuronal cells at the base pair resolution.
J. Ueda[1], M. Bundo[2], F. Sunaga[2], M. Nishioka[2,3], E. Hashimoto[4], W. Ukai[4], T. Saito[4], K. Kasai[3], K. Iwamoto[2], T. Kato[1]. 1) Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Wako, Japan; 2) Department of Molecular Psychiatry, Graduate School of Medicine, The University of Tokyo,Tokyo, Japan; 3) Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, , Tokyo, Japan; 4) Department of Neuropsychiatry, School of Medicine, Sapporo Medical University, Sapporo, JAPAN.

Epigenome information such as DNA methylation and histone modification in the human brain cells reflects their developmental history, neuronal activity, and environmental exposures. Studying the epigenetic modifications in the brain cells is critical to understanding of the role of the genome in brain functions, as well as the pathophysiology of psychiatric disorders. We have previously revealed distinctive DNA methylation patterns of neuronal and non-neuronal cells, those were derived by neuronal marker-based nuclei sorting technique, with promoter-wide tiling arrays. Here we extended our DNA methylation study at the base pair resolution using whole genome-bisulfite sequencing (WGBS). Neuronal and non-neuronal genomic DNA derived from a postmortem prefrontal cortex of subject with no neuropsychiatric disorders was used for bisulfite modification. We obtained 939 and 784 million paired-end reads from sequencing of neuronal and non-neuronal cells, respectively. After mapping with Bismark software, average coverage on CpG sites were 23X for neuronal cells and 17X for non-neuronal cells. Gene Ontology analysis of differentially methylated CpGs between neurons and non-neurons revealed that hyper-methylated genes in neurons compared to non-neurons were enriched to the GTPase and signal transduction-related terms. Strikingly, hypo-methylated genes in neurons compared to non-neurons were enriched to the synapse-related terms. In addition, by comparing DNA methylation data obtained from different platforms including tiling array, reduced representation bisulfite sequencing (RRBS), Illumina HumanMethylation450, and WGBS, we identified characteristics and advantage of each platform.

418T

Methylation of Leukocyte DNA and Ovarian Cancer: Relationship with Disease Status and Outcome.
B. Fridley[1], S. Armasu[2], M. Cicek[2], K. Kalli[2], M. Larson[2], D. Koestler[3], D. Rider[2], V. Shridhar[2], J. Olson[2], J. Cunningham[2], E. Goode[2]. 1) University of Kansas Medical Center, Kansas City, KS, USA; 2) Mayo Clinic, Rochester, MN, USA; 3) Dartmouth College, Lebanon, NH, USA.

We hypothesize that DNA methylation (DNAm) in leukocytes of epithelial ovarian cancer (EOC) cases and controls may differ and that, among cases, this DNAm may vary by disease outcome. Thus, we performed case-control and survival analyses using 336 EOC cases and 398 controls (3 experimental batches) with blood-based DNAm assayed with the Illumina Infinium HumanMethylation27 or 450 BeadChips. To limit spurious results, we limited analysis to CpG probes measured on both arrays, and we removed non-specific probes and probes associated with distributions of white blood cell types. The association of the DNAm levels for 24,520 CpG sites with disease status and overall survival (OS) was completed using linear and Cox-proportional hazards models. Meta-analysis across three batches was completed using a random effect meta-analysis. The top association between DNAm and case-control status was observed for CpG cg04834572 near DUSP13 ($p=1.6\times10^{-14}$, individual batch p range from 2.1×10^{-4} to 1.1×10^{-6}). The top pathway enriched, based on Fisher's Exact test, was the telomerase signaling ($p=1.24\times10^{-3}$). This pathway and variants in TERT have been found to be associated with the development of many cancers, including EOC. TERT has an extensive role in the maintenance of functional telomeres which are important in the protection of chromosomes from DNA damage. The top CpG sites associated with OS were: cg10276549 near GABRE ($p = 5.8\times10^{-5}$); cg06171242 near ACOT13 and TTRAP ($p=4.4\times10^{-4}$); and cg14360897 near RNF167 and SLC25A11 ($p=5.9\times10^{-4}$). The most enriched pathways were: relaxin signaling ($p=7.09\times10^{-5}$) and CXCR4 signaling ($p=1.25\times10^{-4}$) and IL-8 signaling ($p=3.05\times10^{-4}$), all containing GNB1, GNA12, and PIK3R4. GNA12 has been found to be associated with response to platinum/taxane combination therapy, while PIK3R4 is a member of the phosphoinositide 3-kinases family which is involved in multiple cell functions, including proliferation and cell survival. Via a CpG-by-CpG approach accounting for CpGs known to correlate with white blood cell types, we identified genes with differential DNAm signals by EOC disease status and OS. Although pre-treatment case samples were used, a study limitation is the lack of prediagnostic blood samples for EOC. Thus, we are unable to conclude that the CpGs are involved in disease etiology. Future research is needed to functionally determine the biological and epigenetic relevance of the detected CpG sites associated with EOC status and OS.

419F

The combination of genome wide screening and DNA methylation in the determination of bio-marker discovery in ovarian and breast cancer.
M. Poulin, A. Meyer, L. Yan. EpigenDx, Hopkinton, MA.

The association of genetic abnormalities and specific cancers is leading to an increasing number of bio-markers that can be used for early detection of cancers, determination of treatment protocols and predicted prognoses. Additionally, the methylation state of an increasing number of genes is also producing possible predictive bio-markers for specific cancers and stages of cancer as well. To this end, we have analyzed both breast and ovarian cancers and their normal adjacent tissue using both DNA methylation analysis and genome wide array screening for LOH and copy number variations. The combination of these two methods has yielded several groupings of markers that are cancer specific as well as showing specificity for the specific tumor type. We will present the results of the combination of the genome wide analysis and DNA methylation results that may be used in combination to produce potential biomarkers for these two cancer types.

420T

Independent contribution of epigenetic modifications within lipoprotein metabolism genes to plasma lipid profile variability. *S.P. Guay[1,2], D. Brisson[2,3], B. Lamarche[4], D. Gaudet[2,3], L. Bouchard[1,2].* 1) Department of Biochemistry, Université de Sherbrooke, Sherbrooke, Qc, Canada; 2) ECOGENE-21 and Lipid Clinic, Chicoutimi Hospital, Saguenay, Qc, Canada; 3) Department of Medicine, Université de Montréal, Montréal, Qc, Canada; 4) Institute of Nutrition and Functional Foods, Université Laval, Québec, Qc, Canada.

Background: Inheritance plays a central role in the determination of plasma levels of lipids by explaining up to 60% of the interindividual variability. However, the gene polymorphisms identified so far explain less than 25% of the heritability of plasma lipid levels. Recent studies suggest that epigenetic modifications (DNA methylation), a non-traditional hereditary factor, could explain a significant proportion of the missing heritability of complex traits, such as plasma lipid levels. **Objective:** To assess whether the DNA methylation of key genes of the lipoprotein metabolism is associated with changes in fasting plasma lipid levels (high-density lipoprotein cholesterol (HDL C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG)) in patients with familial hypercholesterolemia (FH) carrying the same *LDLR* mutation (p.W66R). **Methods/Results:** In the current study, 98 untreated FH patients (61 men and 37 women) were recruited. Blood DNA methylation levels were measured at the *ABCA1*, *ABCG1*, *CETP*, *LCAT*, *LDLR*, *LIPC*, *LPL*, *PLTP* and *SCARB1* gene loci using bisulfite pyrosequencing. Partial Pearson's correlation analysis showed that DNA methylation levels at the *ABCA1*, *ABCG1*, *CETP*, *LIPC*, *LPL* and *PLTP* gene loci were significantly associated with HDL-C, LDL-C and/or TG levels in a sex-specific manner in FH (all p<0.05). Multivariate linear regression models showed that conventional predictors of the plasma lipid profile (age, anthropometric measurements, blood pressure and fasting glucose levels) explained 33.7% of HDL-C, 17.5% of LDL-C and 26.1% of TG level variability in men (all p<0.001), and 11.4% of HDL-C, 10.7% of LDL-C and 29.3% of TG level variability in women (all p<0.02). When epivariants were added to the statistical models, the variance explained by the combination of conventional and epigenetic factors was found to be increased for both sexes (49.9% for HDL-C, 38.2% for LDL-C and 33.0% for TG levels in men (all p<0.001), and 16.8% for HDL-C, 16.9% for LDL-C and 41.8% for TG levels in women (all p<0.02)). Together, these epivariants independently explained up to 15.8% of HDL-C, 24.0% of LDL-C and 17.4% of TG level variability in FH (all p<0.02). **Conclusion:** These results suggest that epigenetic perturbations of key lipoprotein metabolism genes are associated with fasting HDL-C, LDL-C and TG levels and might contribute to explain the interindividual variability as well as the missing heritability of plasma lipid levels among FH subjects.

421F

Abberant CpG island methylation of Caspase-3 gene in Uterine Leiomyomas. *S. Vaidya[1,2], K. Prabhakar Rao[2], Q. Hasan[1,2], N.A. Shaik[3].* 1) Genetics, Kamineni Hospitals, Hyderabad, Hyderabad, India; 2) Department Of Genetics,Osmania University, Hyderabad,india; 3) Department of Genetic Medicine & Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD).

Uterine Leiomyomas are one of the commonly seen benign growths of reproductive-age group women. These tumors are often associated with excessive bleeding, pelvic pain, infertility, and pregnancy complications in affected women. Role of hormones, genes and environmental factors highlights the multifactorial nature of these complex tumors. Although, numerous genetic alterations are closely associated with the pathogenesis of uterine leiomayomas, the underlying DNA methylation patterns have not been extensively studied in developing these reccurent tumors. Hence, to investigate if epigenetic factors could explain the significant deviations in apoptosis mechanisms, we have examined the methylation status of key apoptotic factors such as Bcl2 and Caspase-3 genes in the pathogenesis of leiomyomas collected from 200 uterine hysterectomies. Using Methylation Specific Restriction Assay, methylation status of P2 promoter region (–1001 to –690) of BCL2 gene and intronic CpG island of Caspase-3 genes were determined for both Leiomyoma and matched myometrium tissues. There were 22 CpG sites in the 312 bp fragment of Bcl2 and 10 CpG sites in the 174 bp Caspase-3 amplicons we have analyzed. Our results have showed that Caspase-3 intronic CpG island is hypomethylated by 2.6 folds in leiomyomas against adjacent myometrium tissues [P = 0.006; OR 2.629 ,95 % CI (1.290 - 5.398)], whereas, Bcl2 gene showed no changes in its methylation pattern [P = 1.000 ; OR 1.000,95 % CI (0.537 - 1.864]. The present data indicate the upregulation of Caspase-3 expression in fibroids suggesting dysregulated apoptotic process in these solid tumors. Furthermore, DNA methylation changes may not be playing a role for overexpression of Bcl2, which is known factor in the development of resistance to apoptosis in UL.

422T

DNA methylation profiling of preeclamptic placentas reveals distinct subgroups. *S.L. Wilson[1,2], J.D. Blair[1,2], C.W. Hanna[1,2], P. von Dadelszen[2,4], S. Langlois[1], D.E. McFadden[2,3], W.P. Robinson[1,2].* 1) Department of Medical Genetics, The University of British Columbia. Vancouver, BC,CA; 2) Child and Family Research Institute. 950W 28th Ave. V5Z 4H4, Vancouver, BC,CA; 3) Department of Pathology, The University of British Columbia. Vancouver,BC,CA; 4) Department of Obstetrics and Gynecology, The University of British Columbia. Vancouver,BC,CA.

Preeclampsia is a leading cause of maternal and perinatal death worldwide, complicating 2-8% of pregnancies. Early-onset preeclampsia (EOPET) arising prior to 34 weeks gestation tends to be more severe and associated with greater placental changes, as compared to late-onset preeclampsia (LOPET) arising after 34 weeks gestation. Intrauterine growth restriction (IUGR) often occurs conjunctively with preeclampsia and it has been suggested that normotensive IUGR (nIUGR) may be caused by similar placental malfunction. Recently, we showed extensive DNA methylation differences in EOPET placentas in comparison to controls. We hypothesized that some DNA methylation changes associated with EOPET placentas may overlap with LOPET and nIUGR associated placentas, but that these groups will also show unique signatures indicating distinct etiologies. The Illumina Infinium HumanMethylation450 BeadChip was used to investigate DNA methylation at greater than 485,000 CpG sites in chorionic villi from placentas associated with EOPET (N=20), LOPET (N=19), normotensive IUGR (N=15) and gestational-age and sex matched controls (N=47). Principal component analysis showed a major effect of gestational age, however, EOPET placentas also clustered distinctly from controls. Interestingly, the LOPET samples with conjunctive IUGR (N=11) clustered separately from those without IUGR (N= 8), the latter of which overlapped controls. The nIUGR group also clustered with the controls. When considering each subgroup compared to gestational-age and sex matched controls separately with a false discovery rate <5% and DNA methylation difference of >5% EOPET showed widespread DNA methylation changes. Fewer significant changes were noted in LOPET with IUGR. CpGs in 102 genes including ECE1 and CYP1A2 were found to be differentially methylated. No significantly different CpGs were identified when comparing nIUGR or LOPET without IUGR to controls. This study provides evidence that EOPET, LOPET, nIUGR, and compound LOPET & IUGR associated placentas have distinct molecular profiles. Further investigation into the differential methylation patterns between EOPET,LOPET, nIUGR and compound LOPET & IUGR may give insight to molecular pathogeneses resulting in preeclampsia and other placental disorders.

423F

MATRIX MODULATES DNA METHYLATION IN HUMAN BLADDER SMOOTH MUSCLE CELLS (BSMC). *K.J. Aitken[1], J.X. Jiang[2], C. Sotiropolous[2], S. Pu[1], T. Kirwan[1], D.J. Bagli[1,2,3].* 1) Sickkids Hospital, Toronto, Ontario, Canada; 2) Department of Physiology, University of Toronto, Toronto, Ontario, Canada; 3) Department of Surgery, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

Extracellular matrix changes are often crucial inciting events for fibroproliferative disease. Epigenetic changes, specifically DNA methylation, are critical factors underlying differentiated phenotypes of smooth muscle and fibroblasts. We examined the dependency of matrix-induced fibroproliferation and SMC phenotype on DNA methyltransferases. The cooperativity of matrix with growth factors, hypoxia and strain was also examined. Primary rat visceral SMC of early passage (0-2) were plated on native collagen or damaged/heat-denatured collagen. Hypoxia was induced with 3% O_2, 5% CO_2 and 95% N_2 over 48 hours. Inhibitors were applied 2-3 hours after cells were plated on matrix, or immediately before hypoxia. Cells were fixed and stained for DNMT3A and α-smooth muscle actin or smooth muscle myosin heavy chain. Western blotting and QPCR were also used to examine DNMT expression on the two matrix environments. Illumina 450K array of CpG sites was performed on bisulfite converted DNA from smooth muscle cells on damaged matrix vs native collagen. Differences in beta values were considered significant after correction for multiple testing by benjamini hochberg analysis, p<0.05. An *a priori* analysis of sites proximal to SMC-related genes was performed using t-test and benjamini hochberg analyses. Damaged matrix up regulates DNMT3A nuclear localization and expression, and influences differentiation in BSMCs exposed to denatured matrix +/– hypoxia. Analysis of DNA methylation signatures showed that matrix caused significant DNA methylation alterations in a discrete number of 14 CpG sites proximal to 12 genes related to SMC differentiation. Matrix has a profound effect on the regulation of SMC phenotype, which is associated with altered expression, localization of DNMTs and discrete changes DNA methylation.

424T

Dose-dependent effect of *in utero* smoking on DNA methylation among Latino children in a methylome-wide association study. *S.S. Oh[1], D. Hu[1], C.R. Gignoux[1], J.M. Galanter[1], S. Huntsman[1], D. Torgerson[1], C. Eng[1], L.A. Roth[1], A. Davis[2], H.J. Farber[3], P.C. Avila[4], E. Brigino-Buenaventura[5], M.A. LeNoir[6], K. Meade[2], D. Serebrisky[7], L.N. Borrell[8], W. Rodríguez-Cintrón[9], R. Kumar[10], J.R. Rodríguez-Santana[11], F. Lurmann[12], E. Burchard[1].* 1) UC San Francisco, San Francisco, CA; 2) Children's Hospital and Research Center Oakland, Oakland, CA; 3) Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 4) Feinberg School of Medicine, Northwestern University, Chicago, IL; 5) Kaiser Permanente-Vallejo Medical Center, Vallejo, CA; 6) Bay Area Pediatrics, Oakland, CA; 7) Jacobi Medical Center, Bronx, NY; 8) City University of New York, Bronx, NY; 9) Veterans Caribbean Health System, San Juan, Puerto Rico; 10) The Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 11) Centro de Neumologia Pediatrica, San Juan, Puerto Rico; 12) Sonoma Technology, Petaluma, CA.

It is known that *in utero* smoke exposure leads to changes in DNA methylation of specific genomic regions (CpG loci). However, the extent to which DNA methylation is affected by the amount of exposure is unknown. We hypothesized that DNA methylation patterns vary by the 'dose' of *in utero* smoking. To investigate the association of DNA methylation with *in utero* smoking, we conducted a methylome-wide association study (MeWAS) on 528 Latino children from the GALA II Study, a nation-wide case-control study of Latino children with and without asthma. Methylation status at >480,000 CpG loci was assessed using the Infinium HumanMethylation450 BeadChip. We used robust linear regression to test the association between CpG methylation and the number of trimesters that children were exposed to *in utero* smoking, adjusting for sex, age, ethnicity, asthma status, plate, position, and the first 10 principal components of variation within our data set.

The most significant differentially methylated locus in our MeWAS was in the first exon of *FAM83A*. For each trimester a mother smoked during pregnancy, her child had 2% less methylation ($p = 4.9E{-}7$; $FDR < 0.05$). Two additional suggestive loci include *ABL2* (1.2% less methylation per trimester, $p = 3.2E{-}6$; $FDR < 0.10$) and WNT3A (0.5% less methylation per trimester, $p = 4.4E{-}6$; $FDR < 0.10$). All three genes are involved in pathways known to be disrupted by tobacco smoke exposure. *FAM83A* is expressed in the lung, and *in vitro* studies have found it to be upregulated in bronchial epithelial cells following exposure to tobacco smoke. ABL2 functions in cytoskeletal rearrangements, and WNT3A is a key regulator of cell fate and patterning during embryogenesis.

In a MeWAS assessing the dose-response effect of *in utero* tobacco smoking, we found that *in utero* smoking was associated with altered methylation at three biologically relevant loci. Furthermore, we found that the degree of methylation was dose-dependent. Smoking during pregnancy is particularly insidious not only for harming the developing fetus but also for its effects manifested in later life. Our findings underscore the importance of tobacco prevention, control, and cessation efforts.

425F

DNA methylation profiling implicates several genes in type 2 diabetes. *M.A. Carless, H. Kulkarni, M.C. Mahaney, H.H.H. Goring, L. Almasy, A.G. Comuzzie, J. Blangero.* Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

Type 2 diabetes (T2D) is a global epidemic, becoming increasingly prevalent due to a rising incidence of obesity caused primarily by poor diet and lack of physical activity. The incidence of T2D is particularly high in Mexican Americans and although the disease is known to be genetically regulated, implicated loci explain only a small portion of the genetic liability. Epigenetic regulation, such as DNA methylation, is a novel mechanism that may lead to gene dysfunction and disease development. Using Illumina HumanMethylation450 BeadChips, we performed genome-wide DNA methylation profiling of >450,000 CpG sites in peripheral blood cells from 859 Mexican Americans from ~40 large pedigrees. In total, 20% of individuals were diagnosed with T2D and all individuals had fasting glucose measures available. Normalized DNA methylation data underwent analysis using SOLAR to test for heritability of each CpG site, and for association with T2D status and fasting glucose (in non-diabetic samples only). We used the Bonferroni method to correct for multiple comparisons. Approximately 24% of CpG sites were found to be significantly heritable (mean $h^2=0.433$, $p<1.06\times10^{-7}$). We identified 12 statistically significant associations between DNA methylation levels and T2D ($p<1.04\times10^{-7}$). The top association was for a CpG site in the *TXNIP* gene ($p=5.57\times10^{-18}$). Txnip is known to modulate glucose metabolism and insulin sensitivity, likely playing a role in T2D and possibly contributing to the therapeutic action of metformin. We also identified a significant association between diabetes and *ABCG1* ($p=5.45\times10^{-12}$), which has been implicated in reverse cholesterol transport, insulin secretion and glucose tolerance. Significant associations between DNA methylation and T2D status were also seen in other genes previously implicated in diabetes, including *CPT1A* ($p=5.92\times10^{-10}$) and *DHCR24* ($p=7.66\times10^{-9}$), as well as in potentially novel genes. Most of these genes were also at least nominally significantly associated with fasting glucose in the non-diabetic portion of our dataset (n=683). Our study has identified a number of genes that show DNA methylation changes associated with T2D in a Mexican American population. We are currently conducting pyrosequencing studies to validate these findings. We show here that DNA methylation profiling is a powerful means to identify genes involved in T2D, which may ultimately aid in the development of targeted therapeutics for diabetic treatment.

426T

Methylation analysis of CpG islands in the *TYR* and *P* genes: correlation with gene expression. *X. Wang[1], L. DSOUZA[1], H. HE[1], C. ANTOLIK[1], Q. ZHANG[1,2].* 1) OGVFB, NEI/NIH, Bethesda, MD; 2) Department of Immunology, Capital Medical University, Beijing, China.

Oculocutaneous albinism types 1 and 2 (OCA1 & OCA2) are autosomal recessive disorders of melanin biosynthesis characterized by hypopigmentation of hair, skin and eyes. OCA1 is caused by mutations in the gene that encodes tyrosinase (TYR; 11q14-q21). OCA2 is caused by mutations in the gene that encodes the P protein (OCA2; 15q11.2-q12). Mutations in these genes have been detected in a majority of OCA patients, but a small proportion do not have any detectable mutations or have only one identifiable mutation despite sequencing all the coding regions of the gene. Epigenetic phenomena have demonstrated roles in the regulation of gene expression. However, the role of DNA methylation in the development of OCA is not known. Three melanoma cell lines (A375, SK-28, and MNT1) and a retinal pigment epithelium cell line (ARPE-19) were previously found to contain different levels of pigmentation and gene expression. The A375 cell line has no TYR and P gene expression and pigmentation. The SK-28 cell line has no pigmentation but TYR and P gene expression. The MNT1 cell line has both TYR and P gene expression as well as pigmentation, and the ARPE-19 cell line has pigmentation and TYR gene expression but no P gene expression. We hypothesized that the DNA methylation status in the candidate CpG islands around promoter regions of the TYR and P genes may play a role in the regulation of gene expression. Therefore, we analyzed TYR and P gene CpG methylation status by bisulphite sequencing of genomic DNA from these cell lines. We found CpG sites to be hypermethylated in TYR in the A375, SK-28 and ARPE-19 cell lines. In contrast, those CpG sites were hypomethylated in the MNT1 cell line. Furthermore, we found that only the A375 cell line was hypermethylated (all the CpG sites were methylated) for the P gene CpG island, and the rest of the cell lines had a hypomethylated CpG island (all the CG sites were unmethylated). Corresponding mRNA levels were also correlated by qRT-PCR. Further analysis should lead to a better understanding of the role of methylation in the regulation of TYR and P gene expression.

427F

The Role of Brain DNA Methylation in the Pathology of Alzheimer's Disease: Evidence of an Interaction Effect. *PL. De Jager[1, 2, 3], G. Srivastava[1, 2, 3], ML. Eaton[3, 4], L. Yu[5], A. Meissner[3, 6], JA. Schneider[5], M. Kellis[3, 4], DA. Bennett[5], LB. Chibnik[1, 2, 3].* 1) Institute for the Neurosciences, Departments of Neurology & Psychiatry, Brigham & Women's Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA; 5) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 6) Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, Cambridge, MA.

The DNA methylome captures the transcriptional potential of a cell or tissue. Differential methylation of validated Alzheimer's disease (AD) loci could influence their effect; however where methylation falls along the disease pathway is unclear. We examine four causal models to assess the role of methylation on the pathology of AD. We utilized data from two longitudinal cohorts, the Religious Order Study and Rush Memory and Aging Project. DNA methylation profiles were generated in samples of dorsolateral prefrontal cortex using Illumina HumanMet450K beadset. We analyzed CpG sites within 25 kb of 11 validated AD susceptibility genes. The outcomes of interest were a measure of neuritic plaque (NP) accumulation and pathologic diagnosis of AD. First, independent associations between CpGs and outcomes were assessed using linear (NP) and logistic (AD) regression. The four hypothesized models are: (1) CpG mediated association, (2) reverse causality, (3) independent associations and (4) SNP by CpG interaction. Both (1) and (2) were assessed using mediation analyses and (4) was assessed for multiplicative interaction followed by stratified analyses. Correction for multiple testing was done using the Benjamini-Hochberg method. A total of 719 subjects were included in the analyses. Nine CpGs across 5 genes (*BIN1* (3), *CLU* (2), *MS4A6A* (2), *ABCA7* (1) and *APOE* (1)) were associated with the outcomes. Together they explain 13.1% of the variability of NP and 14.8% of AD. No gene region showed evidence for either model (1) or (2), however *APOE* showed evidence of (3) with 1 CpG associated with the outcomes but not the APOE haplotype. Most interestingly, a strong interaction effect was seen with the *CR1* in 3 CpGs at the 5' end of the gene, cg10021878 ($p_{interaction}$ = 4.5E−5, 1.3E−4), cg00175709 ($p_{interaction}$ = 0.003, 0.01) and cg05922028 ($p_{interaction}$ = 0.004, 0.04) for NP and AD respectively. Among those with the risk allele rs6656401$^{AT/AA}$ there is an inverse association between methylation and outcome, indicating more methylation is associated with less NP and decreased odds of AD, whereas those with rs6656401^{TT} more methylation is associated with more NP and increased odds of AD. Less significant interactions were also seen in *BIN1*, *ABCA7* and *PICALM*. These observations suggest that, within known AD susceptibility genes, methylation is related to pathologic processes associated with AD and may play a role in influencing gene expression from susceptibility loci.

428T

Array-based assay detects genome-wide 5-methylcytosine and 5-hydroxymethlycytosine in non-human primates and mice. *R.S. Alisch[1], P. Chopra[2], L.A. Papale[1], A.T.J. White[1], A. Hatch[1], P.H. Roseboom[1], M. Brown[1], S.T. Warren[2].* 1) Psychiatry, Univ. of Wisconsin School of Medicine - Madison, Madison, WI; 2) Human Genetics, Biochemistry, and Pediatrics, Emory University School of Medicine, Atlanta, GA.

Murine and non-human primates (e.g. rhesus monkeys) represent excellent model systems to study human health and disease conditions, especially in the brain. However, use of these model systems for genomic profiling studies is limited because most array-based tools have only been developed to survey the human genome. Here we present the optimization of a widely used human DNA methylation array, designed to detect 5-methylcytosine (5-mC), and show that non-human data generated using the optimized array reproducibly distinguishes tissue types within and between chimpanzee, rhesus, and mouse, with correlations near the human DNA level ($R^2 > 0.99$). While using this assay to conduct a genome-wide methylation analysis of rhesus placental and fetal tissues reveals 6,102 differentially methylated loci with pathways analysis significantly overrepresented for developmental processes, restricting the analysis to oncogenes and tumor suppressors genes finds 125 differentially methylated loci, suggesting that rhesus placental tissue carries a cancer epigenetic signature. Further optimization of the assay to detect 5-hydroxymethylcytosine (5-hmC) finds highly reproducible 5-hmC levels within human, rhesus, and mouse brain tissue that is species-specific with a hierarchical abundance among the three species (human > rhesus >> mouse). Together, these data show that this array-based methylation assay is generalizable to all mammals for the detection of both 5-mC and 5-hmC, greatly improving the utility of mammalian model systems to study the role of epigenetics in human health, disease, and evolution.

429F

Airborne Particulate Matter Exposure Modifies the Canonical MAP-Kinase Pathway: From Methylomic Analysis to Biological Implications. *J.J. Carmona[1,2,3,4], T. Sofer[4,5], L. Cantone[6], B. Coull[5], A. Maity[7], J. Schwartz[1,2,3], X. Lin[4,5], A. Baccarelli[1,2,3,4,8,9].* 1) Laboratory of Environmental Epigenetics, Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; 3) Exposure, Epidemiology, & Risk Program, Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA; 4) Program in Quantitative Genomics, Harvard School of Public Health, Boston, MA, USA; 5) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 6) Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, IT; 7) Department of Statistics, North Carolina State University, Raleigh, NC, USA; 8) Dana-Farber/Harvard Cancer Center, Boston, MA, USA; 9) Harvard/Massachusetts General Hospital Center on Genomics, Vulnerable Populations, and Health Disparities, Boston, MA, USA.

Background: Exposure to air particulate matter with an aerodynamic diameter < 2.5 m (PM2.5) is known to elevate blood markers of inflammation and increase cardiopulmonary morbidity and mortality. Major components of PM2.5 are Black Carbon (BC) due to traffic and sulfate from coal-burning power plants. DNA methylation is known to be sensitive to environmental toxins and to mediate environmental effects on clinical outcomes via regulation of gene expression. We hypothesize that exposure to air pollution components affects DNA methylation in blood leukocytes, in genes from inflammatory pathways. Methods: 141 males from the Normative Aging Study (NAS), residing in the Boston area, were selected. Leukocyte DNA samples were hybridized to the RefSeq 385K Promoter tiling array (Roche NimbleGen, Madison, WI) representing the promoter regions of all well-characterized genes in the RefSeq database, as well as all of the UCSC-annotated CpG islands. Sample immunoprecipitation, labeling, hybridization and data extraction were performed according to standard procedures optimized by Roche-NimbleGen. Air pollution components, BC and sulfate, were measured with a sensor located on the roof of the Countway Library at the Harvard Longwood Campus. 30-days moving average values of BC and sulfate were calculated for each participant at his clinical visit date. Sulfate measures were available for 92 of the individuals. Genes associated with the MAP-kinase and NFkB signaling pathways were identified using the BIOCARTA website. Forward stepwise Canonical Correlation Analysis was applied to identify specific genes in the pathways associated with the exposures, and p-values were calculated using a permutation test. Analysis for the effect of BC on the MAPK pathway was adjusted for age and sulfate exposure, and similarly for other pathways/exposures. Results: The MAPK pathway consists of 84 genes. Our analysis identified 10 genes whose methylation was associated with BC exposure, adjusted for sulfate and age (p-value 0.01). The association analysis between sulfate and methylation in this pathway suggested 9 genes, but was not statistically significant (p-value 0.086), which is possibly due to low power. There was no evidence of association between air pollution and methylation in the NFkB pathway. Conclusion: The effects of air pollution may influence inflammatory outcomes via MAPK gene methylation. These results will be validated in a larger subset of men from the NAS cohort.

430T

Epigenomic fetal programming: identifying genomic sites differentially methylated after exposure to maternal gestational diabetes and responsive to its treatment. *A.A. Houde[1,2], S.M. Ruchat[1,2], C. Allard[3], P. Perron[2,3], J.P. Baillargeon[3], J. St-Pierre[2,4], D. Gaudet[2,5], D. Brisson[2], M.F. Hivert[3,6,7], L. Bouchard[1,2].* 1) Department of Biochemistry, Université de Sherbrooke, Sherbrooke, QC, Canada; 2) ECOGENE-21 and Lipid Clinic, Chicoutimi Hospital, Saguenay, QC, Canada; 3) Department of Medicine, Université de Sherbrooke, Sherbrooke, QC, Canada; 4) Department of Pediatrics, Chicoutimi Hospital, Saguenay, QC, Canada; 5) Department of Medicine, Université de Montréal, Montréal, QC, Canada; 6) Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA; 7) Massachusetts General Hospital, Boston, MA.

Background: *In utero* exposition to gestational diabetes mellitus (GDM) is associated with increased lifelong susceptibility to obesity and metabolic disorders for the offspring. Recent evidences showed that epigenetic modifications may be involved in the metabolic health programming of the newborn exposed to GDM. Nevertheless, whether the treatment of GDM women (diet or diet+insulin) has an impact on DNA methylation levels has not been established. **Hypothesis:** DNA methylation at specific gene locus in placenta and cord blood is affected by exposure to GDM and its treatment. **Methods:** Placenta and cord blood samples were obtained from 43 women: 14 with normoglycemia (NGT) and 29 with GDM treated with diet (n=16, GDM-D) or diet+insulin (n=13, GDM-I). GDM was diagnosed between weeks 24-28 of pregnancy according to WHO criteria. DNA methylation was assessed at >485 000 CpG sites using the Infinitum HumanMethylation450 BeadArray. DNA methylation differences between the 3 groups were determined using ANCOVAs (adjustment for infant sex, gestational age, maternal BMI at 1st trimester of pregnancy and history of GDM (P_{adj})) and significance of pairwise comparisons was verified with Tukey's test. **Results:** Women were on average 29 years old. GDM-I women were slightly overweight at 1st trimester (BMI=27.1 kg/m^2) in comparison to GDM-D and NGT (23.9 and 24.6 kg/m^2) (P=0.05). In placenta exposed to GDM, lower levels of methylation were observed at *TOX2* ($P=6.8\times10^{-7}$; $P_{adj}=1.2\times10^{-6}$) and *DPP6* ($P=8.8\times10^{-6}$; $P_{adj}=1.9\times10^{-5}$) compared to NGT-exposed placenta. At *PLB1* locus, lower methylation levels were observed in placenta from GDM-D women (77.5% in NGT vs 73.6% in GDM-D; ANCOVA $P_{adj}=9.6\times10^{-6}$), whereas levels from GDM-I (78.2%) were similar to those from women with NGT. Similarly, in cord blood, we observed that CpG sites near *GNASAS* ($P_{adj}=8.5\times10^{-5}$) and *MYH7B* ($P_{adj}=4.5\times10^{-5}$) were differentially methylated in offspring from GDM-D mothers and were similar in NGT and GDM-I groups suggesting that insulin treatment offsets the impact of GDM exposure at these loci. In cord blood, methylation levels at *STC2* were lower in offspring from mother with GDM, treated either with diet or insulin ($P_{adj}=4.0\times10^{-5}$). **Conclusion:** Our results suggest that exposure to GDM or its treatment may influence DNA methylation at specific locus in the placenta and cord blood. The choice of clinical management of GDM may therefore have long lasting effect on the offspring's epigenome and metabolic health.

431F

Convergence of genetic, epigenetic and environmental factors on CpG-SNPs associated with human disorders: implications for transcriptional regulation in human brain. *D.R. Sarkisyan, I. Bazov, M.M.H. Taqi, H. Watanabe, O. Kononenko, V. Tashbulatov, T. Yakovleva, G. Bakalkin.* Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden.

Interaction between the genome, epigenome and environment is critical for the development of complex disorders. Expression of the genome may be influenced by the environment by shaping epigenetic mechanisms. At CpG-SNP sites, the genetic factors such as SNPs converge with environmentally-influenced epigenetic marks such as CpG dinucleotides, sites for cytosine methylation. The average occurrence rate of SNPs at a CpG site is 10-fold higher than the overall SNP occurrence rate, thus CpG-SNPs are overrepresented in the human genome (Xie et al., 2009). CpG dinucleotides are 7-fold more abundant at SNP sites than expected (Tomso and Bell, 2003). Correlation analysis identified allele specific methylation (ASM) on 30% heterozygous SNPs, and found that up to 88% of ASM regions are dependent on the presence of CpG-SNPs (Kerkel et al., 2008; Shoemaker et al., 2010). Some CpG-SNPs may be associated with a disease, and alterations of their methylation under environmental influences may be a critical factor affecting gene expression and contributing to disease vulnerability. Such CpG-SNPs may be located within DNAse I hypersensitivity sites (DHSs) and targeted by generic and sequence-specific regulatory factors (RF). The CpG-SNP hypothesis received support in recent studies by us and others (John et al., 2011; Kaminsky et al., 2011; Martin-Trujillo et al., 2011; Reynard et al., 2011; Taqi et al., 2011; Ursini et al., 2011). We address this hypothesis by comparing abundance of CpG-SNPs and non-CpG SNPs at DHSs and binding places for 161 RFs in a panel of 91 cell lines (ENCODE releases 1-3). We compared SNPs associated with alcohol dependence, behavioral disorders, neurodegenerative disorders, and the rest of SNPs implicated by NHGRI GWAS Catalog. As a model gene we analyzed methylation of CpG-SNPs in prodynorphin (PDYN), coding for opioid dynorphin peptides. Three PDYN CpG-SNPs were significantly associated with alcohol dependence and differentially methylated in human brain. In the brain of alcoholics, methylation of the C allele of SNP rs2235749 (3'-UTR; C>T; C is non-risk variant) was increased (P<0.001) and positively correlated with PDYN mRNA and dynorphin peptides (P<0.05). A DNA-binding factor that differentially targets the T, risk allele, and the methylated and unmethylated C allele of this SNP, was identified. This findings support CpG-SNP hypothesis. Supported by Swedish FAS, VR and FORMAS. Tissues were received from the NSWTRC, University of Sydney.

432T

Methylation analysis in tongue tissue of BWS patients identifies the (epi)genetic cause in 3 patients with normal methylation levels of H19 and KCNQ1OT1 in blood. *M. Alders[1], S.M. Maas[2], D.M. Kadouch[3], K. van der Lip[1], H.J. Bliek[1], C.M.A.M. van der Horst[4], M.M.A.M. Mannens[1].* 1) Clinical Genetics, Academic Medical Center , Amsterdam, Netherlands; 2) Paediatrics, Academic Medical Center, Amsterdam, The Netherlands; 3) Dermatology, Academic Medical Center, Amsterdam, The Netherlands; 4) Plastic and Reconstructive Surgery, Academic Medical Center, Amsterdam.

Beckwith Wiedemann Syndrome (BWS) is caused by aberrant imprinting of genes in the chromosome 11p15 region. A molecular cause for BWS is found in approximately 80% of patients while 20% remains unexplained. We analyzed the imprinting status of H19 and KCNQ1OT1 in resected tongue tissue of 11 BWS patients, three of which had normal methylation levels in blood. In 8 patients with known methylation defects in blood (6 isolated KCNQ1OT1 hypomethylation, 2 UPD), the same defects were detected in tongue. The methylation levels of H19 in tongue were comparable to those found in blood. The methylation levels of KCNQ1OT1 were relatively higher in tongue than in blood in most patients. In all three patients with normal methylation levels in blood aberrant methylation patterns were found in tongue tissue. In two patients a UPD was detected and the third case had hypermethylation of H19. This result shows that tissue specific mosaic (epi)genetic changes, not present in blood, is the underlying defect in at least a subset of BWS patients without a molecular diagnosis after standard genetic testing.

433F

An Integrated epigenomic-transcriptomic-genetic analysis of schizophrenia brain identifies novel molecular pathways to disease. *J. Mill[1,2], R. Pidsley[2], J. Viana[1], A. Jeffries[2], C. Wong[2], C. Troakes[2], L. Schalkwyk[2].* 1) Exeter University, Exeter, Devon, United Kingdom; 2) Institute of Psychiatry, King's College London, London, United Kingdom.

Schizophrenia (SZ) is a common psychiatric disorder characterized by the presence of psychotic symptoms and altered cognition. Although SZ is highly heritable, the molecular etiology of the disease is largely unknown. In addition to genetic and structural genomic variation, recent evidence supports a role for altered epigenomic and transcriptomic processes in disease pathogenesis. Frontal cortex and cerebellum tissue was obtained from 23 schizophrenia patients and 24 healthy controls. Genome-wide DNA methylation, expression and SNP profiling were performed using the Illumina Infinium Human Methylation450, HumanHT-12v4 Expression, HumanOmniExpress BeadChips respectively. Integrated multi-level analyses provide evidence of SZ-associated DNA methylation and gene expression changes at biologically relevant loci, including GABBR1, RASA3, C8A, NRN1, BNIP3, GAD1 and SERPINA3. Furthermore we identify cis-eQTLs and cis-mQTLs at SZ candidate genes nominated from published GWAS analyses, an increased burden of CNVs in patients with SZ, and a rare NRXN1 deletion in an SZ patient that is associated with altered DNA methylation. Together these results provide important insights into the biological mechanisms underlying SZ and highlight the value of taking an integrated '-omics' approach to complex disease.

434T

Whole Genome Bisulfite Sequencing of Cell Free DNA and its Cellular Contributors Links Placenta Hypomethylated Domains to Gene Deserts. *T. Jensen[1], S. Kim[1], C. Chin[1], Z. Zhu[1], T. Lu[1], C. Deciu[1], D. van den Boom[2], M. Ehrich[2].* 1) Research and Development, Sequenom Center for Molecular Medicine, San Diego, CA; 2) Research and Development, Sequenom, San Diego, CA.

Circulating cell free (ccf) DNA is useful for non-invasive diagnostic testing in prenatal health and oncology. In both cases, the nucleic acid of interest is the minority species and thus needs to be differentiated from the highly abundant ccf DNA background. DNA methylation can serve as a method for distinguishing these; however, this depends on an in depth knowledge of the DNA composition. Whole genome bisulfite sequencing (WGBS) was performed on a set of unmatched samples including ccf DNA from 8 non-pregnant (NP) female donors, genomic DNA from 7 buffy coat and 5 placenta samples, and ccf DNA from 7 pregnant females to gain a comprehensive understanding of the ccf DNA methylome in pregnant plasma. We first created a methylome map of ccf DNA from non-pregnant donors at single base resolution. Consistent with previous work in differentiated cell types, almost all cytosine methylation in NP ccf DNA samples occurred in the CpG context. We also found CpG cytosines within longer fragments were more likely to be methylated, linking DNA methylation and fragment size in ccf DNA. Next, we performed a series of pairwise comparative analyses to identify differentially methylated regions (DMRs). Comparison of the methylomes of placenta and NP ccf DNA enabled the detection of greater than 50000 DMRs, the majority resulting from placenta hypomethylation. We found that >90% of these DMRs were located outside of CpG islands and were often associated with distinct histone tail modifications. Further investigation of the identified DMRs revealed the presence of large domains exhibiting consistent hypomethylation in placenta samples relative to NP ccf DNA across millions of consecutive bases. We found these domains to occur primarily in CpG poor gene deserts. DMRs identified when comparing placenta to NP ccf DNA were recapitulated when comparing pregnant ccf DNA to NP ccf DNA, confirming the ability to detect differential methylation in ccf DNA mixtures. Overall, these data enabled the generation of methylome maps for each sample type at single base resolution, identified a link between local DNA methylation and ccf DNA fragment length, provided comprehensive lists of DMRs between sample groups, and uncovered the presence of megabase-size placenta hypomethylated domains. Furthermore, we anticipate these results to provide a foundation to which future studies based upon discriminatory DNA methylation for non-invasive testing can be compared.

435F

DNA Methylation at CPT1A is Associated with Triglyceride Levels, BMI and WHR. *D.M. Absher[1], M.R. Irvin[2], S. Aslibekyan[2], J. Sha[2], L.L. Waite[1], D. Zhi[3], K. Stanton Thibeault[1], J. Ordovas[4], D.K. Arnett[2].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) Department of Epidemiology, University of Alabama, Birmingham, Birmingham, AL; 3) Department of Biostatistics, University of Alabama, Birmingham, Birmingham, AL; 4) Department of Epidemiology, Atherothrombosis and Imaging, Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain.

Epigenetic variation is thought to be a contributor to complex traits, and is likely to account for some of the missing heritability for phenotypes that have been incompletely explained by genetic variants. As the epigenome is dynamic and environmentally responsive, epigenetic modulators of dietary traits and responses are likely to be critically important to the development of cardiovascular disease and type-2 diabetes. We have undertaken an epigenetic analysis of GOLDN (Genetics of Lipid Lowering Drugs and Diet Network) study participants to identify DNA methylation patterns that contribute to lipid metabolic traits and related phenotypes. Using the Illumina Methylation450 array to measure DNA methylation at ~470,000 CpGs in CD4+ T-cells from 995 individuals in 183 families, we fit mixed effects regression models to identify CpGs where DNA methylation levels were strongly associated with baseline triglyceride levels (TG), body mass index (BMI), and waist-hip-ratio (WHR). We identified a CpG in the CPT1A (carnitine palmitoyltransferase 1A) gene that was significantly associated with all three traits ($p=6.9e{-28}$ for TG, $p=2.8e{-11}$ for BMI, and $p=4.7e{-11}$ for WHR), with ~2.4% of baseline TG variance explained by methylation at this locus. Given that CPT1A is an important regulator of fatty acid metabolism, we hypothesized that the impact of methylation on BMI and WHR would be mediated through its effects on TG. Regression models using TG as a covariate led to a reduced, but not eliminated, significance for BMI ($p=6.8e{-06}$) and WHR ($p=2.3e{-04}$), suggesting that CPT1A methylation has some effects on BMI and WHR that are independent of baseline TG. In addition to CPT1A, we also identified 7 other CpGs that achieved genome-wide significance with BMI exclusively, including CpGs near the CD38, AHRR and PGHDH genes. Furthermore, 1 additional CpG near RPS6KA2 was significantly associated with WHR.

436T

DNA Methylation Profiling is Robust in Different Tissue Types and Reveals Distinct Patterns Across Rheumatoid Arthritis Samples and Phenotypes. *L.F. Barcellos[1], X. Shao[1], E. Elboudwarej[1], A. Baker[1], E. Sinclair[3], L.A. Criswell[2].* 1) Div Epidemiology-SPH, Univ California, Berkeley, Berkeley, CA; 2) Div Rheumatology, Dept Medicine, Univ California, San Francisco, CA; 3) The UCSF-GIVI CFAR Immunology Core, Univ California, San Francisco, CA.

Rheumatoid arthritis (RA) is a chronic inflammatory disease with potential to cause substantial disability, primarily due to the erosive and deforming process in joints. RA etiology is complex with contributions from genetic and non-genetic factors. Epigenetic changes such as altered patterns of DNA methylation (DNAm), are also present in RA. Our goal was twofold: (1) to establish a protocol for performing DNAm profiling of specific immune cell populations isolated from large numbers of samples that is cost and labor efficient, and both accurate and reproducible; (2) to characterize DNAm profiles in RA cases and controls across multiple cell types to identify similarities and differences relevant to disease status, phenotypes and pathogenesis. We generated genome-wide DNAm profiles using Illumina HumanMethylation450 BeadChips (n~22,000 genes, 459704 sites, post QC) in PBMCs, CD14+ monocytes, CD19+ B cells, and CD4+ memory and naïve T cells in 90 individuals (60 female cases and 30 controls) from the UCSF RA Cohort. We investigated the impact of sample storage conditions on: cell count, purity, DNA yields, quality and stability of DNAm profiles. The five outcomes were compared between cells isolated from PBMCs, stained and sorted by FACS on the same day of blood collection, and cells isolated from PBMCs, stained and stored overnight in buffer at 4°C and sorted the second day following collection; all from the same individual. Results show overnight storage did not impact the above outcomes. Background subtraction and beta-mixture quantile normalization was applied to all case and control data. Mean DNAm levels were highly correlated between all cell types in both cases and controls; however, CD4+ memory T cells differed more from CD14+ monocytes ($r<0.94$) DNAm levels within putative RA genes identified through GWAS (43 genes; 883 CpG sites) were similar between cases and controls; however, some differences were noted. Hypermethylation in RA case B cells compared to controls for HLA-DRB1, CCR6, AFF3 and TAGAP CpG sites was observed, whereas, PTPN22 CpG sites were hypomethylated; IRF5 was hypomethylated in case CD4 naïve T cells compared to controls. Visualization of CpG DNAm profiles using multidimensional scaling revealed separation between cell types; results also suggest age and clinical phenotypic associations with DNAm. Our results emphasize the cell specificity of DNAm, which has important implications for understanding human health and disease.

437F

Gene networks for social cognition in Williams syndrome. *L. Dai[1], R. Weiss[2], J.R. Korenberg[1].* 1) Center for Integrated Neuroscience and Human Behavior, Brain Institute, Department of Pediatrics, Univ. of Utah, Salt Lake City, UT; 2) Human Genetics, Univ. of Utah, Salt Lake City, UT.

Williams syndrome (WS), a neurodevelopmental disorder with hypersocial behavior, results from a deletion of ~ 28 genes on 7q11.23, that ultimately disturbs the neural circuitry involving oxytocin and vasopressin. Although cognitive deficits and social-emotional features are ultimately due to the deleted genes, the critical downstream pathways are unknown. Previous studies used rare WS genetic events, partial deletions seen in <2% of WS, to parse DNA contributions, implicating GTF2I and GTF2IRD1 in social behavior, and showing that perturbations of transcriptional networks (34 WS patients and 18 typical controls), implicate disturbed function at synapse and dendritic spine. Here we address the possible role of epigenetic mechanisms in WS using genome-wide analyses of DNA methylation. We employed the Illumina Infinium HumanMethylation450 array to query the methylation of ~500K CpG/non-CpG sites in whole genome using whole blood DNA from 7 WS and 5 parents. The results revealed striking epigenetic changes throughout the genome, with differences from controls in 15,987 methylation sites (p<0.05), corresponding to 7478 genes. 72% of the significant changed methylated sites are in CpG islands of which 67.1% are hypermethylated in WS. Hyper and hypomethylated sites coexist within single genes. Further, the data indicate differences with parental origin and gender: hyper-methylation in maternal deletions (74.6%) vs 25.4% in paternals. Methylation also varies with gender, hypermethylation seen more in males (78.6%) than in females (21.4%), suggesting that the subtle differences in WS genes dramatically alter the epigenetic landscape of the genome, largely through imprinting mechanisms but also related to gender. Preliminary evidence on GTF2I, previously associated with WS social behavior (Dai et al., 2009) and imprinted gene expression (Collette et al., 2009), suggests that methylation effects in WS may affect gene expression and possibly phenotype. Finally, similar to autism spectrum disorders, WS, FZD9 and Wnt signaling, indicated by Partek pathway enrichment analysis, may provide a common pathway of disorders of social behavior. Analyses of larger cohorts are in progress. These data indicate that genes in the WS region alter genome wide epigenetic mechanisms and result in widespread modification of methylation, strongly implicating these in the disturbed development and brain function seen in WS.

438T

Identification of CpG-SNPs Associated with Osteoporosis. *H. Shen[1], C. Qiu[1], L. Zhang[1,2], C. Xu[1], H.W. Deng[1,2].* 1) Center for Bioinformatics and Genomics, Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; 2) Center of Systematic Biomedical Research, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Background Osteoporosis is an increasingly serious public health problem, affecting over 200 million people worldwide. Osteoporosis is mainly characterized by low bone mineral density (BMD), which is under strong genetic control. To date, over 50 BMD-associated loci have been identified, however, all these loci together account for no more than ~5% of BMD variation. In addition, despite the markers identified, specific functional variants are generally unknown. The introduction or disruption of CpG sites has been suggested as a potential mechanism through which some single nucleotide polymorphisms (SNPs) can affect gene function via epigenetics. To identify novel and/or potential functional variants which associated with osteoporosis, we conducted a meta-analysis of genome-wide association studies (GWAS) for BMD, specifically focusing on CpG-SNPs. Methods By leveraging the extensive catalog of common and rare genomic variants identified by the 1000 Genomes Project and our in-house pilot whole-genome high-coverage deep re-sequencing study of 44 Caucasians, we identified 2,197,763 CpG-SNPs throughout the human genome and imputed the genotypes for these CpG-SNPs in five cohorts with a total of 11,140 subjects. Meta-analysis was carried out by using the program METAL. Results Twenty CpG-SNPs in 6 genes, specifically, CCDC170, C7orf76, PTH2R, SHFM1, WNT4, and FAM3C, showed association with hip BMD at the genome-wide significance level (p<=5.0E−08). Moreover, we detected suggestive association (p<=1.0E−06) with hip BMD for additional 85 CpG-SNPs, which were distributed in 21 other genomic regions. Interestingly, most of the BMD-associated CpG-SNPs are mapped to candidate genes (e.g., ESR1, SOX6, WNT16, etc) that have been significantly associated with BMD in previous GWAS and/or have known functional roles in bone biology, suggesting SNP induced alterations of DNA methylation is likely to be an important mechanism mediating the genetic basis of osteoporosis. Conclusion In this study, we identified a number of SNPs that may contribute to BMD variation through altering DNA methylation levels of specific genes. These results provided novel insights into the genetic basis of osteoporosis, and highlighted the close connections between genetic and epigenetic mechanisms of complex diseases.

439F

DNA Methylation in Six Cell and Tissue Types in Sjögren's Syndrome Reveals Distinct Patterns Across Samples and Clustering Based on Disease Status. *A.S. Baker[1], D. Quach, BS[1], H. Quach, BS[1], E. Elboudwarej, MPH[1], L.F. Barcellos, PhD, MPH[1], L.A. Criswell, MD, MPH[2].* 1) Genetic Epidemiology and Genomics Lab, School of Public Health, University of California, Berkeley, CA, USA; 2) Department of Rheumatology, School of Medicine, University of California, San Francisco San Francisco, CA, USA.

Sjögren’s Syndrome (SS) is a chronic, multisystem autoimmune disease characterized by progressive destruction of the exocrine glands, with subsequent mucosal and conjunctival dryness. Increasing evidence supports a role for DNA methylation status in autoimmune disease risk and severity. Our goal was to characterize methylation profiles in SS subjects and healthy controls across multiple cell and tissue types to identify similarities and differences relevant to disease status and pathogenesis. We generated genome-wide DNA methylation profiles using Illumina HumanMethylation450 BeadChips in minor salivary gland biopsy tissue, fresh and banked peripheral blood mononuclear cells (PBMC), CD14+ monocytes, CD19+ B cells, and CD4+ T cells in ten participants (five cases and five controls) from the Sjögren’s International Collaborative Clinical Alliance (SICCA; http://sicca.ucsf.edu/; N01 DE32636) repository. Additionally, we characterized full genome SNP profiles. Sorting of freshly collected blood samples was performed using MACS®technology. DNA yields for all cell and tissue types were high; all samples were background subtracted and normalized with beta-mixture quantile (BMIQ). Methylation of individual measured CpG sites was stable across the six cell types for ~240,000 CpGs (variance< 0.001); however, variance > 0.10 for ~900 CpGs. Mean methylation was calculated within each cell type for the 473,929 CpGs detected by the BeadChip. Mean methylation was highly correlated between all cell types in cases, controls, and in a combined data set (r2: 0.91 - 0.99) with the lowest correlation between salivary gland tissue and CD14+ monocytes and the highest correlation between fresh and banked PBMCs. Methylation levels within 17 putative SS genes identified with GWAS were similar between cases and controls; however, median methylation levels for BLK and TNFSF4 CpGs were 10%–18% less in cases compared to controls within salivary gland tissue. Data visualization of methylation profiles using Multi-Dimensional Scaling revealed distinct separation between cell types in addition to separation based on case status with the greatest separation by case status visible in CD19+ B cells and salivary gland tissue samples. Our results emphasize the cell and tissue specificity of DNA methylation. Additional research, including studies of gene expression for regions associated with disease risk or severity, will be required to fully define the role of DNA methylation in SS.

440T

Epigenetics in cow's milk allergy: a Dutch epidemic. *P. Henneman[1], N.C.M. Petrus[2], A. Venema[1], M.M.A.M. Mannens[1], A.B. Sprikkelman[2].* 1) Clinical Genetics, , DNA-diagnositcs Laboratory, University of Amsterdam, Academic Medical Center, Amsterdam, Netherlands; 2) Pediatric Respiratory Medicine and Allergy, Emma Children’s Hospital, University of Amsterdam, Academic Medical Center, Amsterdam, The Netherlands.

Cow’s milk allergy (CMA) in Dutch infants is more prevalent (~10%) compared to infants from other countries in Europe (~2%). Children suffering from CMA have an increased risk on developing other allergic diseases in later life. Accumulating evidence suggests involvement of gut-microbiota and maturation of the immune-system in CMA. CMA is a complex disorder with an estimated heritability of 15%. Therefore early environmental factors are likely to be involved in CMA. It has been shown that epigenetic gene regulation can be altered by environmental factors. We aimed to detect epigenetic associations with the expression of CMA. We studied a sample of children from the Dutch EuroPrevall birth cohort study (N=20 CMA vs. N=20 controls), age and gender matched (50% girls). Inclusion criteria were based on questionnaires and challenge proven CMA. Bisulfite converted DNA (blood) was analyzed using the 450K infinium DNA-methylation array of Illumina®. Statistical analysis was performed in ‘R’ using IMA. Four groups were analyzed, combined, boys, girls and a subgroup of extremes. Subsequently, we performed pathway-analysis in ‘R’ using the global-test package. Methylation arrays were analyzed successfully. Associations with a P-value <2.6*10−4 were considered as potential candidates (pc), P-values < 2.6*10−6 were assumed significant. In the combined group we detected 13 pc-loci, in boys 15 pc-loci and 1 significant locus (ZNF281, P = 1.1E−06), in girls we detected 6 pc-loci and in the extremes 9 pc-loci. Eight pc-loci and the significant locus were found in both combined groups and males. Two loci overlapped between females and the combined group, suggesting robust association. Associations and pathway-analysis both detected involvement of chromatin-structure, epithelial maturation, intestinal function/absorption and immunological mechanisms. Although the limited statistical power is low in this small sample size study and the character of CMA is complex, we were able to detect plausible loci and pathways involved in chromatin remodeling, epithelial maturation, intestinal function and immunological anti-viral responses. Replication studies have to confirm our findings.

441F

Epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *E.R. Berko[1], M. Suzuki[1], F. Beren[2], C. Lemetre[1], C. Alaimo[3], R.B. Calder[1], K. Ballaban-Gil[4], B. Gounder[2], K. Kampf[2], J. Kirschen[1], S.B. Maqbool[1], Z. Momin[1], D.M. Reynolds[1], N. Russo[3,5], L. Shulman[6], E. Stasiek[1], J. Tozour[1], M. Valicenti-McDermott[6], S. Wang[7], B.S. Abrahams[1,8], J. Hargitai[1], D. Inbar[9], Z. Zhang[1], J.D. Buxbaum[10], S. Molholm[3], J.J. Foxe[3], R.W. Marion[6], A. Auton[1], J.M. Greally[1]*. 1) Center for Epigenomics and Department of Genetics (Division of Computational Genetics), Albert Einstein College of Medicine, Bronx, NY, 10461, USA; 2) Stern College for Women, Yeshiva University, New York, NY 10016, USA; 3) The Sheryl and Daniel R. Tishman Cognitive Neurophysiology Laboratory, Children's Evaluation and Rehabilitation Center, and Departments of Pediatrics and Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA; 4) Department of Neurology, Children's Hospital at Montefiore, Bronx, NY 10467, USA; 5) Department of Psychology, The College of Arts and Sciences, Syracuse University, Syracuse, NY 13244, USA; 6) Children's Evaluation and Rehabilitation Center, Department of Pediatrics, Albert Einstein College of Medicine. Bronx, NY 10461, USA; 7) Information Technology Services, New York University, New York, NY 10003, USA; 8) Department of Neuroscience, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Bronx, NY 10461, USA; 9) Child Development and Rehabilitation Institute, Schneider Children's Medical Center, Petach Tikvah, Israel; 10) Seaver Autism Center for Research and Treatment, Departments of Psychiatry, Neuroscience, and Genetics and Genomic Sciences, and the Friedman Brain Institute, Mount Sinai School of Medicine, New York, NY 10023, USA.

In this study we investigated the role of the epigenome, a possible mediator of environmental effects during development, in the pathogenesis of Autism Spectrum Disorders (ASDs). ASDs encompass a heterogeneous group of diseases characterized by impairments in communication and social interaction, and repetitive stereotyped behaviors. The prevalence of ASDs has increased dramatically, with current epidemiologic estimates citing a diagnosis in 1 out of every 90 individuals. Rates of ASD rise with parental age, with fathers older than 40 and mothers older than 35 each possessing greater independent risk of having a child with an ASD. Recent studies have helped further elucidate the genetic basis of ASDs, investigating the role of copy number, common, and rare variation in the disease. However, these findings explain only a fraction of ASD etiology.

We tested an homogeneous ectodermal cell type from individuals with ASD compared with typically developing (TD) controls born to mothers of ≥35 years, using a quantitative genome-wide DNA methylation assay. We show that DNA methylation patterns are dysregulated in individuals with ASD, performing a stringent analysis that accounted for confounding effects due to subject age, sex and ancestral haplotype, while also excluding mosaic aneuploidy and copy number variability in these subjects. Of note, the loci with altered DNA methylation were found at genes expressed in the brain, genes encoding protein products significantly enriched for interactions with those produced by known ASD-causing genes, representing a perturbation by epigenomic dysregulation of the same networks compromised by DNA mutational mechanisms.

The results indicate the presence of a mosaic subpopulation of epigenetically-dysregulated, ectodermally-derived cells in subjects with ASD. The epigenetic dysregulation observed in these ASD subjects born to older mothers may be associated with aging parental gametes, environmental influences during embryogenesis or could reflect mutations of the chromatin regulatory genes increasingly implicated in ASD. The results indicate that epigenetic dysregulatory mechanisms may complement and interact with DNA mutations in the pathogenesis of the disorder.

442T

Epigenetic and ecogenetic silencing of the FMR gene unrelated to CGG TNR expansion. *J. Kapalanga[1,3,4,5,6], Y. Said[3,7], D. Wong[2,3], A. Gandy[2,3], M. Moyo[5,6], N. Nkiru[4], A. Singh[3]*. 1) Dept Pediatrics, Genetics, Schulich School of Medicine, Western University/Grey Bruce Health Services, Owen Sound, ON, Canada; 2) Paediatrics, Dalhousie University, Halifax, NS Canada; 3) Summerside Medical Center, Summerside, PEI Canada; 4) Grey Bruce Health Services, Owen Sound, ON, Canada; 5) Cambridge Memorial Hospital, Cambridge, ON, Canada; 6) Pediatrics, McMaster University, Hamilton, ON, Canada; 7) Allergy and Pediatric Pulmonology King Fahad Specialist Hospital, Dammam,Saudi Arabia.

Expansions and hypermethylation of a CGG trinucleotide repeat (TNR) at the 5'untranslated region of the FMR1 gene results in transcriptional silencing of the gene. These molecular events have been established as the underlying cause of over 95% of patients with the Fragile-X syndrome (FXS). The commonest features of the FXS phenotype include intellectual disability, hyperactivity, impulsivity, a peculiar jovial speech, autistic features, macrocephaly, macrotia, and macroorchidism. In a behavioral and developmental pediatric practice it is common to see patients with the FXS phenotype but without molecular demonstration of the CGG TNR expansion in the abnormal range. In this study pediatric patients were ascertained through multicenter general, behavioral and developmental pediatrics clinics over a 10 year period. Patients were identified after presenting with a constellation of at least four common FXS phenotypic features including, intellectual disability, hyperactivity, autistic features, and distinctive craniofacial features. During the 10 year period 798 patients with the FXS phenotype, were tested for FXS by currently established molecular methods. Chromosomal studies were also done in all 798 patients. All patients were found to have CGG repeat size of 5 - 45 in their FMR1 gene and chromosomal studies were normal. Point mutations are rare (<1%) and would not be expected to be a significant cause of the FXS phenotype in these patients. These observations suggest that the FXS phenotype can occur unrelated to TNR expansion, chromosomal abnormalities or point mutations. It is conceivable that silencing of the FMR1 gene can occur by epigenetic mechanisms possibly modulated by ecogenetic factors even in the absence of CGG TNR expansion at the 5'untranslated region of the FMR1 gene. This implies that with a certain threshold of epigenetic and ecogenetic influences, silencing of a 'normal' FMR1 gene with repeat size of 5-45 can occur. It is also possible that epi-ecogenetic mechanisms interfere with production of a functional fragile X mental retardation protein (FMRP) by a putatively 'normal' FMR1 gene. We plan a comprehensive analysis of epigenetic influences on expression of the FMR1 gene with simultaneous assay of the FMRP. This could lead to an elucidation of the underlying etiopathogenic mechanisms that result in having the FXS phenotype unrelated to CGG TNR expansion.

443F

Multiple methylation errors at imprinting control regions in patients with S-adenosylhomocysteine hydrolase (AHCY) deficiency.
U. Zechner[1], A. Fitzner[1], J. Knežević[2], M. Polović[2], N. El Hajj[3], E. Schneider[3], R. Belužić[2], S.H. Mudd[4], T. Haaf[3], O. Vugrek[2]. 1) Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany; 2) Division of Molecular Medicine, Institute Ruer Bošković, Zagreb, Croatia; 3) Institute of Human Genetics, University of Würzburg, Germany; 4) Laboratory of Molecular Biology, National Institute of Mental Health, Bethesda, MD.

S-adenosylhomocysteine hydrolase (AHCY) deficiency is a novel human disease, which was first discovered in Croatia in 2004. Main characteristics are psychomotor delay and severe myopathy (hypotonia, absent tendon reflexes and delayed myelination) from birth, associated with hypermethioninaemia, elevated serum creatine kinase levels and increased genome-wide DNA methylation. The prime function of AHCY is the efficient removal of S-adenosylhomocysteine (SAH), the by-product of transmethylation reactions. As SAH is one of the most potent methyltransferase (MT) inhibitors, its rapid removal is crucial to avoid product inhibition of MTs. Thus, AHCY plays a critical role in regulation of biological methylation processes. We set out to more specifically characterize DNA methylation changes in blood DNA samples of seven AHCY-deficient patients as well as HepG2 and HEK293 cell lines after shRNA-mediated knockdown of the AHCY gene by determining the DNA methylation levels at differentially methylated regions (DMRs) of seven imprinted genes (*MEST*, *NESPAS*, *SNRPN*, *LIT1*, *H19*, *GTL2* and *PEG3*) as well as Alu and LINE-1 repetitive elements. Analysis of the imprinted gene DMRs revealed abnormal methylation levels with moderate to strong hypermethylation at several DMRs in three of the seven patients and rather normal differential methylation patterns in the other four patients. The knockdown cell lines also exhibited methylation changes to different degrees at the analyzed DMRs. Methylation analysis of Alu and LINE-1 repetitive elements demonstrated no methylation abnormalities. Microarray-based experiments to analyze the complete DNA methylome of AHCY-deficient patients in comparison to normal individuals are in progress. The finding of hypermethylation in the patients' DNA samples is opposite to what is expected considering the inhibitory effect of SAH on MTs. As an explanation for this finding, it can be speculated that only some MTs are inhibited and, thus, leave excess substrate for other specific DNA MTs not sensitive to increased SAH levels and functioning properly. Our preliminary data indicate that AHCY deficiency may represent a good model disease for studying the biological consequences of multiple methylation errors in epigenetic research. Thus, findings from this study may make an important contribution to develop standard and high-throughput tools for the diagnosis of AHCY deficiency and other diseases associated with aberrant epigenetic modifications.

444T

DNA differential methylation is observed at BRCA1 promoter but not in 8q24.21 in cleft lip and palate.
L. Alvizi, G.S. Kobayashi, C.B.F. Silva, D.Y. Sunaga, D.F. Bueno, M.R.S. Passos-Bueno. Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Sao Paulo, Brazil.

Purpose: DNA methylation is known to be a heritable regulatory mechanism in gene expression and influenced by both genetic and environmental factors. It is also known that impairment in gene methylation status may lead to gene expression dysregulation and thus disease. In this context, cleft lip and palate (CL/P) is a congenital craniofacial malformation with high incidence (1:700 live births) strongly determined by the genetic and environmental interplay in which epigenetic factors such as DNA methylation are very plausible factors in the malformation etiology. Aiming to investigate DNA methylation at specific sites to CL/P, we investigated whether BRCA1 , previously associated to CL/P (Kobayashi and Alvizi et al, 2013 PloS ONE), and 8q24.21 CL/P risk region were differentially methylated in CL/P samples in comparison to control samples. Methods: Bisulfite sequencing analysis for BRCA1 promoter was performed in a DNA sample set obtained from dental pulp stem cells (DPSC) of 18 CL/P and 12 controls and for 8q24 region in a DNA sample set from white blood cells DNA of 34 CL/P and 44 controls. A total of 300 clones for BRCA1 promoter and 780 clones for 8q24.21 were sequenced and analysis was performed using BISMA (Bisulfite Methylation Analysis - BPCD online tool). BRCA1 expression was also assessed by qRT-PCR in the DPSC sample. Results/Conclusions: Total BRCA1 promoter methylation was significantly higher (+1,4%) in the DPSC CL/P sample. Besides, BRCA1 promoter CpGs 1, 2 and 11 were the most hipermethylated in the CL/P sample (17.8%, 30.2% and 23.1%, respectively). As expected, BRCA1 expression was significantly reduced in comparison to controls (p=0.001). No evidence of differential methylation at the 8q24.21 cleft lip risk locus was found in the white blood cells DNA sample of CLP patients as compared to controls. Our results suggest that downregulation of BRCA1 in CL/P samples may be driven by increased BRCA1 promoter methylation and the causative factors in this hypermethylation should be next investigated. BRCA1 expression rescue by promoter demethylation is being conducted in DPSC CL/P samples. FAPESP/CNPq-MCT.

445F

Widespread changes in DNA methylation at CpG island shores and distal regulatory regions in response to a bacterial infection.
L.B. Barreiro[1,2], A. Pacis[2,3], L. Tailleaux[4], V. Yotova[2], J.C. Grenier[2], R. Pique-Regi[6], K.D. Hansen[7], Y. Gilad[5]. 1) Department of Paediatrics, Faculty of Medicine, University of Montréal, Montréal, Canada; 2) Ste-Justine Hospital Research Centre, Montreal, Canada; 3) Department of Bioinformatics, Faculty of Medicine, University of Montréal, Montréal, Canada; 4) Unité de GénétiqueMycobactérienne, Pasteur Institute, Paris, France; 5) Department of Human Genetics, University of Chicago, Chicago, USA; 6) Department of Molecular Medicine and Genetics, Wayne State University, Detroit, USA; 7) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA.

DNA methylation is an essential epigenetic modification for gene regulation, development, and disease processes. Recent studies have reported the dynamic nature of DNA methylation in response to different environmental conditions in mammalian cells. To better understand the role of DNA methylation in immune responses to infection, we collected single-base pair resolution methylation profiles of dendritic cells (DCs) before and after infection with *Mycobacterium tuberculosis* (MTB), the causative agent of tuberculosis (TB). We identified 1695 differentially methylated regions (DMRs) genomewide that span 300 hundred base pairs on average. Our findings show that changes in methylation do not occur in CpG islands and promoters but rather in low CpG-density regions, namely in CpG shores and distal regulatory regions. By using DNAse hypersensitivity sites (DHS) data we show that DMRs are enriched among DHS, which are strong predictors of open chromatin and active regulatory regions. We also observed a significant overlap of DMRs with enhancer regions as well as regions bound by transcription factors of key importance in immune responses, such as NF-κB. Finally, we show that DMRs are enriched within close proximity of genes that were differentially expressed genes after MTB infection. Specifically, we show that 25% of the genes that change expression after infection also have a DMR in close proximity. Genes close to DMRs that are hypomethylated after MTB infection tend to be up-regulated upon infection, whereas genes close to hyper-methylated DMRs tend to decrease their expression levels. These findings suggest a mechanistic link between changes in DNA methylation and changes in gene expression after MTB infection. Importantly, our results suggest that methylation levels might be more dynamic than previously thought, particularly in response to an infectious agent.

446T

Acceleration of age-associated methylation patterns in peripheral blood of HIV-1-infected adults.
R.M. Baxter[1], T.M. Rickabaugh[2], M. Sehl[2], J.S. Sinsheimer[1], O. Martinez-Masa[3], S. Horvath[1], E. Vilain[1], B.D. Jamison[2]. 1) Human Genetics, UCLA, Los Angeles, CA; 2) Medicine, division of Hem./Onc., UCLA, Los Angeles, CA; 3) Medicine, divsion of OB & GYN, UCLA, Los Angeles, CA.

Young HIV-1-infected adults, even when successfully treated with antiretroviral therapy, are prone to diseases more commonly associated with older uninfected adults. In addition, HIV-1-infection has been shown to have additive, detrimental effects, with aging on both cell number and telomere length within peripheral lymphocytes suggesting that HIV-1-infection accelerates aging in these cells by at least 10 years. However, it remains unknown whether aging and HIV-1-infection exert these effects through similar, or disparate, mechanisms. As we had previously identified methylation patterns associated with aging in uninfected adults, we tested whether HIV-1-infection would induce methylation changes associated with aging. Utilizing Infinium methylation arrays, we evaluated methylation levels at more than 450,000 CpG sites in DNA isolated from peripheral blood mononuclear cells (PBMC) obtained from the Multicenter AIDS Cohort Study (MACS), a longitudinal study of HIV-1 infection. Samples were from young (20–24 years) and older adults (48–56 years). Each group consisted of 12 HIV-1-infected individuals and 12 age-matched uninfected controls. We examined the relationship between differential methylation across age and HIV-1 infection and found that CpG sites with a positive age correlation were often also hypermethylated in HIV-1 infected individuals. A smaller number of sites correlated with decreased methylation in both aging and HIV-1-infeciton. Weighted gene co-methylation network analysis (WGCNA) identified 11 modules in the data based on methylation levels. CpG sites within module 7 were significantly correlated with both age and HIV-1 status. Using this module we demonstrated that HIV-1 infection accelerated age-related methylation by about 13 years. Examination of the genes related to the CpGs in this module showed enrichment for polycomb group targets, genes such as *SOX1, SOX8, PENK, MYOD1*, and *NPTX2* that are known to be involved in cell renewal and aging. These data demonstrate that HIV-1 infection is associated with methylation patterns that are similar to those associated with aging in the general population. The acceleration of aging due to HIV-1 infection by 13 years fits well with other studies on the effects of HIV-1 in the immune system. Taken together these data suggest that HIV-1-infection does accelerate some aspects of aging and that general aging and HIV-1 related aging work through at least some common mechanisms.

447F

Methylation QTLs often show opposite allelic directions when comparing different tissues. *M.J. Bonder[1], S. Kasela[2,3], K. Kirotar[3], M. Kals[2], M. Ivanov[4], A. Metspalu[2,3], M. Ingelman-Sundberg[4], C. Wijmenga[1], A. Zhernakova[1], L. Milani[2], L. Franke[1].* 1) Genetics, University Medical Center Groningen, Groningen, Groningen, Netherlands; 2) Estonian Genome Center, University of Tartu, Tartu, Estonia; 3) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 4) Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden.

Introduction It is clear that many disease-associated genetic variants affect gene expression (eQTL mapping). However, this effect is often tissue specific. We recently observed that within the same individuals eQTLs can show opposite allelic directions when comparing different tissues (Fu et al, PLoS Genetics 2012, Fairfax et al, Nature Genetics 2012). Here we investigated whether the same phenomena can be observed when investigating the effects of genetic variants on methylation (meQTL mapping)

Material and Methods We collected genotype, expression and methylation data from 94 Swedish liver samples. This was combined with a set of 84 Dutch individuals for whom we collected genotype, expression and methylation from liver, saturated adipose tissue, visceral adipose tissue and muscle. We performed eQTL and mQTL mapping in each of these tissues.

Results In the liver data we identified 2,920 significant meQTL probes and 443 significant eQTL probes after stringent multiple testing correction (estimated false discovery rate = 0). When investigating the other three tissues, we found that around 80% of the meQTLs and around 50% of the eQTLs that were identified in the non-liver datasets were also present in liver. We then assessed whether these overlapping QTL signals had consistent allelic directions. For the eQTLs we did not identify probes which showed an opposite effects in liver as compared to the other tissues. However, we observed 14 unique methylation probes, which gave significant opposite allelic effects in liver as compared to the three other tissue types. As we did not observe any opposite allelic effects when comparing the meQTLs detected in the individual Swedish and Dutch liver samples, we believe the 14 probes with opposite allelic effects reflect true positive results.

Conclusion In this study we found that around 80% of the meQTLs are shared between different tissues, but we also identified a few meQTLs that showed completely opposite allelic effects when comparing different tissues. We thus conclude that careful selection of the tissue of interest is crucial when it comes to interpretation of both methylation and expression QTL results.

448T

Predicting Prostate Cancer Progression through Gene Network Analysis of Methylation Data. *L. Briollais, K. Kron, B. Bapat, H. Ozcelik.* Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, M5T 1L9, Canada.

Promoter and 5' end methylation regulation of tumour suppressor genes is a common feature of many cancers. Such occurrences often lead to the silencing of these key genes and thus they may contribute to the development of cancer, including prostate cancer. In order to identify methylation changes in prostate cancer progression, we performed a genome-wide analysis of DNA methylation using Agilent human CpG island arrays available on 20 patients (10 with Gleason score 6 and 10 with Gleason score 8). Our first set of analyses identified a large number of potential epigenetic biomarkers of prostate cancer progression, including various genes belonging to the Homeobox family. The second set of analyses aimed at constructing a gene network around the Homeobox genes and use this information as a predictive tool for prostate cancer progression. The different models found in the second stage of our analysis showed an excellent predictive ability and these models were further validated in an independent data set of methylation data as well as in a gene expression data set. We finally discuss various statistical approaches for gene network analysis including graphical models. Our conclusion is that gene network analysis can provide a very sensible and comprehensive framework for understanding the genetic basis of complex human diseases and for identifying individuals the more susceptible to disease progression.

449F

A pilot study testing DNA methylation profiles in Samoan obese and lean young adult males. *O.D. Buhule[1], N.L. Hawley[5], M. Medvedovic[3], R.L. Minster[2], G. Sun[3], H. Cheng[4], S. Viali[6], R. Deka[3], D.E. Weeks[1,2], S.T. McGarvey[7].* 1) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio 45367, USA; 4) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH, USA; 5) Weight Control and Diabetes Research Center, The Miriam Hospital, Providence, RI, USA & The Alpert Medical School, Brown University, Providence, RI, USA; 6) Medical Specialist Clinic and National Health Services, Government of Samoa, Apia, Samoa; 7) International Health Institute and Department of Epidemiology, Brown University School of Public Health, Providence, RI 02912, USA.

Background and Objective: Methylation levels, which influence gene expression, can be influenced by environment and life style. Obesity, as a product of both nutritional environment and life style, could be related to methylation levels. Here we present preliminary findings from a pilot study examining DNA methylation patterns across the genome in young obese and lean male Samoans to identify epigenetic loci associated with obesity. Methods: DNA was extracted from whole peripheral blood from 46 obese (BMI >=32 & Abdominal Circumference >=92.5cm) and 46 lean (BMI <26 & Abdominal Circumference <92.5cm) in the age range 25–40 years. The blood was drawn while fasting the morning after the anthropometric measures were taken. We used methylumi and wateRmelon packages in R available through Bioconductor to analyze data from the 450K Human Methylation Array. The M-values (log2 of the methylation levels) were calculated, as these are preferred to beta values. The M-values were adjusted for chip and row batch effects using the ComBat package. Probe filtering was done to exclude any outlying samples and sites before computing the M-values. To identify differentially methylated positions between obese and lean males at each locus or probe, we used a moderated-t-test that adjusted for multiple testing using the Benjamini-Hochberg method (Limma package). Results: A total of 469,507 CpG loci across the genome in 92 individuals were investigated. A total of 91,805 probes emerged as differentially methylated in obese males when compared to lean males at an q-value threshold of 0.05; 3,454 of these had p-values less than the Bonferroni threshold of 1.06 × 10–7. The 3,454 top hits represent approximately 1,345 different regions, several of which have previously been associated with obesity. An additional data set of 92 individuals is being generated. Support: This work was supported by NIH R01 HL093093 (PI: Stephen McGarvey).

450T

DNA Methylation Alterations in CHARGE Patients with Heterozygous *CHD7* Mutations. *D.T. Butcher[1], D. Grafodatskaya[1], D.W.X. Wei[1], W. Reardon[2], B. Gilbert-Dussardier[3], A. Verloes[4], F. Bilan[5], B. Papsin[6,7], R. Badilla-Porras[8], R. Mendoza-Londono[8], R. Weksberg[1,8,9].* 1) Genetics & Genome Biology, Sickkids Research Institute, Toronto, Ontario, Canada; 2) National Centre for Medical Genetics, Our Lady's Children's Hospital, Dublin, Ireland; 3) Service de Génétique, Centre de Référence Anomalies du Développement de l'Ouest, CHU Poitiers, France; 4) AP-HP, Groupe Hospitalier Pitié-Salpêtrière, UF de Génétique Clinique, Paris, France; 5) Institut de Physiologie et Biologie Cellulaires, Centre National de la Recherche Scientifique Unité Mixte de Recherche, Université de Poitiers, CHU Poitiers, France; 6) Otolaryngology, The Hospital for Sick Children, Toronto, Canada; 7) Department of Otolaryngology, The University of Toronto, Toronto, Canada; 8) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 9) Department of Paediatrics, The University of Toronto, Toronto, Canada.

CHARGE syndrome (CHARGE) is a rare autosomal dominant genetic disorder, with an incidence of 1 in 8500–10000 births. Clinical diagnosis for CHARGE is based on non-random associations of the following congenital abnormalities: Coloboma of the eye, Heart defects, Atresia of the choanae, Retarded growth and development, Genital abnormalities, Ear abnormalities/deafness/vestibular disorder. In the majority of cases, CHARGE is the result of haploinsufficiency due to a nonsense, missense, or deletion in the gene encoding Chromodomain Helicase DNA-binding protein (CHD7). Targeted mutational studies in Drosophila (kismet) and mouse (Chd7) have found phenotypes similar to those found in human. In Drosophila reduced expression of kismet/CHD7 results in deficits in axonal pruning, guidance and extension as well as defects in memory and motor function. Normal mammalian growth and development depend on the correct epigenetic programming of the genome. Epigenetic patterns evolve across development utilizing mechanisms such as DNA methylation and covalent modifications of histone proteins. In Drosophila, kismet the ortholog of human CHD7 has been demonstrated to regulate the repressive histone H3 methylation mark of lysine 27. In human cell lines, CHD7 has been shown to bind to chromatin regions that are active as demonstrated by histone H3 lysine 4 methylation and DNAse1 hypersensitivity of these binding sites. CHD7 also interacts with RNA polymerase II, forming complexes that alter chromatin structure to facilitate access for transcriptional machinery. These epigenetic modifications of histone H3 are tightly linked to DNA methylation patterns. We hypothesized that specific DNA methylation alterations occur as a result of the heterozygous CHD7 mutations and could reveal critical downstream targets associated with CHARGE clinical features. We have analyzed cases with CHD7 mutations comparing their methylation alterations to age and sex-matched controls using the Illumina Inifinium Methylation450 BeadChip array. Data were analyzed using the IMA package in R and Genome Studio software from Illumina. We identified both gain and loss of methylation in genes that play a role in growth and neurodevelopment. The identification of these epigenetic modifications could lead to an improved understanding of the pathophysiology of CHARGE and the type of chromatin regions to which CHD proteins are recruited.

451F

Contribution of DNA methylation to gene expression varies by tissue and age. *C. Chen[1,2], C. Zhang[1,2], L. Cheng[1,2], J. Badner[4], E. Gershon[4], J. Sweeney[5], J. Reilly[6], J. Bishop[1,3], C. Liu[1,2].* 1) Psychiatry, University of Illinois at Chicago, Chicago, IL; 2) Institute of Human Genetics, University of Illinois at Chicago, Chicago, IL; 3) Department of Pharmacy, University of Illinois at Chicago, Chicago, IL; 4) Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, IL; 5) Department of Psychiatry, University of Texas Southwestern, Medical Cetner, Dallas, TX; 6) Psychiatry, Northwestern University, Chicago, IL.

DNA methylation, as an epigenetic mark on CpG dinucleotides, was considered to simply block the binding of transcription factors in promoter region and repress gene expression. However, recent studies showed that DNA methylation functions vary with genomic context and tissues. We systematically evaluated correlations between expression of individual genes and DNA methylation at cis-regions in brain cerebellum (CB), prefrontal cortex (PFC), and blood and lymphatic endothelial cells (LEC). We found that less than 5% of the genes were significantly correlated with DNA methylation level within one specific tissue. Many other genes have their expression correlated with methylation by tissue types or across life span while their expressions have little variation within one tissue of limited age range. Positive correlations were observed in all tests. CpG sites from correlated pairs were more dynamic and tissue-specific in CpG-poor regions. Genes involved in age-dependent methylation regulation were enriched with brain development functions. CpG sites in promoter CpG Island (CGI) were more likely to be consistently unmethylated, and corresponding regulated genes were enriched for acetylation functions. These diverse correlations suggested complexity of the roles of DNA methylation in regulating gene expression. DNA methylation of most CpG sites may have been used to define spatiotemporal gene expression patterns, i.e., tissue and age variations, while much fewer CpG sites mediate variations of gene expression within tissue in human population for individual differences.

452T

Effective adjustment of differential cell populations in epigenome-wide association studies. *J. Chen[1], J. Huang[2], L. Liang[1,2], X. Lin[1].* 1) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA.

Epigenome-wide association studies, which investigate an association between epigenetic variation and phenotypic variation, have attracted increasing attention recently, especially given the availability of the Infinium HumanMethylation450 BeadChip. Blood samples are routinely collected for measuring DNA methylation. However, blood is a mixture of different cell types, each with a unique methylation pattern. Differential cell populations between case and control samples can potentially confound the associations of interest. Several approaches have been proposed to estimate cell proportions and include them as covariates in a regression model. The effectiveness of this strategy depends on the quality and completeness of a reference panel of purified cell types, which are not always available. We propose to adjust for the confounding of cell mixtures using PCA (Principle Component Analysis), SVA (Surrogate Variable Analysis) and ISVA (Independent Surrogate Variable Analysis). Our proposed method is easily to be carried out and does not require a reference panel. By including the estimated principle components (PCs) or surrogate variables (SVs) in regression models, these methods correct the inflation of QQ-plots under the null due to cell mixture confounding and improve power when true signal presents. We demonstrate by simulation that the proposed methods can capture the cell proportions as low as 3% in presence of batch effects and population stratification, and ISVA is more efficient in capturing the proportions of rare cell types, while all methods perform similarly for dominant cell types. We further show that PCA but not SVA/ISVA breaks down when a large number of true signals are mixed with false signals due to confounding. We apply the methods to several 450K methylation data sets and find that the first few PC/SVs are sufficient to capture the proportions of dominant cell types - neutrophil and lymphocyte, while rare cell types such as eosinophil require more PC/SVs. Our results show that SVA/ISVA provide a convenient and effective approach to adjust for differential cell populations as well as other batch effects and population stratification in epigenome-wide association studies.

453F

A specific DNA methylation signature associated with NSD1+/- mutations in Sotos syndrome reveals a significant genome-wide loss of DNA methylation (DNAm) targeting CGs in regulatory regions of key developmental genes. *S. Choufani[1], C. Cytrynbaum[2], A.L. Turinsky[3, 4], Y.A. Chen[1], D. Grafodatskaya[1], J. Xiang[1], M. Feigenberg[2], B.Y.H. Chung[5], D.J. Stavropoulos[6], R. Mendoza-Londono[2], D. Chitayat[2], W.T. Gibson[7], M. Reardon[8], M. Brudno[1,4,9], R. Weksberg[1,2].* 1) Program in Genetics and Genome Biology, Hosp Sick Children, Toronto ON, Canada; 2) Div Clin & Metabolic Gen, Hosp Sick Children, Toronto, ON, Canada; 3) Molecular Structure & Function, Hosp Sick Children, Toronto, ON, Canada; 4) Centre for Computational Medicine, Hosp Sick Children, Toronto, ON, Canada; 5) Dept of Paediatrics & Adolescent Med, Li Ka Shing Faculty of Medicine, Hong Kong; 6) Paediatric Laboratory Medicine, Hosp Sick Children, Toronto, ON, Canada; 7) Dept. of Medical Genetics, UBC, Child and Family Research Institute, Vancouver, BC, Canada; 8) Our Lady's Hospital for Sick Children, Crumlin, Dublin 12, Ireland; 9) Department of Computer Science and Donnelly Centre, University of Toronto, Toronto, ON, Canada.

Sotos syndrome (SS) is characterized by somatic overgrowth and intellectual disability. Most SS cases (>75%) have mutations in NSD1 (nuclear receptor-binding SET domain protein 1). NSD1 binds near promoter elements and regulates transcription initiation and elongation via interactions with H3-K36Me and RNA polymerase II. To determine if NSD1 mutations impact stable epigenetic marks such as DNA methylation (DNAm), we compared DNAm in peripheral blood DNA from SS cases with NSD1 mutations (NSD1+/–; n=20) to controls (n=30) using the Illumina Infinium450methylation BeadChip. Differential DNAm analysis using non-parametric statistics (with correction for multiple testing) coupled with permutation analyses identified a surprisingly high number (n=2157) of differentially methylated (DM) CG sites (with >20% difference in DNAm) between SS and controls. These sites were distributed across the genome; 95% demonstrated loss of DNAm. Using unsupervised hierarchical clustering of the 2157 DM CG sites, all SS cases with NSD1 +/– clustered as a distinct group separate from controls. Moreover, DNAm at these sites clearly distinguished SS (NSD1+/–) from Weaver syndrome (EZH2+/-, n=5), another overgrowth syndrome which has considerable phenotypic overlap with SS. These results suggest that these DM CG sites constitute a DNAm signature that is specific for NSD1+/–. Also, the DNAm signature was successfully used to reclassify NSD1 variants of unknown significance (VUS) in six cases of SS into functionally damaging (n=1) and non-pathogenic (n=5) variants. The majority of these DM CGs mapped to enhancers and CpG island shores. Analysis of ChIP-seq data showed that NSD1+/– specific CG sites are associated with reduced H3K36me3 marks in both normal blood and embryonic stem cells. Also, Ingenuity analysis showed enrichment in neural and cellular development pathways (p<0.001). We then searched for binding motifs enriched in these NSD1+/– DNAm targets using MEME and JASPAR CORE database; SP1 was the most enriched with binding sites in 41% of the targets (NCOR= 0.62). This is the first report of an NSD1+/– specific DNAm signature in SS and that loss-of-function mutations in NSD1 can deregulate the intricate transcriptional balance of key developmental genes. Further elucidation of this signature will significantly impact our understanding of the molecular pathophysiology of SS and identify the specific molecular targets for NSD1 that govern its action in early development.

454T

Comparison of methylation profiles in human blood and lung tissue identifies tissue specific CpG methylation sites. *D. Daley[1], K. Ushey[1], L. Akhabir[1], A. Saferali[1], S.M. Mah[2], A. Sandford[1], M.S. Kobor[2], P. Parè[1].* 1) James Hogg Research Center Department of Medicine University of British Columbia Vancouver, BC, Canada V6G 1Y6; 2) Department of Medical Genetics University of British Columbia Vancouver, BC, Canada V5Z 4H4.

We hypothesize that differences in methylation patterns between individuals may contribute to the etiology of asthma, COPD and other lung related traits. As the majority of genetic studies collect and maintain blood samples from study participants, blood is likely to be the most widely available resource for studies looking to investigate the relationship between phenotype and methylation status. Given our interest in pulmonary phenotypes, and the knowledge that methylation patterns are tissue specific we wanted to identify CpG sites that demonstrate tissue specific methylation (TSM) patterns between blood and lung tissues. **Methods:** To evaluate TSM we used 36 paired samples (blood and lung tissue from the same individual) and 22 lung only samples. Methylation of over 450 thousand CpG sites throughout the genome was evaluated using the Illumina Infinium HumanMethylation450K bead chip array. Methylation levels were assessed using beta values which range from 0 (no methylation) to 1 (complete methylation). Principal components were used to identify 3 samples which clustered differently, all 3 samples were noted to have low and/or poor quality DNA and were removed from subsequent analyses. Correlations between beta values were examined using principal components, heatmaps, and a mixture model of three distributions (unmethylated, variable methylation, complete methylation). **Results:** We identified 49,376 CpG sites with apparent tissue specific methylation. There are 31 CpG sites with extreme differences in methylation, at these sites there is complete methylation in lung tissue but no methylation in blood tissue. We further identified 4,541 CpG sites that are unmethylated in lung tissue, and demonstrate intermediate (variable) methylation patterns in blood tissue, 34,974 CpG sites with complete methylation in blood and intermediate levels of methylation in lung tissue and finally 9,830 sites where there is complete methylation in lung tissue and intermediate levels in blood tissue. Next we evaluated specific genes which have been associated in genome-wide association studies with lung related diseases. We noted good correlation between blood and lung methylation values at *ORMDL3* asthma locus but poor concordance at the Thymic Stromal Lymphopoietin (*TSLP*) locus. **Conclusion:** We have identified 49,376 CpG sites with tissue specific methyation, including CpG sites in the asthma gene *TSLP*.

455F

Hippocampal gene expression, miRNA expression and DNA methylation changes in a mouse model of Fetal Alcohol Spectrum Disorder (FASD). *E.J. Diehl, Bl. Laufer, SM. Singh.* Biology, Western University, London, Ontario, Canada.

Alcohol abuse during pregnancy can lead to a range of neurological abnormities termed Fetal Alcohol Spectrum Disorder (FASD). The mechanisms by which alcohol (ethanol) induces FASD are poorly understood; however, disruption of the regulation of genetic programs in the brain may be involved. We have previously developed a model to generate FASD-like mice showing characteristic learning and memory impairment. Furthermore, as young adults these mice show persistent changes in whole-brain gene expression. However, the mechanisms maintaining these long-term alterations remain unknown. Epigenetic phenomena are increasingly implicated in FASD; however, few studies have integrated multiple genome-wide epigenetic analyses in one study. Here, we examined global difference gene expression, DNA methylation, microRNA (miRNA), and histone methylation in the hippocampi of ethanol-exposed mice. We hypothesized that altered epigenetic mechanisms induce gene expression changes in the hippocampus which may underlie learning and memory impairment. Mouse pups were injected with saline or ethanol on postnatal days 4 and 7, the period equivalent to human trimester three. At 70 days of age, hippocami were isolated and used for gene and miRNA expression microarray, and methylated DNA immunoprecipitation & microarray (MeDIP-chip). ChIP-chip was used to assess changes in histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 27 trimethylation (H3K27me3). Fifty nine genes were differentially expressed at p<0.05 and fold cut off 1.2. Ingenuity Pathway Analysis (IPA) identified the top affected biological pathway as ‘Free Radical Scavenging, Gene Expression, Dermatological Diseases and Conditions’ with 13/40 molecules affected. Five genes had gene expression and DNA methylation changes; one such gene was Mafg, downregulated by 1.21 fold. MAFG is a transcription factor involved in oxidative stress response. Two miRNAs which target Mafg, miR-130a and miR-200b were upregulated. Mafg may be downregulation as a result of miRNA and DNA methylation changes induced by ethanol. Ethanol is known to induce oxidative stress in the developing brain through a variety a mechanisms including reduction of antioxidant levels and increasing reactive oxygen species (ROS) production. Alteration in the epigenetic regulation of Mafg and other oxidative stress response genes may represent a novel point of interface between the epigenetic and oxidative stress mechanisms of FASD generation.

456T

Genome-wide DNA methylation profiles in fruit flies and the effect of huntingtin knockout. *S. Erdin[1], K. Dietz[1,2], A. Ragavendran[1], M.E. Talkowski[1,3,5], J.A. Walker[1,3], J.F. Gusella[1,4,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Louisiana State University Health Sciences Center, Shreveport, LA; 3) Department of Neurology, Harvard Medical School, Boston, MA; 4) Department of Genetics, Harvard Medical School, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.

The definitive presence of DNA methylation in Drosophila melanogaster has not yet been established and, if present, the epigenetic role at specific loci will need to be determined. We previously detected different DNA methylation levels in male and female wild type fruit flies using immunological methods with antibodies specific for 5'-Methylcystosine (5mC) and 5'-Hydroxymethlcystosine, suggesting genome-wide methylation is present. In this study, we hypothesized that (a) DNA methylation is present in the male and female adult flies, (b) the pattern of DNA methylation is gender-specific, and (c) the loss of huntingtin protein results in genome-wide differential methylation compared to wild-type flies. We conducted a methylated DNA immunoprecipitation (meDIP) sequencing experiment that profiles methylation patterns on a genome-wide scale based on enrichment using antibodies specific for 5mC. Extracting DNA from the adult fly's brain, we prepared eight sequencing libraries with 47.3 million 50 bp paired-end reads on average for each of wild-type and huntingtin-null male and female flies with 5mC and 5hmC specific antibodies and their counterparts with no antibodies for comparison. For analysis, we followed a computational protocol involving: quality filtering of reads by Sickle, sequence alignment to the fruit fly reference genome by BWA and subsequent filtering of alignments by SamTools. To identify peaks and differentially methylated regions and their annotation, we used standard tools specifically designed for meDIP-seq analysis, Medips and peak callers, MACS and Homer. Our preliminary results based on differentially methylated regions identified by Medips relying on edgeR's statistical analysis (FDR < 0.001) and methylation peaks identified by MACS (p-value < 1e−5) confirm the presence of global DNA methylation in the fly, reveals genome-wide gender-specific differences, and suggest differential methylation associated with loss of huntingtin. Replication of these findings and further downstream analyses are ongoing, but these data confirm the presence of methylated DNA sites in Drosophila melanogaster and suggest a significant role of huntingtin in epigenetic modification.

457F

A novel method for identification and quantification of consistently differentially methylated genomic regions. *C. Fann[1], C.L. Hsiao[1], C.J. Chang[2].* 1) Epidemiology & Genetics, Inst Biomed Sci, Acad Sinica, Taipei, Taiwan; 2) Graduate Institute of Clinical Medical Science, Chang Gung University, Taoyuan, Taiwan.

Advances in biotechnology have resulted in large-scale studies of DNA methylation. A differentially methylated region (DMR) is a genomic region with multiple adjacent CpG sites that exhibit different methylation statuses among multiple samples. Many so-called 'supervised' methods have been established to identify DMRs between two or more comparison groups. Methods for the identification of DMRs without reference to phenotypic information are, however, less well studied. An alternative 'unsupervised' approach was proposed, in which DMRs in studied samples were identified with consideration of nature dependence structure of methylation measurements between neighboring probes from tiling arrays. Through simulation study, we investigated effects of dependencies between neighboring probes on determining DMRs where a lot of spurious signals would be produced if the methylation data were analyzed independently of the probe. In contrast, our newly proposed method could successfully correct for this effect with a well-controlled type I error and a comparable statistical power. Identification of DMRs in a population of samples is vital for understanding methylation variation within a dataset. By applying to two real datasets, we demonstrate that our method provides a more global picture of methylation variation either between groups or between individuals in a single screen.

458T

Epigenetic changes in relation to asbestos exposure in malignant pleural mesothelioma. *G. Fiorito[1,2], S. Guarrera[1], E. Casalone[1,2], M. Betti[3], E. Aldieri[4], D. Ferrante[5], C. Di Gaetano[1,2], F. Rosa[1], A. Russo[1,2], S. Tunesi[5], M. Padoan[5], A. Aspesi[3], C. Casadio[6], F. Ardissone[7], E. Ruffini[8], P.G. Betta[9], R. Libener[9], R. Guaschino[10], E. Piccolini[11], D. Mirabelli[12,13], C. Magnani[5,13], I. Dianzani[3,13], G. Matullo[1,2].* 1) Human Genetics Foundation, HuGeF, I-10126 Turin, Italy; 2) Department of Medical Sciences, University of Turin, I-10100, Turin, Italy; 3) Laboratory of Genetic Pathology, Department Health Sciences, University of Piemonte Orientale, I-28100, Novara, Italy; 4) Department of Oncology, University of Turin, I-10126, Turin, Italy; 5) CPO-Piemonte and Unit of Medical Statistics and Epidemiology, Department Translational Medicine, University of Piemonte Orientale, I-28100, Novara, Italy; 6) Thoracic Surgery Unit, University of Piemonte Orientale, I-28100, Novara, Italy; 7) Chest Surgery, Department of Clinical and Biological Sciences, University of Turin, I-10043, Orbassano, Italy; 8) Thoracic Surgery Unit, University of Turin, I-10126, Turin, Italy; 9) Pathology Unit, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, I-15121, Alessandria, Italy; 10) Transfusion Centre, Azienda Ospedaliera Nazionale SS, Antonio e Biagio e Cesare Arrigo, I-15121, Alessandria, Italy; 11) Pneumology Unit, Santo Spirito Hospital, I-15033, Casale Monferrato, Italy; 12) Unit of Cancer Epidemiology, CPO-Piemonte and University of Turin, I-10126, Turin, Italy; 13) Interdepartmental Center for Studies on Asbestos and other Toxic Particulates 'G. Scansetti', University of Turin, I-10125,Turin, Italy.

Malignant pleural mesothelioma (MPM) is a rare and aggressive tumor strongly associated with asbestos exposure. Only 5–17% of individuals exposed to asbestos develop MPM, suggesting the involvement of other environmental, genetic and epigenetic risk factors. DNA methylation is an important mechanism of gene silencing in human malignancies. The relationship between aberrant DNA methylation and inflammation has been documented in many types of cancers, including MPM. Asbestos exposure may contribute to MPM onset through this relationship. We conducted an epigenome-wide scan to identify differentially methylated regions (DMR) in 40 MPM cases versus 40 controls, and in asbestos high-exposed versus low-exposed subjects. Methylation status was measured for about 470K CpG sites in DNA from whole blood, using the HumanMethylation450 BeadChip (Illumina, S. Diego, CA). Logistic regression analysis after adjustment for age, gender and center of recruitment showed no significant association with MPM of any single CpG methylation profile. However, a regional analysis showed multiple significant signals in several genomic regions. In particular, a significant decreased methylation (P<10−7) has been identified in asbestos exposed subjects compared with non-exposed for BLCAP (bladder cancer associated protein), a tumor suppressor gene, and NNAT (neuronatin), which was reported as overexpressed in non-small cell lung carcinoma, both of them located on opposite strands in the same locus on Chr20. No significant difference has been found between the overall cases and controls at the regional analysis. To verify if blood DNA could indeed be a proxy for MPM tissue, we compared the methylation profiles of DNA from whole blood and from corresponding MPM cell lines of 4 MPM cases. The methylation profiles at the BLCAP/NNAT locus proved to be comparable in DNA from blood or MPM cells. These results suggest that different DNA methylation profiles can be related to asbestos exposure induced tumorigenesis of MPM and that epigenetic changes are detectable also in blood DNA.

459F

Genetic Ancestry Explains Differences in Global and Local Methylation Patterns in the GALA II Study.
J.M. Galanter[1], C.R. Gignoux[1], S.S. Oh[1], D.G. Torgerson[1], C. Eng[1], S. Huntsman[1], L. Roth[1], D. Hu[1], S. Sen[1], M. Pino-Yanes[1], E. Nguyen[1], P. Avila[2], H.J. Farber[3], A. Davis[4], E. Birgino-Buenaventura[5], M.A. Lenoir[6], K. Meade[4], D. Serebrisky[7], S. Thyne[8], W. Rodriguez-Cintrón[9], R. Kumar[10], J.R. Rodriguez-Santana[11], E.G. Burchard[1]. 1) University of California, San Francisco San Francisco, CA; 2) Feinberg School of Medicine, Northwestern University, Chicago, IL; 3) Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 4) Children's Hospital and Research Center Oakland, Oakland, CA; 5) Kaiser Permanente-Vallejo Medical Center, Vallejo, CA; 6) Bay Area Pediatrics, Oakland, CA; 7) Jacobi Medical Center, Bronx, NY; 8) San Francisco General Hospital, San Francisco, CA; 9) Veterans Carribean Health System, San Juan, PR; 10) The Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 11) Centro de Neumologia Pediatrica, San Juan, PR.

Epigenetic modification of the genome through methylation plays a key role in the regulation of diverse cellular processes. Changes in DNA methylation patterns have been associated with many complex diseases. Recent studies have found significant differences in the methylation patterns of peripheral blood between African Americans and non-Hispanic Whites. In this study, we leveraged estimates of genomic ancestry in 575 Latino children of multiple Latino ethnicities (Puerto Rican, Mexican, and other) enrolled in the GALA II study of childhood asthma to determine whether differences in global and local methylation patterns between ethnic groups could be explained by ancestry. We measured DNA methylation at ~450,000 markers using the Illumina Infinium HumanMethylation450 BeadChip. We used multidimensional scaling to determine global methylation patterns. Ethnicity was significantly associated with the sixth principal coordinate ($p < 2 \times 10{-}16$ for the overall effect of ethnicity on PC6 and for the comparison of Puerto Ricans to Mexicans). Native American ancestry, when added to the model, was also highly associated with the principal coordinate, and its inclusion in the model eliminated the significance of the association between ethnicity and the methylation pattern measured by PC6. We then performed an epigenome-wide association study between ethnicity and methylation at each site. There was a significant association between ethnicity and local methylation patterns at 1356 sites at a Bonferoni corrected significance level ($1.4 \times 10{-}7$). We performed a mediation analysis to determine the extent to which genomic ancestry mediated the effect of ethnicity on local methylation. Of the 316 methylation sites with a p-value less than $1 \times 10{-}10$, 40 (13%) were significantly mediated by Native American ancestry. The median proportion of the effect of ethnicity on methylation mediated by Native American ancestry was 66% across all sites. An epigenome-wide association study between Native American Ancestry and methylation found 309 sites with a Bonferoni corrected significance level of $1.4 \times 10{-}7$ or below. These findings have broad implications for the study of methylation patterns across populations and for disease association studies. There significant differences in methylation patterns between ethnic groups that are due to ancestry differences in those groups. These differences should be accounted for in performing epigenome wide disease association studies.

460T

A fast and simple method for whole genome bisulfite sequencing library preparation from ultra-low DNA input: Pico-MethylSeq library preparation kit.
K. Giang, T. Chung, X. Sun, X. Jia. Zymo Research Corporation, Irvine, CA.

The distribution of 5-methylcytosine (5-mC) in DNA within the eukaryotic genome is known to greatly affect gene regulation and is currently a major topic of research. Studies on DNA methylation have been aided by advancements in bisulfite conversion and next-gen sequencing technologies which, when coupled, provide single-base resolution of 5-mC in the whole genome. Many whole-genome bisulfite sequencing (WGBS) library preparation protocols designed to analyze 5-mC distribution in the whole genome employ bisulfite to convert unmethylated cytosine bases to uracil after the library preparation and while these protocols produce reliable results, degradation of DNA is inherent to bisulfite conversion. As such, a large proportion of the adapterized library is fragmented and can no longer be amplified, which requires these protocols to call for large amounts of starting input DNA that is often times impossible to obtain. By rearranging the order of library preparation and bisulfite conversion, we developed a streamlined protocol that reveals whole-genome methylation patterns at single-base resolution. The work-flow leads with the degradation inherent to bisulfite conversion to randomly fragment the DNA prior to the library preparation, which allows the protocol to accommodate for pico-gram quantities of starting input, making it ideal for analysis in precious and limited samples. Comparisons of sequencing data from this WGBS library preparation method with the established Reduced Representation Bisulfite Sequencing method using human DNA showed a correlation coefficient of 0.95 for CpG sites with more than 10X coverage. With slight modifications, this protocol is versatile in its ability to prepare libraries for ChIP-seq and RNA-seq.

461F

Evaluation of Illumina 450K Methylation Chip using Technical Replicates.
W. Guan[1], M. Bose[1], J.S. Pankow[1], E.W. Demerath[1], J. Bressler[2], M. Fornage[2], M. Grove[2], T. Mosley[3], C. Hicks[3], K. North[4], E. Boerwinkle[2]. 1) Univ of Minnesota, Minneapolis, MN; 2) Univ of Texas Health Sciences Ctr at Houston, Houston, TX; 3) Univ of Mississippi, Jackson, MS; 4) Univ of North Carolina-Chapel Hill, Chapel Hill, NC.

DNA methylation is a widely studied epigenetic mechanism that alterations in methylation patterns may be involved in the development of common diseases.. As part of the Atherosclerosis Risk in Communities (ARIC) Study, the Illumina Infinium HumanMethylation450 (HM450) BeadChip was used to measure DNA methylation in peripheral blood obtained from ~3000 African American participants. Over 480,000 cytosine-guanine (CpG) dinucleotide sites were surveyed on the HM450 BeadChip. For most of the CpG sites (>90%), we observed significant differences in methylation levels between samples on different plates and chips, largely due to technical error. To evaluate the impact of technical errors, we used 137 technical replicates included in the study. For each CpG site, we calculated an intraclass correlation coefficient (ICC) to compare variation of methylation levels within- and between-replicate pairs, ranging between 0 and 1. The median of observed ICC values is .30, with interquartile range (IQR) of .52 (.11-.63). We observed a relatively large proportion (7.7%) of assayed CpG sites with ICC of 0. The distribution of ICC was modeled as a mixture of censored or truncated normal and normal distributions, which was fitted using an EM algorithm. The CpG sites were clustered into zero-, low-, and high-reproducibility groups, according to the calculated posterior probabilities. The proportions of sites in the three groups are approximately 8%, 50%, and 42%, respectively. We then demonstrated the performance of proposed clustering on CpG sites on their associations with smoking status of samples in the ARIC study. Our results suggested that association at CpG sites with low reproducibility may be under-estimated given the impact of technical variation and needs to be interpreted carefully.

462T

Edematous severe childhood malnutrition is associated with widespread DNA hypomethylation.
N.A. Hanchard[1,2], S. Howell[3], K. Marshall[3], X. Wang[1], T.E. Forrester[3], M.E. Reid[3], J.W. Belmont[1,2], C.A. McKenzie[3]. 1) USDA/ARS/Children's Nutrition Research Center, Baylor College of Medicine, 1100 Bates Ave, Houston, TX, 77030, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030, USA; 3) Tropical Metabolism Research Unit, Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston 5, Jamaica.

Severe childhood malnutrition (SCM) is a major global health problem that contributes to more than two million childhood deaths worldwide each year. SCM occurs in two clinically distinct forms — the more severe edematous SCM (ESCM), and the milder non-edematous SCM (NESCM). The etiology of this clinical dichotomy is unclear; however, previous studies have observed differences between acutely-ill ESCM and NESCM patients in the flux of methyl-groups through the 1-carbon conversion of methionine to homocysteine. Because this reaction contributes significantly to the maintenance of CpG dinucleotide methylation during mitosis, we hypothesized that in high-turnover tissues, DNA methylation in ESCM patients might differ relative to NESCM patients. We used the Illumina Infinium HumanMethylation450 BeadChip array to evaluate methylation at ~485,000 genome-wide CpG sites in buccal DNA samples collected from Jamaican children recruited shortly after recovery from acute SCM, and Jamaican adults recruited 18 or more years after having SCM. After data normalization and quality control, linear regression models were used to determine CpG sites with differential methylation between ESCM and NESCM patients. Immediately post-SCM recovery, we found that samples from ESCM children (N=9) showed significant differential hypomethylation relative to NESCM (N=8) at 21,000 CpG sites genome-wide (false discovery rate, FDR, p<0.01), involving 1,461 genes with more than two significant probes. In pathway analyses, the top 1,000 differentially hypomethylated probes were 4-fold enriched for genes involved in GTPase regulator activity (FDR, $p<1 \times 10{-}3$). Conversely, there were no statistically significant differences in methylation between ESCM (N=10) and NESCM (N=14) individuals sampled as adults. Our results provide evidence for genome-wide differential DNA methylation in ESCM that is directly related to the nutritional insult, and provide a basis for larger studies to replicate our observations. Further, these results suggest that methylation may be a useful intermediate phenotype with which to probe possible genetic trait loci that could provide insight into why some children develop ESCM while others develop NESCM, despite similar environmental and dietary exposures.

463F

Epigenetic Alterations and an Increased Frequency of Micronuclei in Women With Fibromyalgia.
C. Jackson-Cook[1,2], V. Menzies[3], K.J. Archer[4], J. Brumelle[1], K.H. Jones[1,5], Q. Zhou[4], G. Gao[4], T.P. York[2], D.E. Lyon[3]. 1) Dept Pathology, VA Commonwealth Univ, Richmond, VA; 2) Dept Human & Molecular Genetics, Virginia Commonwealth Univ, Richmond, VA; 3) School of Nursing, VA Commonwealth Univ, Richmond, VA; 4) Dept Biostatistics, Va Commonwealth Univ, Richmond, Va; 5) NeoDiagnostix, Inc., Rockville, MD.

Fibromyalgia (FM) is a debilitating condition associated with multiple symptoms, including chronic, widespread pain; dyscognition; fatigue; sleep disturbances; and depression. These symptoms often culminate in decreased workplace productivity, increased health/disability costs, and compromises in the overall quality of life for people with FM. While the adverse health and economic impacts of FM have been well recognized, its etiology remains enigmatic. We hypothesized that FM is associated with the development of acquired somatic epigenetic and/or chromosomal alterations. To test this hypothesis we compared the frequency of spontaneously occurring micronuclei (MN) and genome-wide methylation patterns (Illumina 450K HumanMethylation Chip) in leukocytes of women with FM (N=10) to those seen in comparably aged healthy controls (N=42 [MN]; N=8 [methylation]). A 3.26-fold significantly increased mean frequency of MN ($\psi 2$ =45.552; df= 1; p=1.49×10−11) was observed in the patients with FM [mean=51.4, sd= 21.9] when compared to the controls [mean=15.8, sd=8.5]. Significant differences (N=69 sites) in methylation patterns were also observed between cases and controls (5% false discovery rate), with 91% of these differentially methylated regions (DMRs) being attributable to increased methylation values in the women with FM. The DMRs involved 15 significant biological function groups, including (but not limited to) genes involved in neuron differentiation, nervous system development, skeletal/organ system development, and chromatin compaction. Specific genes with a DMR(s) included the brain-derived neurotrophic factor (BDNF) gene (noted to play an important neuromodulatory role in pain transduction, learning and memory), as well as genes involved in muscle contraction (NR4A3; HDAC4; FEZ1; PRKG1), chromatin compaction (NAT15; HDAC4; UHRF1), and DNA damage or chromosomal segregation (SOD3; UHRF1; NAT15). In summary, these results support the role of epigenetic and acquired chromosomal alterations as a possible biological mechanism leading to or resulting from symptoms associated with FM. Ultimately, since epigenetic changes demonstrate plasticity, recognition of consistent epigenetic alterations associated with FM could provide a means for developing future therapeutic or intervention strategies to treat this debilitating condition.

464T

Epigenome-wide methylation patterns across multiple fetal tissues.
A.C. Just[1], A.A. Baccarelli[1], R.J. Wright[2], H.H. Burris[3], G. Estrada[4], L. Schnaas[4], R.O. Wright[2]. 1) Harvard School of Public Health, Boston, MA, USA; 2) Icahn School of Medicine at Mount Sinai, NY, USA; 3) Department of Neonatology, Beth Israel Deaconess Medical Center, Division of Newborn Medicine, Children's Hospital Boston, Harvard Medical School, Boston, MA, USA; 4) National Institute of Perinatology, Mexico City, Mexico.

Rationale: Methylation is a common epigenetic mark with regulatory functions important for cell differentiation, normal development, and disease. Methylation is also responsive to environmental exposures. New tools enable measures of methylation with single-base resolution across the genome which can be utilized in large epidemiologic studies. A limitation however is that typically only surrogate tissues such as white blood cells have been studied. Human genomic methylation patterns in many potential target tissues have not yet been characterized. Objectives: To compare genomic methylation across multiple fetal tissues (cord blood, placenta, umbilical vein, umbilical artery) collected from the same individuals and identify similar and distinct patterns across samples. Methods: All four fetal tissues were collected for DNA extraction from participants (n=9, total tissue samples= 36) in Mexico City. Samples were run on the Illumina Human Methylation 450K beadchip and resulting % methylation values (betas) were adjusted with subset quantile normalization (Touliemat and Tost 2012). Methylation profiles were compared using multi-dimensional scaling and Pearson's correlation analysis using all probes passing QC as well as subsets defined based on genomic context. Measurements and Main Results: After initial quality control 428,055 CpG sites passed filtering for all 36 samples. Subsetting to autosomal probes, the largest Pearson's correlations between tissues, calculated for each participant, were for umbilical vein and umbilical artery (range 0.975 to 0.995) and the lowest for cord blood and placenta (range 0.796 to 0.834) with similar values when restricted to sites located in CpG islands. Conclusions: Correlation analysis demonstrates that DNA from umbilical vein and artery have more similar patterns of methylation than do cord blood versus placental DNA. In a larger cohort analysis currently underway (n = 130 participants), methylation measures in these fetal tissues will be related back to prenatal exposure to environmental stressors (e.g. lead, air pollution, psychosocial stress) and as mediators of associations between exposures and birth outcomes.

465F

Mapping the DNA unmethylome with mTAG, a novel method to investigate the epigenome.
V. Labrie[1], E. Kriukienė[2], T. Khare[1], G. Urbanaviěiūtė[2], A. Lapinaitė[2], K. Konceviěius[3], D. Li[4], T. Wang[4], S. Pai[1], C. Ptak[1], J. Gordeviěius[5], S.C. Wang[6], A. Petronis[1], S. Klimašauskas[2]. 1) The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, Toronto, Canada; 2) Department of Biological DNA Modification, Institute of Biotechnology, Vilnius University, Vilnius, Lithuania; 3) Faculty of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania; 4) Department of Genetics, Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, U.S.A; 5) Institute of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania; 6) Institute of Systems Biology and Bioinformatics, National Central University, Chungli, Taiwan.

DNA methylation is an epigenetic mechanism that contributes to the regulation of eukaryote genomes, and impacts normal phenotypic variation and disease risk in humans. We have developed a novel method to investigate DNA methylation, through the mapping of unmethylated regions of the genome. Our techniques known as DNA methyltransferase-direct transfer of activated groups (mTAG) uses covalent tagging of unmodified CpG sites, followed by affinity enrichment. The enriched unmodified DNA fraction is then analyzed with tiling microarrays or next-generation sequencing. Numerous control experiments and studies done with human genomic DNA from cultured cells and tissues demonstrate that the mTAG approach reveals unique cross-section through the complex epigenomic landscape, and offers high precision and robustness compared to existing affinity-based techniques. Overall, the mTAG technique is a valuable addition to the toolbox of epigenomic studies.

466T

Functional data analysis for identifying differentially methylated regions.
M. LeBlanc[1,2], C. Page[2,3], A. Frigessi[2], B. Kulle Andreassen[1,2]. 1) Epi-Gen, Institute of Clinical Medicine, Akershus University Hospital, University of Oslo and Akershus University Hospital, Oslo, Norway; 2) Department of Biostatistics, University of Oslo, Oslo, Norway; 3) Department of Neurology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

There is general agreement that methylation studies need to report results for differentially methylated regions (DMR). As such, methylation studies are well suited for the application of statistical methods that simultaneously analyze all available data for a given region of interest. One approach for this is functional data analysis (FDA), where statistics are calculated using functions describing the pattern of methylation over a defined region. In contrast, most methylation studies to date either use statistics based on single methylation probes or summarize single probe values within a region of interest with one value (e.g. maximum or mean). There are also methods using various smoothing methods, where the statistical approach operates with a summary measure based on the smoothed values (e.g. Jaffee 2012). We apply statistical methods based on FDA to methylation data. Using simulated data, we explore the utilities of FDA in the context of methylation and compare it to other methods used in the literature. In addition, we apply the proposed method to a recent dataset. The results encourage the use of FDA for methylation data in predefined regions of interest.

467F

DNA methylation analysis of iPS cells using whole-genome bisulphite sequencing. *D. Lee[1,2], J. Shin[1], P. Tonge[4], M. Puri[4], A. Nagy[4,5], J. Seo[1,2,3].* 1) GMI, Seoul national university, Seoul, Seoul, South Korea; 2) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea;; 3) Macrogen Inc., Seoul 153-023, Korea;; 4) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada.

Somatic cell reprogramming involves epigenetic remodeling of chromatin architecture including DNA methylation, conferring induced pluripotent stem cells (iPSCs) with characteristics similar to embryonic stem (ES) cells. However, it remains unclear how the epigenetic pattern changes during the reprogramming process and how it controls the messenger RNA expression. To address these questions, we utilized a secondary inducible reprogramming system developed in mouse using piggyBac-mediated transposition of the four reprogramming factors. Here we report whole-genome DNA methylation profiles at single-base resolution for mouse ES cells, secondary mouse embryonic fibroblast (MEF), secondary factor independent iPSCs induced from these MEFs, and cells undergoing reprogramming from Day 2 to 18, using whole genome bisulfite sequencing, along with comparative analysis of RNA expression and Histone modification (H3K4me3, H3K27me3, H3K36me3) using NGS. We clarified how the epigenetic change controls RNA expression and showed CpG methylation blocks the activation of core pluripotency genes. We defined 7990 differentially methylated regions (DMRs) in CpG context. Ectopic expression of reprogramming factors leads to a dynamic CpG methylation change from start of the process. Gain of CpG methylation occurs rapid and massive around PRCs (RING1, Suz12, EZH2) binding sites both in ESC like and aberrant direction. CpG methylation of transcription factor binding sites of 4 factors (Oct4, Sox2, Klf4, cMyc) or the factors activated early in reprogramming (Nanog) showed demethylation during high dox treatment, but only at the exact binding sites. However the samples became pluripotent showed demethylation in wide region around binding sites of all the core pluripotency factors, including Esrrb, Tcfcp2l1, and Zfx, We also observed that H3K4me3 and H3K27me3 are biased to regions showing low CpG methylation. These regions show dynamic change of both histone modifications and RNA expression. Thus, the regions with high CpG methylation are restricted to non-Histone modification and the genes having high CpG methylation promoters showed low expression. Our data provide evidence for how ectopic reprogramming factors affect CpG methylation in early stage of reprogramming and what is essential to gain pluripotency.

468T

Correlation and Null Hypothesis in Epigenome-wide Association Studies (EWAS). *B. Lehne[1], A. Drong[2], M. Loh[1], W. Zhang[1], W. Scott[1], M.R. Jarvelin[1,3,4,5], P. Elliott[1,3], M.I. McCarthy[2,6], J.S. Kooner[7], J.C. Chambers[1].* 1) Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom; 4) Institute of Health Sciences, University of Oulu, Oulu, Finland; 5) National Institute of Health and Welfare, Oulu, Finland; 6) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 7) National Heart and Lung Institute, Imperial College London, London, United Kingdom.

DNA methylation plays an important role in the regulation of gene expression and may be involved in the molecular mechanisms that lead to many human diseases. Currently multiple large-scale Epigenome-wide Association Studies (EWAS) are in progress to investigate the relationships between DNA methylation and phenotypic variation. In this work we analyse the p-value distribution that arises in an EWAS under no association, using permutation testing of Infinium 450K data for 2,660 individuals.

Under no association we observe a substantial deflation of test statistics (median genomic inflation factor λ_{median}=0.92). We demonstrate that this apparent deflation is caused by close correlation between methylation markers, even if these markers have entirely different genomic locations (mean pairwise correlation r_{mean}=0.23 for 1,000 randomly selected markers). This correlation structure substantially reduces the number of independent tests and affects the distribution of test statistics under the Null Hypothesis.

We identified multiple adjustments that reduce correlation between markers; Quantile Normalisation notably reduces correlation (r_{mean}=0.071) resulting in a reduction of statistical deflation (λ_{median}=0.968). Including technical confounding factors (bisulfite-batch, control probe intensities), biological confounding factors (gender, age, white-blood cells) and principal components (PC 1-5) as linear predictors in the regression model leads to a further reduction in correlation and statistical deflation. Together these adjustments remove the majority of correlation between markers (r_{mean}=0.018 for 1,000 randomly chosen markers) and re-establish the Null Hypothesis (λ_{median}= 1.003).

We conclude that correlation between markers is the consequence of biological and technical confounders, each of which affect the methylation status of multiple markers simultaneously. Correlation between markers affects the Null Hypothesis underlying an EWAS. Analysis of EWAS data is enhanced by careful adjustment for these confounding factors.

469F

Importance of Batch and White Blood Cell Subtypes Correction in Analysis of Illumina Infinium 450K Methylation Arrays. *M. Loh[1,2], B. Lehne[1], A. Drong[3], W. Scott[1], W. Zhang[1], M.-R. Jarvelin[1,2,4,5], P. Elliott[1,4], M.I. McCarthy[3,6], J.S. Kooner[7], J.C. Chambers[1].* 1) Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 2) Institute of Health Sciences, University of Oulu, Oulu, Finland; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 4) MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom; 5) National Institute of Health and Welfare, University of Oulu, Oulu, Finland; 6) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 7) National Heart and Lung Institute, Imperial College London, London, United Kingdom.

DNA methylation plays a key role in regulation of gene expression. Dense microarrays such as the Infinium 450K are currently in use to investigate relationships between methylation and complex disease, however most studies are limited to analysis of DNA from heterogeneous cell populations such as peripheral blood. We explore the impact of batch effect and variation in white blood cell (WBC) subsets on methylation association signals in studies of peripheral blood, and evaluate strategies for correcting for technical and biological confounding. We measured methylation in DNA extracted from peripheral blood from 1,072 people with incident Type-2 diabetes (T2D; cases) and 1,615 controls using the 450K array. 36 DNA samples were analysed in duplicate. Measured WBC subsets (basophils, eosinophils, lymphocytes, monocytes and neutrophils) were available for all participants. In addition we estimated WBC subsets (B-cells, granulocytes, natural killer [NK], monocytes and T-cells [CD4+ and CD8+]) using the method proposed by Houseman *et al.* (2012). Regression analysis was used to quantify the associations between methylation and replicate or case-control status, and adjustment for technical and biological confounding. Amongst the duplicate samples, we observed strong association between replicate batch and methylation (genomic inflation factor λ=2.188). Adjustment for 24 different in-built control probes on the array reduces the degree of statistical inflation, with bisulfite conversion control showing the strongest effect (λ=1.292). Simultaneous correction for all control probes corrects for majority of the inflation (λ=1.098). In the case-control study, there was high concordance between measured and imputed WBC subsets. We found close association of WBC subsets with DNA methylation (λ=3.090). The association of methylation with T2D case-control status showed substantial departure from null expectation (λ=1.130). Adjustment for inferred WBC subsets displayed better reduction in inflation (λ=1.076) compared to measured values (λ=1.134). In particular, adjustment for NK cells (λ=1.108) and monocytes (λ=1.118) had the strongest effects. Epigenome-wide association studies of DNA methylation are confounded by technical and biological factors, including the distribution of cell subtypes. Corrections for batch via the use of control probes, as well as cell subtypes at least partially reduces confounding and statistical inflation in analysis of these complex data.

470T

The effects of a therapeutic equivalent dose of olanzapine on genome-wide DNA methylation in hippocampus and cerebellum, in a rat model in vivo. *M.G. Melka[1], B.I. Laufer[1], P. McDonald[1], C.A. Castellani[1], N. Rajakumar[2], R. O'Reilly[2], S.M. Singh[1].* 1) Molecular Genetics Unit, Department of Biology; 2) Department of Psychiatry, University of Western Ontario, London, Ontario, Canada, N6A 5B7.

Background: The mechanism of action of olanzapine in treating schizophrenia is not clear. This research reports the effects of therapeutic equivalent treatment of olanzapine on DNA methylation in a rat model in vivo. Methods: Genome-wide DNA methylation was assessed using MeDIP-Chip analysis. All methylated DNA immunoprecipitation (MeDIP), sample labeling, hybridization, and processing were performed at Arraystar Inc. (Rockville, Maryland, USA). The identified gene promoters showing significant alterations to DNA methylation were then subjected to ingenuity pathway analysis (Ingenuity System Inc, CA, USA). Results: The results show that antipsychotic medications cause an increase in methylation in 1159, 1310 and 1329 genes and a decrease in methylation in 638, 567 and 533 genes in hippocampus, cerebellum and liver, respectively. Most genes affected are tissue-specific. Only 41 genes (~3%) affected showed an increase and no gene showed a decrease in methylation in all three tissues. Further, the two brain regions shared 127 (~10%) genes affected. Interestingly, the genes affected are enriched in pathways affecting dopamine signalling, molecular transport, nervous system development and functions in hippocampus; ephrin receptor signalling and synaptic long term potentiation in cerebellum; tissue morphology, cellular assembly and organization in liver. Also, the affected genes included those previously implicated in psychosis. Conclusions: The known functions of affected genes suggest that the observed epigenetic changes may underlie amelioration of symptoms as well as account for certain adverse effects including the metabolic syndrome. The results offer novel insight into the mechanisms of actions of olanzapine, in the understanding of therapeutic as well as side effects of antipsychotics.

471F

Astrocytic abnormalities and global DNA methylation patterns in depression and suicide. *C. Nagy[1,2], M. Suderman[3,4], M. Szyf[4], N. Mechawar[1,2], C. Ernst[1,2], G. Turecki[1,2].* 1) Medicine, McGill University, Verdun, Quebec, Canada; 2) McGill Group for Suicide Studies, Douglas Mental Health University Institute; 3) McGill Centre for Bioinformatics; 4) Department of Pharmacology and Therapeutics.

Background: Astrocytes are central nervous system specific cells involved in numerous brain functions including the regulation of synaptic transmission and of immune function. There is mounting evidence suggesting astrocytic dysfunction in major depression and suicide. Objective: To identify gene-based changes associated with astrocytes and to determine if DNA methylation plays a role in these abnormalities. Design: MBD2-based methylome sequencing of cases with astrocytic abnormalities. Participants, setting, and main outcome measures: We first characterized prefrontal cortex samples from 121 individuals (76 who died by suicide and 45 sudden death healthy controls) for the astrocytic markers GFAP, ALDH1L1, SOX9, GLUL, SCL1A3, GJA1, and GJB6. A subset of 22 cases with consistently downregulated astrocytic markers was then compared to 17 matched controls using MBD2-sequencing followed by validation with high resolution melting and bisulfite Sanger sequencing. With these data, we generated a genome-wide methylation map unique to altered astrocyte-associated depressive psychopathology and suicide. Results: The map revealed regions of differential methylation between cases and controls, the majority of which displayed reductions in methylation levels in cases. A large number of our differentially methylated regions overlapped with known ENCODE identified regulatory elements. Conclusions: Taken together, our data indicate significant differences in methylation patterns specific to astrocytic dysfunction associated with depression and suicide, providing a potential framework for better understanding this phenotype.

472T

The effects of perinatal testosterone exposure on DNA methylation in the brain are late-emerging and dynamic. *T.C. Ngun[1,2], N.M. Ghahramani[1,2], P.Y. Chen[3], S. Krishnan[1,2], S. Muir[1,2], T. TeSlaa[3], L. Rubbi[3], A.P. Arnold[2,4], G.J. de Vries[5], N.G. Forger[5], M. Pelligrini[3], E. Vilain[1,2].* 1) Department of Human Genetics, David Geffen School of Medicine at University of California Los Angeles (UCLA), Los Angeles, CA, USA; 2) Laboratory of Neuroendocrinology of the Brain Research Institute, UCLA, Los Angeles, CA, USA; 3) Department of Molecular, Cellular, and Developmental Biology, UCLA, Los Angeles, CA, USA; 4) Department of Integrative Biology and Physiology, UCLA, Los Angeles, CA, USA; 5) Neuroscience Institute, Georgia State University, Atlanta, GA, USA.

Sexual differentiation of the rodent brain is driven mainly by the actions of gonadal hormones, primarily testosterone (T) and estradiol. Perinatal exposure to T can have long-lasting and irreversible consequences on brain anatomy. These have been termed organizational effects, and lead to masculinization of certain brain regions, including those that control sexually dimorphic traits. The molecular mechanisms underlying these effects are still poorly understood. We hypothesized that perinatal T exposure will affect DNA methylation patterns in two sexually dimorphic brain regions - the bed nucleus of the stria terminalis/preoptic area (BNST/POA) and the striatum - and that this exposure contributes to the establishment of sex differences in the methylome. We established the first genome-wide methylation profiles for males (XY), females (XX), and females treated with T at birth (XX+T) at two time points. The first was postnatal day 4 (PN4) (within the critical window for T's organizational effects) and the second was PN60 (adulthood), to evaluate the longer term effects of T. Our approach was validated by the finding that the majority of X chromosome CpG sites differing between XX and XY mice at PN60 were hypermethylated in XX animals in both regions of the brain, consistent with X chromosome inactivation. The short-term effect of T exposure was relatively modest: 45 genes were differentially methylated between XX and XX+T in the BNST/POA, while 68 genes were influenced by T in the striatum. However, by PN60, this number had grown dramatically. The number of T-influenced genes was 760 in the BNST/POA and 1377 in the striatum. The traditional view of brain organization is that organization by T is stable. Therefore our finding that only a few of the genes affected by T at PN4 were the same as those affected at PN60 in either region was unexpected. Methylation at sexually dimorphic CpG sites in XX+T was more masculinized at PN60 than PN4 in both regions. However, the masculinizing effect of T on the methylome was stronger in the striatum than in the BNST/POA. Taken together, these data suggest that the organizational effect of T on the brain methylome is late-emerging and dynamic. Our data also suggest that organization by T may occur via early programming on relatively few genes and that this small initial effect sets up the brain to respond in a particular fashion to later developmental events. Drs. Ngun and Ghahramani contributed equally to this work.

473F

Genome-wide DNA Methylation Network Analysis for Osteoporosis Risk. *c. qiu[1], H. Shen[1], J. Li[1], H.W. Deng[1,2].* 1) Center for Bioinformatics and Genomics, Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; 2) Center of Systematic Biomedical Research, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Background: Osteoporosis is a common disease mainly characterized by low bone mineral density (BMD) and increased risk of fractures. Peripheral blood monocytes (PBMs) may act as precursors of osteoclasts, the bone resorption cells, and also produce cytokines important for osteoclast activity, and thus represent major systemic target cells for bone metabolism. Alterations in DNA methylation has been implicated as a key regulatory mechanism in the etiology of human complex diseases. Recent studies suggested that DNA methylome is organized into modules of co-methylated features. In this study, we carried out a network analysis to construct modules of highly co-methylated gene promoters in PBMs and identify modules that are significantly associated with BMD. Methods: Genome-wide DNA methylation profiles were generated by MeDIP-seq in PBMs from 18 unrelated Caucasian postmenopausal females with extremely high (n=9) and low (n=9) hip BMDs. MeDIP-seq signals were normalized and quantified using the MEDIPS analysis package. By focusing on the promoter DNA methylation data, we applied a weighted correlation network analysis (WGCNA) to identify the co-methylation modules and summarize the methylation profiles of each module into a single representative eigengene value. The eigengene values of individual modules were compared between the high and low BMD groups to identify co-methylation modules associated with BMD. Results: We identified a total of 18 co-methylation modules, each ranging in size of 32-251 gene promoters. Specifically, the overall methylation level of module-18 was significantly higher in the low-BMD group (p=0.002). Gene ontology analysis suggested that module-18 was highly enriched for genes belonging to a number of interesting biological processes, such as 'cellular response to vitamin D' (p=1.75E−16), 'blood vessel endothelial cell migration' (p= 4.33E−12), and 'cellular response to mechanical stimulus' (p=2.07E−09). Interestingly, several of the module-18 genes (e.g., AQP9, ITGB1) have been associated with BMD variation through previous genome-wide association studies. Conclusions: Using system level network analysis, we reconstructed the promoter co-methylation network in PBMs and identified a co-methylation module that may mediate variation in risk to osteoporosis. Our results highlighted the advantages of using systems-level network analysis to add value to the traditional DNA methylation analysis.

474T

Genome-wide placental DNA methylation analysis of severely growth-discordant monozygotic monochorionic twins reveals a novel epigenetic signature for IUGR. *M. Roifman[1], S. Choufani[2], S. Drewlo[3], S. Keating[4], J. Kingdom[5], R. Weksberg[1,2].* 1) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 2) Genetics and Genome Biology Program, Hospital for Sick Children Research Institute, University of Toronto, Toronto, Canada; 3) Research Centre for Women's and Infants' Health, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Canada; 4) Department of Laboratory Medicine and Pathology, Mount Sinai Hospital, University of Toronto, Toronto, Canada; 5) Maternal-Fetal Medicine Division, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada.

Background: Intrauterine growth restriction (IUGR) refers to reduced fetal growth in the context of placental insufficiency. IUGR is not only associated with perinatal morbidity and mortality, but has also been linked to adult-onset diseases, such as obesity, hypertension and diabetes, and thus poses a major health burden. Placental epigenetic dysregulation has been implicated in IUGR; however, clear pathophysiological mechanisms remain to be elucidated. Monozygotic monochorionic (MZ-MC) twins are particularly affected by IUGR in the setting of severe discordant growth with unequal placental share. Because MZ twins have the same genotype at conception, and a shared maternal environment, they provide a unique model for studying epigenetic regulation of the placenta (i.e. the non-shared environment). **Objective:** We compared genome-wide placental DNA methylation patterns of severely growth-discordant MZ-MC twins to identify novel candidate genes for IUGR in the context of unequal placental share. **Methods:** We studied 8 severely growth-discordant MZ-MC twin pairs, each with unequal placental share. Snap frozen placental samples were obtained at delivery from each twin. Placental pathology was histologically determined. We used the Infinium HumanMethylation450 BeadChip array platform to identify genomic regions exhibiting differential methylation in IUGR versus normal twins. **Results:** Our analysis of larger twins (n=8) versus smaller twins (n=8) identified 138 candidate genes with statistically significant ($p<0.05$) differential methylation (using a cutoff of 10% difference in DNA methylation). Across these 138 genes, the lower weight twins exhibit a distinctly different methylation signature compared to their higher weight counterparts. These genes, many linked to IUGR for the first time in this study, confer a variety of functions, including cellular development, signaling, angiogenesis, and lipid metabolism. Largest methylation differences between the two groups were found in the *DECR1*, *LEPR* and *SPG7* genes, implicating lipid metabolism and mitochondrial dysfunction as prominent pathophysiological mechanisms in this form of IUGR. **Conclusion:** We propose a novel epigenetic signature for IUGR in the context of unequal placental share in MZ-MC growth-discordant twins, highlighting lipid metabolism and mitochondrial dysfunction as major contributors, and suggesting an underlying mechanism for the fetal reprogramming associated with adult-onset diseases.

475F

Genetics of global DNA methylation patterns in multiple tissues from twins. *J.K. Sandling[1,8], E. Grundberg[1,2,9], E. Meduri[1,2,10], A.K. Hedman[3], S. Keildson[3], J. Nisbet[1], A. Barrett[4], K.S. Small[2], B. Ge[5], E.T. Dermitzakis[6], M.I. McCarthy[3,4,7], T.D. Spector[2], J.T. Bell[2], P. Deloukas[1], the MuTHER consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK; 5) McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 6) Department of Genetic Medicine and Development and Institute for Genetics and Genomics in Geneva, University of Geneva Medical School, Geneva, Switzerland; 7) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK; 8) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 9) Present address: Department of Human Genetics, McGill University, McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 10) Present address: Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK.

Epigenetic mechanisms, such as DNA methylation, are attractive candidate processes for explaining the interplay of genetic and environmental factors in complex traits. We explored the associations between genetics and epigenetics utilizing genome-wide DNA methylation profiles (Illumina HumanMethylation450) from skin (N=469) and adipose tissue (N=648) from female twins in the MuTHER study. We found that methylation levels were generally highly correlated between skin and adipose tissue from the same individual ($r_{s\ median}$=0.96); this dropped significantly if only the (10%) most variable sites were analysed ($r_{s\ median}$=0.78). For around 1% of CpG sites we observed marked differences in average methylation levels between the tissues (Δbeta>0.3). We found methylation variability to be suppressed in regions known to impact gene regulation such as promoters (adipose: P<1.E-5; skin: P<1.E-4). We then studied the degree of association between adipose DNA methylation and expression of nearby genes in this tissue, and of 13,532 genes we found 17% to have at least one significant association with a slight enrichment of negative correlations. Concordance in methylation for both adipose and skin tissue was greater between MZ twins than DZ twins, indicating genetic influences on DNA methylation. Narrow-sense heritability analysis showed evidence for DNA methylation being heritable (h^2_{mean} adipose=0.19; skin=0.13) and half of the heritable adipose CpG sites were associated with common variants (metQTLs) located in close proximity to the methylation site (FDR1%, median distance 12kb). Cross-tissue comparison revealed that the vast majority (63%) of adipose metQTLs could be replicated in skin (P<0.05 and same direction of effect; estimated proportion of significant tests π1=0.74). Our results show that there is widespread genetic regulation of CpG methylation in adipose and skin tissue. Integrating our adipose metQTL data with adipose reference epigenomes (H3K4me1 and H3K4me3) from the NIH Roadmap Epigenomics and the GWAS catalogue (NHGRI) revealed significant enrichment of metQTLs overlapping with metabolic disease loci in enhancer elements. Finally, we have been scanning cardiometabolic and skin related traits available in the MuTHER study for association to methylation changes. As expected we find strong tissue-phenotype dependency; for example, 3.5-fold more methylation sites in adipose tissue associated with body mass index than in skin (FDR1%).

476T

The Highs and Lows (of Methylation) in the Placental Epigenome. *V. Shridhar, P. Shaw, K. Bunce, T. Chu, D. Peters.* Obstetrics and Gynecology, University of Pittsburgh, Pittsburgh, PA.

Fetal health and development during gestation have long-term consequences upon the overall health of an individual. Regulation of fetal development is strongly influenced by gestational age-appropriate cues from the uterine environment. These cues often take the form of epigenetic marks at the feto-maternal interface on the placenta. There is increasing evidence that the placental epigenome has far-reaching effects on both prenatal and postnatal life of the fetus.

We generated the methylation profiles of 6 placental (from 3 male and 3 female fetuses) as well as 3 unmatched maternal blood cell samples, using Agilent Technology's SureSelect Methyl-seq targeted enrichment kits. We found that placental DNA is hypo-methylated, as compared to the maternal blood cell DNA, in accordance with previously published results. There were smaller-sized tracts in placental DNA that were methylated, compared with maternal DNA. We classified the placental epigenome into methylated (>80%) and unmethylated (<20%) domains. We correlated methylation levels of the genes within these domains to publically available expression level data from gestational-age matched placental tissue. At a gene level, there was a positive correlation between methylation and expression. Gene ontology groups that were over-represented in the methylated domains included embryonic morphogenesis, endopeptidase activity, cell fate commitment, organ development and regulation of apoptosis (p<0.05). Ontology groups over-represented in the unmethylated domains included metal ion homeostasis, GPCR activity and heterochromatin formation among others (p<0.004). There were some gender-related differences in placental methylation levels. Analysis is currently ongoing to better define these regions of high and low methylation as well as the physiological relevance of these methylation sites. Term placentas are being harvested to study the changes in epigenome over time. Additionally, we intend to extend this study for comparison of methylation profiles of placentas from pregnancies with complications, such as preeclampsia and pre-term labor.

We anticipate that these studies will yield novel insights into placental gene expression over the course of gestation and into the epigenetic mechanisms of preeclampsia and pre-term birth.

477F

Whole-genome bisulfite sequencing (WGBS): A novel `post-bisulfite conversion' library construction method from low gDNA inputs. *R. Sooknanan, J. Hitchen, D. Gabel, V. Ruotti.* Epicentre® (an Illu1EPICENTRE Biotechnologies. LLC (an Illumina® company), Madison, WI, USA mina® company), Madison, WI.

Genome-wide analysis of 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC) are possible with whole-genome bisulfite sequencing (WGBS), where unmethylated cytosine residues are converted to uracil. Further, with the use of T4 β-glucosyltransferase (T4-BGT) and Ten-eleven Translocation Gene Protein 1 (Tet1) enzymes, 5-hmC can be discriminated from 5-mC on a genome-wide basis. However, a major challenge in WGBS is the degradation of DNA that occurs during bisulfite conversion under conditions required for complete conversion. Typically, ~90% of input DNA is degraded and thus, is especially problematic with limited starting amounts of DNA. Additionally, regions that are rich in unmethylated cytosines are more sensitive to strand breaks. As a consequence, a majority of DNA fragments contained in di-tagged NGS DNA libraries treated with bisulfite 'post-library construction' can be rendered inactive due to strand breaks in the DNA sequence flanked by the adapter sequences. These mono-tagged templates are then excluded during library enrichment resulting in incomplete coverage and bias when performing whole genome bisulfite sequencing. Here, we describe a novel 'post-bisulfite conversion' library construction method for preparing NGS libraries from genomic DNA prior to the addition of the adapters. This 'post-bisulfite conversion' library construction method uses the untagged single-stranded DNA as template for the subsequent addition of adapter sequences required for NGS. Thus, single-stranded DNA fragments independent of size and position of strand breaks remain as viable templates for library construction, eliminating the loss of fragments and the selection bias associated with a 'post-library construction' bisulfite conversion strategy. This novel 'post-bisulfite conversion' library construction method exhibits increased sensitivity and efficiency and improved coverage required for WGBS for detecting 5-mC and 5-hmC marks.

478T

A novel sequencing method for genome-wide profiling of 5-hydroxymethylcytosine with single-base resolution. *D. Tan, H. Chung, X. Sun, X. Jia.* Zymo Research Corp., Irvine, CA.

5-hydroxymethylcytosine (5hmC) is an epigenetic mark abundant in embryo stem cells and brain tissues. The exact biological functions of 5hmC are still under close investigation although several lines of evidence have indicated it could be involved in active DNA demethylation. Meanwhile, extensive studies have been carried out to determine its genomic distribution. A number of approaches have been developed using either affinity based enrichment, such as hMeDIP, that rely on antibody and other specific binding proteins to target 5hmC, or modified bisulfite sequencing, namely oxidative bisulfite sequencing (OxBS) and TET assisted bisulfite sequencing (TAB-sequencing). However, all those methods have limitations which hamper their application. For example, affinity based methods lack single base resolution while modified bisulfite sequencing methods require efficient chemical or enzymatic oxidation which cannot be easily achieved or guaranteed. As an alternative, we have developed a novel genome-wide sequencing method that utilizes an enzyme based modification approach coupled with bisulfite-sequencing for detecting 5hmC. This methodology allows quantification of 5hmC levels with single CpG resolution and can also be employed for locus-specific assays. Using this method, we were able to map and quantify 5hmC sites at the genomic scale for several different biological samples. This novel method can determine the exact location and abundance of 5hmC, which will facilitate our understanding of 5hmC in regulating gene expression in different biological contexts.

479F

The role of DNA methylation in B°AT1 transcriptional regulation along crypt-villus axis. *E. Tumer, T. Juelich, S. Broer.* Research School of Biology, The Australian National University, Canberra, Australia.

Hartnup disorder is an autosomal recessive disorder caused by mutations in the B°AT1 (Slc6a19) gene encoding the major transport system for neutral amino acids in the intestinal brush-border membrane. B°AT1 protein is expressed at the apical membrane of enterocytes in increasing amounts along the crypt-villus axis. Previously, we showed that HNF1a and HNF4a bind to the Slc6a19 promoter up-regulating transcription. Sox9, by contrast, suppressed the promoter activity induced by HNF1a and HNF4a. Sox9 is a transcription factor known to be involved in the differentiation of stem cells into mature enterocytes, and is highly expressed in the crypt region, but absent at the villus tip. Thus Sox9 expression could explain the gradient of B°AT1 in the intestine. However, despite HNF1a and HNF4a expression in the liver, B°AT1 is not expressed, while Sox9 is absent. As a result we considered DNA methylation as a mechanism to regulate B°AT1 expression. DNA methylation of a transcription factor binding sites can prevent transcription factor from binding to the DNA. Kikuchi et al (2010) suggested that HNF1a binding to the B°AT1 promoter could be prevented by methylation in the liver. However, they observed particular methylation in the liver at position -1080, which is 940 bp upstream of the HNF1a binding site. In this study we investigated the DNA methylation status of the B°AT1 promoter, using bisulfite sequencing. DNA methylation was determined in a 1.2 kb region upstream of the B°AT1 transcriptional start site in liver, kidney and crypt-villus preparations. From each experiment a minimum 10 independent clones were sequenced. Twenty CpG sites were analysed and found to be differentially methylated in the chosen tissues. We found that CpG dinucleotides around HNF1a, HNF4a and TATA binding sites were hypomethylated in the villus, whilst being hypermethylated in the crypt and in liver tissue. These results indicate that DNA methylation might play a role together with Sox9 in the repression of B°AT1 expression in the crypt, whereas in the liver cells DNA methylation itself might be sufficient. Kikuchi, R., S. Yagi, et al. (2010). "Genome-wide analysis of epigenetic signatures for kidney-specific transporters." Kidney international 78(6): 569–577.

480T

Epigenome-wide differences in DNA methylation of autosomes in Klinefelter's Syndrome. *E.S. Wan[1], W. Qiu[1], J. Morrow[1], T.H. Beatty[2], J. Hetmanski[2], E.K. Silverman[1], D.L. DeMeo[1] on behalf of the COPDGene Investigators.* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Rationale: Klinefelter's Syndrome (47 XXY) affects approximately 1 in every 600 human male births and is associated with hypogonadism and impaired fertility, tall stature with eunuchoid proportions, and a variety of cognitive and behavioral disorders. The degree to which Klinefelter's (KF) subjects present with this clinical spectrum is highly variable and may suggest a role for epigenetics. While the role of DNA methylation in X-inactivation among KF subjects has been explored, differential methylation at autosomal sites has not previously been described.

Methods: All subjects were recruited from the COPDGene study, a multi-center, cross-sectional cohort of non-Hispanic white and African American current and former smokers with ≥10 pack-years of smoking. Six KF subjects were identified during genome-wide genotyping. Genome-wide DNA methylation data was obtained on 5 KF subjects (the 6th had insufficient DNA) and 105 male control subjects using the Illumina HumanMethylation450K Beadchip. Preprocessing, quantile normalization, and adjustment for batch effects were performed using *limma*, *minfi* (v 1.4.0) and *sva* (v 3.4.0) packages in R (release 2.15); all pre-processing and subsequent analyses were performed separately for Infinium I and Infinium II probe types. Empirical Bayes methods were used to test for differential methylation between KF subjects and controls at 470,326 autosomal CpG sites using linear models adjusted for age, race, and body mass index (BMI). A false-discovery rate of <0.05 was used to denote significance.

Results: There were no significant differences by race or current smoking status between KF subjects and controls; KF subjects were significantly younger and had a higher mean BMI than controls. 204 Infinium I sites and 275 Infinium II sites, including sites annotated to the G-protein coupled receptor 27 (*GPR27*) and Fas associated factor 1 (*FAF1*), met the *a priori* threshold for significance.

Conclusions: Differential methylation at non-imprinted autosomal CpG sites exists in Klinefelter's syndrome and may highlight novel loci associated with the clinical manifestations of the syndrome.

481F

Ascorbate is a possible cofactor for Tet proteins to hydroxylate 5mC to 5hmC. *G. Wang, E. Minor, K. Dickson, C. Gustafson, B. Court, J. Young.* Dr. John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst.for Human Genomics, Univ Miami, Miami, FL.

Ascorbate (vitamin C) is best known for its role in scurvy, in which the hydroxylation of collagen catalyzed by collagen hydroxylases is incomplete due to ascorbate deficiency. Recently, Tet (ten-eleven translocation) family dioxygenases (also known as Tet proteins) have been identified to hydroxylate 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA. Both Tet proteins and collagen hydroxylases belong to the iron and 2-oxoglutarate-dependent dioxygenase superfamily. These dioxygenases utilize Fe^{2+} as a cofactor, 2-oxoglutarate as a co-substrate, and some of them including collagen hydroxylases require ascorbate as another cofactor for full catalytic activity. We found that the content of 5hmC was extremely low in mouse embryonic fibroblasts (MEFs) cultured in ascorbate-free medium. Additions of ascorbate dose- and time-dependently enhanced the generation of 5hmC, without any effects on the expression of Tet genes. Treatment with another reducer glutathione (GSH) did not change the level of 5hmC. Further, blocking ascorbate entry into cells by phloretin and knocking down Tet1-3 expression by short interference RNAs (siRNA) significantly inhibited the effect of ascorbate on 5hmC. The effect of ascorbate on 5hmC generation appeared independent on iron uptake by cells or the production of 2-oxoglutarate. These results suggest that ascorbate enhances 5hmC generation, most likely by acting as a cofactor for Tet proteins to hydroxylate 5mC to 5hmC. Thus, we have uncovered a novel role for ascorbate in modulating the epigenetic control of genome activity.

482T

A new target enrichment system generating ultra-high complexity probe pools for targeted bisulfite sequencing. *J. Wendt[1], M. Suzuki[2], T. Richmond[1], N. Patterson[2], T. Millard[1], D. Green[1], R. Bannen[1], R. Selzer[1], T. Albert[1], J. Jeddeloh[1], J. Greally[2], D. Burgess[1].* 1) Roche NimbleGen Inc., Development and Technology Innovation Groups, Madison, WI; 2) Albert Einstein College of Medicine, Center for Epigenomics, Bronx, NY.

DNA methylation has been shown to have a role in a host of biological processes, including silencing of transposable elements, stem cell differentiation, embryonic development, genomic imprinting, and inflammation, as well as, many diseases, including cancer, cardiovascular disease, and neurologic diseases. Epigenetic modifications can also affect drug efficacy by modulating the expression of genes involved in the metabolism and distribution of drugs, as well as, the expression of drug targets, contributing to variability in drug responses among individuals. There are currently a number of tools to study DNA methylation status, either at a single locus level, using methods like methylation-specific PCR or MALDI-TOF-MS, or at a broader, genome-wide level, using DNA microarrays, reduced representation bisulfite sequencing (RRBS), or even whole genome shotgun bisulfite sequencing. The latter method is preferred by many researchers, as it provides DNA methylation status at base pair resolution and allows for the assessment of percent methylation at each position in the genome. However, it is expensive in terms of money and analysis to generate such data for the entire genome, when generally only a subset of the genome is of interest to most researchers. We describe a system for the targeted enrichment of bisulfite treated DNA, allowing researchers to focus on a subset of the genome for high resolution cytosine methylation analysis. Regions ranging in size from 10 kb to 75 Mb may be targeted, and multiple samples may be multiplexed and sequenced together to provide an inexpensive method of generating methylation data for a large number of samples in a high throughput fashion. Innovations that differentiate our system, including probe design, selection and manufacture, as well as optimization techniques to improve capture uniformity, will be highlighted.

483F

Epigenome-wide association study of Autism Spectrum Disorder using Post-Mortem Brain Tissue. *C.C.Y. Wong[1], D.H. Geschwind[2,3,4], N. Parikshak[2], C. Troakes[5], J. Viana[6], D. Condliffe[1], T.G. Belgard[2], S. Prabhakar[7], J. Mill[1,6].* 1) King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, London, UK; 2) Center for Autism Research and Treatment, and Program in Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA; 3) Department of Human Genetics, University of California, Los Angeles, CA, USA; 4) Program in Neurogenetics, Department of Neurology, University of California, Los Angeles, CA, USA; 5) King's College London, MRC Centre for Neurodegeneration Research, Institute of Psychiatry, De Crespigny Park, London, UK; 6) University of Exeter Medical School, Exeter University, St Luke's Campus, Exeter, UK; 7) Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis St, Singapore 138672, Singapore.

Autism Spectrum disorders (ASD) is a group of complex neurodevelopmental disorders characterized by considerable etiological heterogeneity. Although genetic variation in several dozen ASD genes has been implicated in the development of ASD, collectively accounting for 10-20% of ASD cases, there is emerging evidence to suggest a role for epigenetic variation in the disorder. In this study, we performed an epigenome-wide association study (EWAS) of ASD using a large collection of post-mortem brain tissue (n=139) obtained from four brain regions. DNA methylation was quantified using the Illumina Infinium 450K Human Methylation array in conjunction with an analysis pipeline developed by our group. We identified a number of significant disease-relevant differentially methylated regions (DMRs), including several located in the vicinity of genes previously implicated in psychiatric disorder. Pathway analysis of our top loci highlighted a significant enrichment of epigenetic disruption in biological networks and pathways relevant to disease and neurodevelopment. Overall, our data provide further evidence to support a role of differential DNA methylation in the etiology of ASD.

484T

Racial and Ethnic Variation in DNA Methylation of Human Peripheral Blood. *Y. Yang[1], R. Sebra[1], Y. Li[2], J.F. DeCoteau[3], I. Peter[1], S.A. Scott[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Anesthesiology, Xuzhou Medical Colleage, Xuzhou, Jiangsu, China; 3) Cancer Stem Cell Research Group, University of Saskatchewan, Saskatoon, SK, Canada.

Genome-wide DNA methylation profiles can differ between human cell types and tissues; however, little is known about CpG methylation variability between racial and ethnic populations. To determine if significant ethnic-specific CpG methylation differences exist in germline DNA, 60 African-American, Caucasian, Chinese, and Hispanic peripheral blood DNA samples from unrelated healthy adults were subjected to DNA methylation profiling using the HumanMethylation450 BeadChip (450K-array; Illumina). 450K-array probes that contained known single-nucleotide polymorphisms or mapped ambiguously to the human genome were removed, and batch effects controlled for by quantile normalization. To validate these data, selected samples were subjected to targeted genome-wide bisulfite sequencing (84 Mb; 24x average coverage; Agilent Technologies), which had high concordance with 450K-array methylation profiling (r=0.976). Unsupervised hierarchical clustering of 450K-array probes with high beta-value (β) standard deviation (SD>0.15) among all 60 samples suggested ethnic-specific methylation profiles for the four tested populations. Moreover, probes were tested for differential ethnic-specific methylation using the Kruskal-Wallis test and a population mean $|\Delta\beta|>0.15$, which identified 407 CpG methylation sites that were significantly differentially methylated between at least two of the four populations ($p<0.001$). Six selected regions subsequently were validated by third-generation bisulfite single-molecule real-time (SMRT®) sequencing of population-pooled ~500 bp amplicons (1123x average coverage; Pacific Biosciences), which confirmed ethnic-specific 450K-array probe and/or region methylation differences. Among the 407 identified ethnically variable CpG methylation sites (located in or near 265 genes), the majority were located outside of CpG islands and their methylation profiles were sufficient to accurately separate all four populations by principal component analysis. Although DNA methylation can change with age and be modulated by environmental stimuli, these data suggest that a small proportion of CpG sites in the human genome are epigenetically stable in peripheral blood mononuclear cells and can reflect racial and/or ethnic ancestry. Additionally, these data have important implications for epigenome-wide association studies of complex human diseases and suggest that further investigation on the mechanism of ethnically variable CpG methylation is warranted.

485F

Gene based association analysis of brain DNA methylation with Alzheimer's disease pathology using random permutation. *L. Yu[1], G. Srivastava[2,3,4], L. Chibnik[2,3,4], M. Eaton[4,5], S. Leurgans[1], A. Meissner[4,6], P.L. De Jager[2,3,4], D.A. Bennett[1].* 1) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 2) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology & Psychiatry, Brigham and Women's Hospital, Boston, MA; 3) Harvard Medical School, Boston MA; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA; 6) Department of Stem Cell and Regenerative Biology, Harvard University, Harvard Stem Cell Institute, Boston, MA.

Background: Gene based analysis of DNA methylation serves as a useful tool to test the global hypothesis for regional association of CpGs with traits of interest. In order to minimize the inflation of Type I error, the correlations between CpGs must be considered. However, the correlation structure of CpGs has not been fully understood. We propose a permutation based method to evaluate methylation associations by gene regions controlling for correlations among CpGs. Methods: We analyzed gene based associations of brain DNA methylation at 11 validated AD loci, *APOE*, *CR1*, *BIN1*, *CD33*, *CLU*, *ABCA7*, *CD2AP*, *PICALM*, *EPHA1*, *MS4A6A* and *MS4A4A*, with the pathologic diagnosis of Alzheimer's Disease (AD). We utilized data from two ongoing longitudinal cohort studies of aging and dementia, the Religious Orders Study (ROS) and the Memory and Aging Project (MAP). Genome wide methylation profiles were generated from dorsolateral prefrontal cortex tissue samples using Illumina HumanMethylation450 beadset. We first applied logistic regression models with AD diagnosis according to NIA Reagan criteria as the binary outcome, and examined the associations of individual CpGs at each of the 11 loci. Next, for each locus we combined the significance values of individual CpGs using the truncated product method; and the observed statistics were tested against corresponding empirical null distributions derived by random permutation of the phenotype to obtain one methylation p-value per gene. Results: 719 ROS and MAP individuals were included in the analysis. An average age at death was 88 years (SD=6.6 years), 63.4% were females and 60.1% met criteria for pathologic AD. The most significant genes for the association of DNA methylation with AD were *ABCA7* (p<10-6) and *BIN1* (p=8.4×10-5). Similar associations were nominally significant for *CLU* (p=0.036) and *MS4A6A* (p=0.054). Little evidence of association of DNA methylation with AD was seen with the remaining AD loci. Conclusion: These findings suggest that gene based disease associations can be driven by genomic and epigenetic mechanisms. Gene based association analysis using random permutation helps to investigate regional association while accounting for the internal structure of genetic markers.

486T

Down-regulated microRNA-199a-3p induces aberrant DNA methylation in testicular cancer by directly targeting DNMT3A. *B. Chen, S. Gu, Y.K. Suen, L. Li, W.Y. Chan.* School of Biomedical Sciences, the Chinese University of Hong Kong, Shatin,, N.T.,, Hong Kong.

It was previously demonstrated that miR-199a was downregulated in testicular germ cell tumor (TGCT) probably caused by hypermethylation of its promoter. Further studies found that re-expression of miR-199a in testicular tumor cells (NT2) led to suppression of cell growth, cell migration, invasion and tumor metastasis. More detailed analyses showed that these properties of miR-199a could be assigned to miR-199a-5p, one of its two derivatives. The biological role of the other derivatives, miR-199a-3p in TGCT, remains largely uncharacterized. In this report we identified DNA (cytosine-5)-methyltransferase 3A (DNMT3A), the de novo methyltransferase, as a target of miR-199a-3p. We demonstrated that after transient transfection of miR-199a-3p into NT2 cells, DNMT3A expression (especially the DNMT3A gene isoform 2, DNMT3A2) was significantly decreased at mRNA and protein levels. The dual luciferase reporter assay results showed that miR-199a-3p significantly reduced the luciferase activity in NT2 cells when co-transfected with miR-199a-3p and Luc-DNMT3A construct harboring the potential binding site for miR-199a-3p. In clinical samples, DNMT3A2 was significantly overexpressed in malignant testicular tumor tissues, and the expression of DNMT3A2 was inversely correlated with the expression of miR-199a-3p. However, DNMT3A expression did not negatively regulate miR-199a expression in NT2 cells. Further characterization of miR-199a-3p revealed that it negatively regulated DNA methylation partly through targeting DNMT3A. Overexpression of miR-199a-3p restored the expression of APC and MGMT tumor-suppressor genes in NT2 cells via affecting DNA methylation of their promoter regions. Our studies demonstrated the deregulation of miR-199a-3p expression in TGCT may provide novel insights into the mechanism of TGCT carcinogenesis and suggested a potentially therapeutic use of synthetic miR-199a-3p oligonucleotides as effective hypomethylating compounds in the treatment of TGCT.

487F

The utility of DNA extracted from saliva for methylation studies of psychiatric traits. *A.K. Smith[1,3], V. Kilaru[1], T. Klengel[2], K.B. Mercer[3], K.N. Conneely[3,4], K.J. Ressler[1], E.B. Binder[1,2].* 1) Psychiatry & Behavioral Sciences, Emory University, Atlanta, GA; 2) Max Planck Institute of Psychiatry, Munich, Germany; 3) Genetics and Molecular Biology Program, Emory University, Atlanta, GA; 4) Human Genetics, Emory University, Atlanta, GA.

DNA methylation has become increasingly recognized in the etiology of psychiatric disorders. Because brain tissue is not accessible in living humans, epigenetic studies are often conducted in blood, but few are conducted in children because even a blood draw may be too invasive. Saliva is readily collectable, but the proportion of epithelial cells and leukocytes varies between individuals and represents a significant confound. This study will evaluate whether DNA isolated from saliva is comparable to DNA from blood for outcomes relevant to psychopathology; because of the wealth of adult and animal data, child abuse will be used as an example. Saliva and blood samples were collected from 64 African American participants in the Grady Trauma Project, each of which completed the Childhood Trauma Questionnaire to assess abuse. DNA methylation was interrogated with the HumanMethylation450 BeadChip. The proportion of epithelial cells in saliva and specific leukocyte types in blood were estimated. We examined the association of each CpG site with child abuse using linear models that adjusted confounding factors such as age, sex, race, batch effects and cellular heterogeneity. Similar models were used to test whether DNA methylation levels in saliva predicted those of blood. For all analyses, the FDR was controlled at 5 percent. Hierarchical clustering segregated blood samples from saliva samples. Overall, DNA methylation of only ~10 percent of CpG sites from saliva predicted that of blood. While no individual CpG site remained associated with child abuse after adjustment for multiple testing, the test statistics for analyses of saliva and blood were moderately correlated (r=.21). We evaluated CpG sites in 11 genes that were reported as differentially methylated in the blood of those with a history of child abuse. The majority of those genes contained CpG sites whose saliva methylation levels were highly predictive of those in blood ($p<1\times10^{-7}$) though certain areas of the gene (i.e. promoters) were more likely to contain such sites. These results have applications for longitudinal and biomarker studies as well as large-scale DNA methylation studies of childhood psychiatric disorders. DNA derived from saliva may be informative for research questions that can be assessed in blood, though only a small fraction of CpG sites can be considered correlated. Tissue-specific differences seem more prominent than those related to shared genetic or environmental factors.

488T

Rapid DNA Methylation Analysis of the FXN gene in Friedreich ataxia cells and tissues. *S. Al-Mahdawi, C. Sandi, M.A. Pook.* Biosciences, Brunel University, Uxbridge, United Kingdom.

Friedreich ataxia (FRDA) is caused by a homozygous GAA repeat expansion mutation within intron 1 of the FXN gene, leading to reduced expression of frataxin protein. Evidence suggests that the mutation induces epigenetic changes and heterochromatin formation, thereby impeding gene transcription. Thus, studies using FRDA lymphocytes and lymphoblastoid cell lines have detected increased DNA methylation of specific CpG sites upstream of the GAA repeat, together with histone acetylation and methylation changes in regions flanking the GAA repeat. We have previously reported DNA methylation changes in FRDA brain, cerebellum and heart tissues, which are the primary affected systems of the disorder. Bisulfite sequence analysis of the FXN flanking GAA regions revealed a shift in the FRDA DNA methylation profile, with upstream CpG sites becoming consistently hypermethylated and downstream CpG sites becoming consistently hypomethylated. However, bisulfite sequencing is a time-consuming procedure. Therefore, we have now developed specific 'methylscreen' restriction enzyme digestion and qPCR-based protocols to more rapidly quantify DNA methylation at four of the upstream CpG sites. Analysis of human cerebellum and heart tissue samples was used to validate the technique by comparison with previous bisulfite sequencing results. Increased DNA methylation in both FRDA cerebellum and heart tissue samples compared with unaffected control samples was confirmed at all 4 CpG sites. Increased DNA methylation at all four CpG sites in the upstream GAA repeat region was also confirmed in human FRDA fibroblast cells, which have not previously been characterized. The levels of DNA methylation at two CpG sites in the upstream GAA repeat region, together with transgenic FXN mRNA levels and frataxin expression levels were determined in Y47R control and YG8R FRDA mouse fibroblasts, neural stem cells (NSCs) and neuronal/glial differentiated NSCs. In all three cell types, there is increased DNA methylation in YG8R FRDA cells compared with Y47R control cells, which is associated in all cases with decreased FXN mRNA and frataxin protein expression in YG8R FRDA cells compared with Y47R control cells. These methylscreen protocols will allow the rapid detection of DNA methylation changes at the FXN locus in FRDA cells and tissues subsequent to epigenetic-based therapeutic testing.

489F

Systematic classification of common disease-associated regulatory DNA variations by their epigenomic relationship. *M. Dozmorov, C. Giles, J. Wren.* Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK.

Background: The success of genome-wide association studies (GWASs) in finding causative genomic variants for Mendelian phenotypes is contrasted with their inability to accurately elucidate complex patterns and biological roles of mutations underlying non-Mendelian inheritance. Our motivation was to find common epigenomic elements enriched with sets of disease-specific SNPs, and to systematically classify the diseases by their epigenomic background. Methods: Human disease-specific sets of SNPs were extracted from the UCSC GWAS catalog. We used our method, GenomeRunner (http://sourceforge.net/projects/genomerunner/) to test them for statistically significant associations with epigenomic data extracted from the UCSC genome database. Disease-specific epigenomic associations were compared with random associations, obtained by testing random sets of SNPs sampled from the GWAS catalog. P-values of enriched associations were calculated using Fisher's exact test, and corrected for multiple testing using Benjamini-Hochberg procedure. Results: 212 disease and 363 trait/phenotype associated sets of SNPs were tested for associations with >4,000 genome annotation data. We identified that diseases/traits of similar origin (immunological, neurological, metabolic) tend to act within similar epigenomic context, and can be distinguished by their epigenomic associations. Our results suggest that alterations of specific epigenomic regulators may underlie disease susceptibility, paving a road for future epigenomic drug design and therapeutic targets. Conclusion: The vast and growing amount of genome annotation data contains enormous potential to interpret sets of disease-associated mutations within a common, unifying theme of epigenomic regulators. Considering these themes will empower us to interpret the results of GWASs in terms of unifying mechanisms, complementing SNP-gene-pathway approaches. Conversely, similarities and differences in epigenomic context of disease- and trait-associated SNPs provide a new means to classify phenotypes and understand their common epigenomic denominators.

490T

Integrated epigenetic analysis of APOBEC gene family in breast cancer. *Y. Zhang, R. Delahanty, W. Zheng, J. Long.* Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.

The APOBEC gene family was initially reported to play critical roles in virus restriction. We recently found that a common germline deletion in the *APOBEC3B* gene was strongly associated with breast cancer risk. It has been reported that the *APOBEC3B* gene may play a significant role in C-to-T mutations in breast cancer. However, it remain unknown how *APOBEC3* family genes are regulated. In this study, we aimed at determining epigenetic profile of activation and inactivation of APOBEC family members in breast cancer cell lines through large-scale integrated analyses. All data were downloaded from the GEO/ENCODE databases. RNA-seq data were available from 12 cell lines, including three estrogen receptor negative (ER-), eight estrogen receptor positive (ER+) breast cancer cell lines as well as one normal breast epithelia cell line. Data of epigenetic markers based on ChIP-seq, including H3K4me3, H3K27ac, H3K4me1, H3K36me3, H3K9me3 and H3K27me3, were available from both ER+; and ER- breast cancer cell lines. DNA methyl-seq data were obtained only in one ER- cell line. RNA-seq data showed that among *APOBEC* genes, the *APOBEC3B* gene was the only one showing up-regulation compared with normal breast cell lines. Other APOBEC members, including *APOBEC1*, *APOBEC2*, *APOBEC3A*, *APOBEC3D*, *APOBEC3F*, *APOBEC3G*, *APOBEC3H*, *APOBEC4* and *AID* were either not expressed or down-regulated in breast cancer cell lines with an exception of the constantly high expression of the APOBEC3C gene observed in ER- breast cancer cell lines but down-regulation in ER+; breast cancer cell lines. Activated epigenetic markers, including H3K4me3 and H3K27ac and H3K36me3, were observed in the *APOBEC3B* gene, in both ER+; and ER- breast cancer cell lines. The activated markers H3K4me3 and H3K36me3 were also observed in the *APOBEC3C* gene in the ER- cell lines. Except the *APOBEC3B* and *APOBEC3C* genes, all other *APOBEC* family members showed DNA hyper-methylation at their promoters in the ER- cell line, which may contribute to their no expression or down-regulation. In conclusion, our integrated epigenetic analysis showed that both histone modification and DNA methylation may regulate gene expression patterns in APOBEC family members in breast cancer cell lines.

491F

Role of CTCF in epigenetic regulation of 4qD4Z4 macrosatellite repeat. *J. Lim[1], J.M. Moore[1], N.A. Rabaia[1], Y.D. Krom[2], S.J. Tapscott[1,3], S.M. van der Maarel[2], G.N. Filippova[1].* 1) Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 3) Department of Neurology, University of Washington, Seattle, WA.

Facioscapulohumeral muscular dystrophy (FSHD) is caused by incomplete repression of the D4Z4 macrosatellite repeat array on the disease-permissive chromosome 4q that results in aberrant expression of DUX4, the candidate FSHD gene imbedded within the D4Z4 repeat. Loss of repressive chromatin modifications at the D4Z4 array in FSHD has been reported to be associated with increased binding of the chromatin insulator protein CTCF (Zeng, et al., 2009 and Ottaviani, et al., 2009). However, due to repetitive nature of the D4Z4 array, there is very limited information available on CTCF binding at this region. Here we characterize CTCF binding at the 4q35 D4Z4 locus and identify several CTCF binding sites within each D4Z4 repeat unit with two sites located at the boundary with the proximal p13E-11 sequence as well as a cluster of five CTCF sites at the distal D4Z4/pLAM region flanking the pLAM polyadenylation site that is critical for generation of polyadenylated DUX4 transcripts. Moreover, one of the CTCF binding sites at the pLAM region shows FHSD-related haplotype-specific binding. We confirm that CTCF binds to the D4Z4 repeats in both FSHD1 and FSHD2 muscle cells but not in controls. Consistent with the role of CTCF in protection of genes from epigenetic silencing, we observe a significant loss of both histone H3K9 trimethylation and DNA methylation at CTCF binding sites at D4Z4 in FSHD cells in comparison to controls. Depletion of CTCF results in increase of histone H3K9 methylation at the D4Z4 and reduction of the DUX4 transcripts levels in FSHD myoblasts. Our findings suggest that CTCF binding at the D4Z4 locus in FSHD may interfere with the setting and/or maintenance of repressive epigenetic marks and result in inappropriate DUX4 transcription.

492T

Subtelomeric deletions exert telomere position effects on nearby genes. *J. Gerfen, H.M. Mason-Suares, M.K. Rudd.* Department of Human Genetics, Emory University, Atlanta, GA.

Copy number variations (CNVs) are a major cause of autism, intellectual disability, and birth defects. In some cases, loss or gain of genes within the CNV can explain clinical features in affected individuals. However, many pathogenic CNVs are associated with patient phenotypes without pinpointing the gene(s) involved. Position effect, where genes nearby but outside the CNV have altered expression, is another process that could be involved in disease etiology. In particular, telomere position effect (TPE) is a well-described epigenetic process in model systems where genes are silenced by proximity to a heterochromatic telomere. We measured TPE in lymphoblastoid cell lines derived from patients with pathogenic subtelomeric deletions using ChIP-chip and RNA-seq followed by targeted analysis with qPCR and next-generation sequencing of PCR amplicons. Analysis of four different subtelomeric deletions demonstrated a variety of chromatin and expression changes in genes adjacent to deletions. In one 5.5-Mb deletion of the short arm of chromosome 4 we detected an ~400-kb region enriched in histone three lysine nine trimethylation, a marker of heterochromatin, directly adjacent to the chromosome breakpoint. RNA-seq analysis of *EVC*, a gene ~150-kb from the breakpoint and within the region of heterochromatin, revealed allele-specific skewing of expression. We found that one allele accounts for 80% of the transcripts, suggesting that the *EVC* locus on the other chromosome is at least partially silenced. Other genes outside the region of heterochromatin were expressed at similar levels from both alleles. Silencing of *EVC* may be responsible for the atrial septal defect in this patient. Heterozygous mutations of *EVC* have been reported in a family with single atria, and recessive mutations lead to Ellis-van Creveld syndrome, which includes structural defects of the atria. These data support our hypothesis that heterochromatin originating from a new telomere can silence genes adjacent to a subtelomeric deletion. Other deletions in our study showed weaker TPE. We are currently characterizing additional chromatin modifications associated with human TPE and measuring the distance of TPE spreading in other CNVs. Our experiments are the first to demonstrate TPE from a native human telomere. These data will elucidate the role of CNV position effects in disease and may reveal new genes involved in patient phenotypes.

493F

Chromatin state characterization of GWAS results of different neuropsychiatric traits is suggestive of brain-specific as well as non-neuronal origins of disease. *R.A. Ophoff[1], J. Ernst[2].* 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, California; 2) Depts of Biological Chemistry and Computer Science, UCLA, Los Angeles, California.

The Psychiatric Genomics Consortium (PGC) has been instrumental in GWAS of neuropsychiatric traits. GWAS results of the different disease groups have been encouraging with partly overlapping findings between the disorders. A significant proportion of GWAS identified disease variants localize outside known genic regions but within regulatory elements as described through genome-wide functional data as generated by the ENCODE project. As part of our effort to better understand the PGC GWAS findings at the functional level we set out to study the chromatin state of the top SNP findings for bipolar disorder, schizophrenia and major depressive disorder. We obtained the PGC GWAS results of the different disorders and generated a clumped ranked list of independent SNPs based on association signal and removing any SNPs in LD. The ranked list of SNPs was compared with chromatin state maps defined by NIH Epigenomics Roadmap Consortium based on applying ChromHMM to multiple histone modifications mapped across 90 samples covering a wide range of different primary cell types. We focused the comparison on a canonical enhancer state, associated with high cell type specificity, to assess whether biologically relevant cell types were preferentially associated with GWAS prioritized SNPs. We computed the significance of the overlap of the number of GWAS SNPs overlapping the enhancer state using a binomial distribution where the probability of overlap was based on the frequency of the enhancer state among all SNPs in the clumped ranked list. Instead of focusing on one specific cutoff in the GWAS ranked list, we compared the overlap at each position within the top several thousand and ranked the cell types based on the most significant p-value obtained at any cutoff. We observed enrichment of GWAS signal highlighting neuronal-derived cell types for bipolar disorder and schizophrenia. However, ChromHMM analysis of the GWAS data of major depressive disorder completely lacked evidence of involvement of neuronal cell types. Schizophrenia and bipolar disorder share a substantial genetic risk, which may explain our findings of enrichment for neuronal cells, even though our results may indicate different classes of neuronal cells to be involved. The lack of enrichment of neuronal cells for GWAS of major depressive disorders could imply a non-neuronal origin for the disease.

494T

Histone H2A variants: one family, different roles in the human genome. *M.Y. Tolstorukov[1,2], J.A. Goldman[1], C. Gilbert[3], V. Ogryzko[3], N. Volfovsky[4], R.M. Stephens[4], R.E. Kingston[1], P.J. Park[5].* 1) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA; 2) Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) Oncogenese, Differenciation et Transduction du Signal, Institut Andre Lwoff, Villejuif, France; 4) Advanced Biomedical Computing Center, SAIC-Frederick, NCI, Frederick, MD, USA; 5) Center for Biomedical Informatics, Harvard Medical School, Boston, MA.

Readout of genomic information is controlled and modulated by chromatin structure, which at the basic level is represented by the DNA wrapped around the histone core. Although chromatin structure in human cells has been extensively investigated in recent years, the biological role and genomic distribution of the replacement histone variants remain poorly understood. Using publicly available and newly generated data, we focus on the variants of histone H2A, one of the most diverse histone families. In particular, we produced genome-wide profiles of the variants H2A.Z, macroH2A and H2A.Bbd using HeLa cell lines that stably express affinity-tagged versions of the corresponding histones. We report that nucleosomes bearing variant H2A.Bbd protect less DNA and are enriched inside actively transcribed genes. This is in contrast to macroH2A nucleosomes, which are enriched in repressed genes. At the same time, H2A.Bbd and macroH2A are not mutually exclusive and a detectable fraction of the genome is enriched for both variants. To further investigate the role of the recently discovered variant H2A.Bbd we preformed a comparative analysis of the transcription products in the cells where H2A.Bbd was depleted with shRNA and in control. This analysis showed that the H2A.Bbd depletion results in the 'net' down-regulation of gene expression and in the disruption of mRNA splicing pathways. Thus, our analysis suggests that H2A.Bbd may be involved in the formation of a specific chromatin structure that facilitates transcription elongation and initial mRNA processing. We also observe that specific chromatin organization involving histone variants may affect the level of conservation of the underlying DNA sequence. For instance, we report that the loci preferentially occupied by the nucleosomes bearing H2A.Z variant as well as such histone modifications as H3K4me3 show decreased frequency of SNPs as compared to the loci associated with 'bulk' nucleosomes. Taken together, our results demonstrate that H2A variants play highly specialized roles in human chromatin and that their distribution is evolutionary conserved.

495F

Spread of repressive chromatin from the expanded GAA trinucleotide repeat mutation contributes to gene silencing in Friedreich ataxia. *Y.K. Chutake[1], S.I. Bidichandani[1, 2].* 1) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Department of Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Friedreich ataxia (FRDA) is a neurodegenerative disease caused by an expanded GAA trinucleotide repeat (TNR) in intron 1 of the *FXN* gene. Reduction in levels of *FXN* mRNA and subsequently the mitochondrial protein, frataxin, has been attributed to the expanded TNR. The expanded GAA TNR interferes with transcriptional elongation via at least two mechanisms: formation of one or more abnormal DNA structures, and by formation of repressive chromatin. Together these result repeat length dependent transcriptional deficiency downstream of the GAA TNR. However, consistent with the expanded GAA TNR being a source of position effect variegation in mammalian cells, we demonstrate via mononucleosomal chromatin immunoprecipitation that the repressive chromatin spreads upstream from the GAA TNR thus encompassing the transcription start site and promoter of the *FXN* gene. Indeed, Nucleosome Occupancy and Methylome sequencing (NOMe-seq) experiments revealed reduced accessibility via tight nucleosomal occupancy in the vicinity of the transcriptional start site in FRDA cells versus controls. The spread of repressive chromatin is associated with reduced binding of the chromatin insulator protein CTCF and Myc, which are known regulators of *FXN* gene transcription. Indeed, FRDA patients show a deficiency of transcription, reduced occupancy of RNA pol II, and reduced H2A.Z upstream of the expanded GAA TNR, indicating that a less permissive transcriptional landscape extends far upstream from the expanded GAA TNR. Therapeutic strategies for FRDA would have to effectively target the repressive chromatin spanning a larger region of the *FXN* gene than previously recognized.

496T

Complex molecular findings in 11p15-associated Imprinting Disorders require a broadening of testing strategies. *M. Begemann, S. Spengler, L. Soellner, T. Eggermann.* RWTH Aachen, Aachen, Germany.

The chromosomal region 11p15 is one key player in molecular processes regulated by genomic imprinting. Genomic as well as epigenetic disturbances affecting the two imprinting control regions (ICRs) 11p15 are associated either with Silver-Russell syndrome (SRS) or Beckwith-Wiedemann syndrome (BWS): SRS patients carry ICR1 hypomethylation, maternal 11p15 duplications or uniparental disomies (UPD), whereas the opposite alterations are characteristic for BWS. In the last years, a growing number of patients suffering from imprinting disorders (IDs) has been reported showing a hypomethylation at the two ICRs in 11p15 as well as at further imprinted loci on other chromosomes. The molecular basis of these multilocus methylation defects (MLMDs) is widely unknown, however an interaction between trans-localized imprinted genes via a so-called Imprinted Gene Network (IGN) has been suggested. Cases with other types of molecular disturbances than aberrant methylation (e.g. uniparental disomy (UPD), chromosomal imbalances) support this IGN concept. The complex molecular alterations as well as the overlapping and sometimes ambiguous clinical findings in ID patients often make the decision for a specific ID test difficult. As aforementioned, the 11p15 loci are consistently affected in all ID patients with more complex alterations, but the pattern of affected loci is nearly unpredictable. We therefore suggest to implement molecular tests in routine ID diagnostics which allow the detection of a broad range of (epi)mutation types (epimutations, UPDs, chromosomal imbalances) and cover the currently known ID loci. The need to apply multilocus tests (methylation-specific MLPA, SNuPE) is corroborated by our experience from routine diagnostics in more than 650 patients referred as SRS or BWS: (a) Multilocus tests increase the detection rates as the tests ascertain numerous loci. Thereby the chance to identify even slight mosaic hypomethylation patterns is growing. (b) Patients with unusual phenotypes and unexpected molecular alterations will be detected. (c) The testing of rare imprinting disorders becomes more efficient and quality of molecular diagnosis increases. (d) The tests identify MLMDs. In future, the detailed characterization of clinical and molecular findings in ID patients will help us to decipher the complex regulation of imprinting as the basis of an individualized therapeutic management in IDs.

497F

Allelespecific analysis of putative imprinted gene network members in human. *D. Prawitt[1], F. Bohne[1], D. Langer[1], U. Martiné[1], K. Oexle[2], B.U. Zabel[3], T. Enklaar[1].* 1) Molecular Genetics Lab, Center for Pediatrics and Adolescent Medicine, Univ. Medical Center, Mainz, Germany; 2) Institute of Human Genetics, Technical University, Munich, Germany; 3) Center for Pediatrics and Adolescent Medicine, University Medical Center, Freiburg, Germany.

The two imprinting disorders Beckwith-Wiedemann syndrome (BWS) and Silver-Russell syndrome (SRS) present with marked intrauterine and postnatal overgrowth (BWS) or growth retardation (SRS). Molecular defects are heterogeneously associated and are mainly due to epigenetic changes in imprinting control regions (ICR) of the 11p15.5 region. ICR1 regulates the monoallelic expression of *IGF2* and *H19 in cis*. A subgroup of SRS and BWS patients present with epigenetic alterations of other chromosomal regions, arguing for a functional dependence of the affected imprinted genes. To date the molecular mechanisms of the ICR1 effects are incompletely defined and their analysis often has to be performed using genetically engineered model organisms. The description of a network of imprinted genes (IGN) in mice raised the possibility to improve the comprehension of multilocus imprinting defects in humans and depicts candidate genes involved in associated clinical symptoms. The presented work offers an unusual mechanistic insight into ICR1-driven regulation of imprinted genes in human cells. By utilizing unique fibroblasts with paternal (BWS) or maternal (SRS) uniparental disomy 11p15, we uncovered a set of transcriptionally co-regulated imprinted genes on different chromosomes. According to findings for murine orthologues the genes can be part of an IGN involved in regulating developmental growth.

498T

The allelic expression of Gnas is affected by mutations in Mecp2 and maternal strain. *H. Stitik[1,2], R.D. O'Connor[1,2], A. Kumar[2], N.C. Schanen[1,2].* 1) Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA; 2) Nemours Biomedical Research, Alfred I duPont Hospital for Children, Wilmington, DE 19803, USA.

The GNAS gene encodes the G protein stimulatory alpha subunit of (Gsα) and shows tissue specific monoallelic expression that is regulated through a methylation-dependent process. To determine whether the Methyl CpG Binding Protein 2 (MeCP2) was involved in regulation of imprinting of the Gnas transcript, we examined allelic expression of Gnas in tissues from male progeny of an outcross between C57Bl/6J females, wild type or heterozygous for a null allele for Mecp2 (Mecp2tm1.1Bird/J), and Mus castaneus (CAST) males. Using a strain specific polymorphism, we determined the parental origin of the Gnas transcript in tissues at postnatal days 3 (P3) and 28 (P28). In keeping with previous reports, our control cross showed a preference for maternal expression of Gnas in tissues that are imprinted. However, unexpectedly, we found that allelic expression of Gnas was altered in both wildtype and mutant progeny of the Mecp2+/- mothers, showing preferential paternal expression of the Gnas transcript in a number of tissues. Furthermore, we observed both a loss as well as a gain of imprinting at P3 in the kidney and calvaria, respectively. Since a strain effect has been reported in other imprinted loci, we sought to determine whether the strains used in this outcross could be affecting the expression of Gnas by performing a reciprocal cross between CAST female X C57Bl/6J male mice. These studies revealed expression patterns consistent the previous wildtype outcross with the exception of the spleen, which showed preferential expression of the C57Bl/6J allele. Additionally, we examined the differentially methylated region (DMR) in exon 1A that is responsible for controlling expression of Gnas using sodium bisulfite conversion and sequencing analysis to determine if there was a developmental shift in methylation in wild type C57Bl/6J mice. In all tissues examined, methylation increased significantly from P3 to P28. These studies suggest a transgenerational effect of maternal Mecp2 genotype on imprinting of Gnas in mice that occurs in a tissue specific manner and demonstrate epigenetic regulatory processes are impacted by strain and age.

499F

Exploring the role of NLRP7 in reprogramming of CpG methylation in the oocyte and early embryo. *S. Mahadevan[1,2,3], S. Wen[2,4], Y. Wooi[2,3], Z. Liu[3,5], M. Kyba[6], B. Sadikovic[4,7], I. Van den Veyver[1,2,3,4].* 1) Interdepartmental Program in Translational Biology and Molecular Medicine, Baylor College of Medicine, Houston, TX; 2) Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX; 3) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX,; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics (Neurology), Baylor College of Medicine, Houston, TX; 6) Lillehei Heart Institute and Department of Pediatrics, University of Minnesota, Minneapolis, MN; 7) Department of Pathology and Laboratory Medicine, McMaster University, Hamilton, ON, Canada.

Maternal effect mutations in NLRP7 cause recurrent, biparental hydatidiform moles (BiHM), a trophoblastic disease characterized by loss of imprinting at maternally imprinted loci. The existing paradigm suggests a loss of methylation at maternally imprinted loci, while studies from our group in hESC revealed a global, genome wide shift in methylation dynamics causing us to further investigate the extent to which loss of NLRP7 affects the cellular epigenetic state. We had previously shown that NLRP7 physically interacts with the transcription regulator YY1. ChIP-qPCR of YY1 in HEK293 cells with overexpression of NLRP7 revealed an increased occupancy of YY1 at imprinted loci such as PEG3 and SNRPN. This increased occupancy of YY1 at its target sites is not caused by an upregulation in YY1 expression or intra-cellular redistribution within the cell. Fortuitously, it was also observed that NLRP7, which was previously thought to be entirely cytoplasmic in localization, was capable of migration into the nucleus. Given that YY1 is also nuclear in localization, NLRP7's presence in the nucleus likely contributes to the observed differences in YY1 occupancy at its target sites. Computational interrogation of the protein sequence of NLRP7 revealed a putative bipartite nuclear localization sequence (NLS). In the absence of NLRP7's ability to localize within the nucleus, it will be informative to assess whether YY1 occupancy at its target sites is also affected. This information is not only valuable from the perspective of understanding NLRP7 function but is suggestive of NLRP7's potential of being a novel YY1 cofactor. The observation that the loss of methylation at maternally imprinted loci is caused by a maternal loss of NLRP7 argues in favor of a role for this protein in the oocyte and / or early embryo in the process of imprint acquisition and/or maintenance. A systematic interrogation of proteins with known roles in imprint acquisition and/or maintenance revealed physical interaction of NLRP7 with CG Binding Protein 1 (CGBP / CFP1). Given that distinct domains of CFP1 function as mediators of crosstalk between histone and DNA methylation, current efforts are focused on delineating which domains of CFP1 interact with NLRP7. The overall scope of this project is to gain insight into the roles of these proteins in the reprogramming of CpG methylation in the oocyte to embryo transition.

500T

Epigenetic effects of the endocrine disruptor phthalate: influence of the genetic background. *A. Paoloni-Giacobino[1,2], Ch. Stouder[1], E. Somm[3].* 1) Department of Genetic Medicine and Development, Geneva University Medical School, Geneva, Switzerland; 2) Swiss Center for Applied Human Toxicology, Geneva, Switzerland; 3) Department of Pediatrics, Geneva University Medical School, Geneva, Switzerland.

Di-(2-ethylhexyl)phtalate (DEHP) is a plasticizer with endocrine disrupting properties found ubiquitously in the environment. The aim of the present study was to evaluate the possible deleterious effects of DHEP administration in pregnant mice, on imprinted genes in the offspring. Mice of 2 different genetic backgrounds (C57BL/6 and FVB/N) were chosen and treated with DHEP during the time of embryo sex determination. Potential effects of DHEP on the methylation of the differentially methylated domains (DMDs) of 4 paternally expressed genes, Snrpn, Peg1, Peg3 and Igf2r and of 3 maternally expressed genes, H19, Gtl and Igf2 were investigated. The degrees of methylation of the imprinted genes were analysed in the sperm as well as in the liver and skeletal muscle. In parallel, the mRNA expressions of Snrpn, Peg3, Igf2r in these same tissues were studied to assess the link between the imprinted gene methylation and expression. In the sperm DEHP did not affect the degree of methylation of the paternally expressed genes except that of Igf2r in the C57BL/6 mice but induced decreases in the degree of methylation of all the maternally expressed genes in the C57BL/6 mice whereas it had no effect on these same genes in the FVB/N mice. In the liver, DEHP induced ubiquitous effects, potentially representing a relaxation of the monoallelic methylation pattern, in all or in 6 out 7 of the imprinted genes tested in C57BL/6 and FVB/N mice, respectively. In the muscle, 5 out of 7 and 3 out of 7 imprinted genes were affected by DEHP in the C57BL/6 and FVB/N mice, respectively. Eventually, the mRNA expressions in the sperm, liver and muscle of 2 imprinted genes, Snrpn and Peg3 were increased by DEHP in FVB/N mice, in the absence of any change of DMD methylations. In conclusion, DEHP has an effect on sperm DNA methylations that seems to be dependent on the genetic background. The direct effects of DEHP on somatic cells affect practically all imprinted genes studied, especially in the liver. The effects of DEHP on gene expressions are either direct or mediated by epigenetic effects other than DNA methylation.

501F

Establishing induced pluripotent stem cell-derived neurons as an appropriate cellular model for studying mechanisms of myoclonus-dystonia. *K. Freimann, A. Westenberger, P. Seibler, A. Weissbach, N. Brueggemann, K. Lohmann, C. Klein, A. Gruenewald.* Institute of Neurogenetics, University of Luebeck, Luebeck, Germany.

Myoclonus-dystonia (M-D) is a movement disorder presenting with a combination of dystonic features and brief myoclonic jerks. Mutations in the *epsilon sarcoglycan* (*SGCE*) gene have been found to cause this autosomal dominantly inherited disorder. Due to maternal imprinting the penetrance of M-D is reduced. *SGCE* encodes one ubiquitously expressed and one brain-specific isoform of a membrane-associated glycoprotein. In order to understand the molecular mechanisms leading to M-D it is important to develop a cellular model that recapitulates the conditions (including the presence of methylation - imprinting) in neuronal cells of patients. To establish whether induced pluripotent stem cell (iPSC)-derived neurons are such an appropriate model, we compared the imprinting status of *SGCE* in fibroblasts, iPSCs, and iPSC-derived neurons of one *SGCE* mutation carrier and two healthy controls. To distinguish between methylated and unmethylated sequences, we treated DNA extracted from these cells with bisulfite. Furthermore, we analyzed the presence of the two different *SGCE* isoforms by direct sequencing of cDNA reversely transcribed from mRNA extracted from fibroblasts and iPSC-derived neurons of one control. Upon bisulfite treatment, differential methylation of the *SGCE* promoter was detected in fibroblasts, iPSCs, and iPSC-derived neurons of two controls and one M-D patient, resembling the methylation status previously shown for lymphoblast-derived DNA. Sequencing revealed that fibroblasts express only the ubiquitously present *SGCE* isoform, whereas iPSC-derived neurons express the ubiquitous isoform as well as the brain-specific *SGCE* isoform. These findings suggest that the process of reprogramming of fibroblasts to iPSCs and further differentiation of iPSCs into neurons does not alter the imprinting status of *SGCE* in these cells. Additionally, the finding of the brain-specific *SGCE* isoform in iPSC-derived neurons, that was not detected in non-neuronal cells taken from patients (i.e. fibroblasts), suggests that splicing of *SGCE* changes in a tissue-specific manner when patient-derived fibroblasts are transformed into neurons. Thus, neurons derived from iPSCs of *SGCE* mutation-carrying patients can be considered not only a biologically relevant but also functionally adequate cellular model for studying molecular mechanisms leading to M-D.

502T

Full resolution DNA methylome analysis in multiple tissues from twins. *S. Busche[1,2], M. Caron[2], T. Kwan[2], V. Forgettta[1,3], B. Ge[2], S. Westfall[2], J. Qi[1,2], M-M. Simon[2], J.T Bell[4], . MuTHERConsortium[4,5], . UK10KConsortium[4,5], J.B. Richards[1,3,4], G. Bourque[1,2], M. Lathrop[1,2], P. Deloukas[5], T.D Spector[4], T. Pastinen[1,2], E. Grundberg[1,2].* 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 3) Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 4) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 5) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Numerous population studies applying methylome analyses by the Illumina450K array are on-going, but no comprehensive or unbiased analysis of CpG variation at the population level has been carried out to date. Using adipose tissue (AT) and whole-blood (WB) from monozygotic (MZ) and dizygotic (DZ) twins of the well characterized MuTHER/TwinsUK cohort (e.g. dense phenotype, 450K and WGS data available) we aim to unravel the impact of stochastic, environmental, and genetic factors underlying methylation (mCpG) variation and the impact on disease susceptibility using whole genome bisulfite sequencing (WGBS). So far, we have analyzed 26 AT and 21 WB samples (6 MZ pairs, 5 DZ pairs, 7 singletons) generating a total of 1.97 Tbp sequence at on average 6.5-fold coverage for AT and 9.7-fold for WB samples. Totally, we identified 21M and 24M CpG sites in AT and WB with an average methylation level of 85%. CpG values detected by WGBS at > 4-fold coverage were highly concordant with overlapping Illumina 450K data (Pearson's r = 0.93-0.95) but only <1% of global mCpG variation is accessible through the 450K array. Pairwise comparisons of MZs (<20% vs. >80% methylated) allowed us to identify the proportion of mCpG variation that is of environmental origin, which corresponded to on average 0.03% in AT and 0.015% in WB. Similar comparisons of DZs and unrelated yielded a ~5- and ~10-fold increase. Sequence variants at the CpG site appeared to underlie most non-environmental mCpG, and thorough variant removal reduced this increase to ~2-fold. Further analysis of environmental driven differentially methylated regions (eDMRs; ≥ 3 consecutive CpGs within 500 bp and ΔmCpG>40%) using MZs revealed on average 90 eDMRs per tissue and pair with no overlap across tissue, highlighting tissue-specificity with likely correlation with phenotypic discordance. We also studied tissue-specific DMRs (tDMRs) and found AT-specific, hypomethylated tDMRs to be highly enriched (17-fold) in human adipocyte enhancer elements (H3K4me1, NIH Roadmap Epigenomics Project). Finally, the existence of WGS (UK10K project) will allow us to define the largest allele-specific methylation (ASM) dataset in multiple tissues, including random, parent-of-origin and sequence-dependent effects. Preliminary data indicates ASM effects accounting for 1% of all CpGs, where a minor proportion (3.4%) corresponds to known imprinted loci with the remaining effect most likely being due to sequence-dependent ASM effects.

503F

Epigenome-wide association study on identical twins discordant for birth weight. *Q. Tan[1], M. Frost[2], L. Christiansen[1].* 1) University of Southern Denmark, Odense, Denmark; 2) Odense University Hospital.

Epidemiological evidences have shown that early life exposure to adverse environment, e.g. bad nutrition, and stress during pregnancy of mother which usually resulting in low birth weight, could have long-term health consequences including metabolic disorders, diabetes, hypertension, and even mental diseases in adult life. Epigenetics is a molecular mechanism that explains the acquired changes or modification in gene function independent of DNA sequence variation with DNA methylation as one of several epigenetic mechanisms most intensely studied. In order to examine if low birth weight induces persistent epigenetic modification detectable at adult ages, we performed a genome-wide DNA methylation profiling in peripheral blood of 150 pairs of identical Danish twins discordant for birth weight using the Illumina Infinium HumanMethylation450 BeadChip featuring 485,000 CpG sites across the genome. After quality control and data preprocessing using free R package minfi, data were analysed by a mixed effects model including fixed effect variables such as birth weight difference, age and sex of twin pairs; random effect variables such as batch, well, and sample position on the array, etc. Statistical analysis revealed 12 probes with p value<1e-05, among them 1 probe with p value<1e-06. Biological pathway analysis using these probes showed no significant functional cluster involved. In conclusion, our analysis suggests that discordant birth weight in identical twins is not associated with measured DNA methylation level at adult ages. Similar studies on tissues other than peripheral blood are required in order to re-confirm and generalize our conclusion.

504T

Identification of a differentially methylated locus in monozygotic twins discordant for esophageal atresia. *L. Boutaud de la Combe[1], A. Pelet[1], C. Bole[2], J. Tost[3], W. Carpentier[4], N. Cagnard[2], R. Smigiel[5], V. Rousseau[6], F. Gaudrin[7], J. Amiel[1], S. Lyonnet[1].* 1) INSERM U781,Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris 15, Paris, France; 2) Plateforme de génomique et de bioinformatique, Université Paris Descartes-Sorbonne Paris Cité, Institut Imagine, Paris 15, Paris, France; 3) Centre National de Génotypage CEA - Institut de Génomique 2 rue Gaston Crémieux CP5721 91057 Evry Cedex France; 4) Plateforme P3S,UPMC Pitié Salpétrière 91 Blv de l'Hôpital 75013 Paris France; 5) Genetics Department Wroclaw Medical University Marcinkowskiego 1 PL 50-368 Wroclaw, Poland; 6) Service de Chirurgie Viscérale Pédiatrique Hôpital Necker 149 rue de Sèvres 75743 Paris Cedex 15, France; 7) AFAO 56 rue Cécile 94700 Maisons-Alfort France.

Until recently monozygotic twins were considered to be genetically identical. However the increasing description of monozygotic twins discordant for a phenotype leads us to change our point of view. Numerous causes can explain the discordance between twins. In addition to environmental causes, genomic and epigenomic modifications can occur in only one of the sibs, such as mosaic mutations, post-zygotic alterations and DNA methylation. In order to understand the molecular bases of isolated esophageal atresia, we are studying monozygotic twins discordant for this serious and frequent congenital malformation. Our work aims at finding the origin of the esophageal atresia that occurs in only one child of each twin pair. We decided to investigate two hypotheses: genomic and epigenomic. Eight twin pairs were studied; one twin of each pair has esophageal atresia with or without other congenital abnormalities (such as anomalies of the heart or kidney). All the pregnancies were the result of natural conception and the parents were not consanguineous. Twins were raised together and the children are 2 to 7 years of age. Thus, we assume that environmental differences are negligible. Genomic and epigenomic hypothesis were tested searching first for copy number variations (CNVs) in the twins DNA and then for methylation differences between the child suffering from esophageal atresia and his healthy co-twin. No CNVs were found, however, using a cytosine microarray (Illumina Infinium HumanMethylation450 BeadChip) we have compared 450,000 CpG loci and have found a modification of methylation ($p < 0.05$) at a locus that could explain the phenotypical discordance between twins. This locus shows both a hypermethylation in the promoter and a hypomethylation in the gene body in the affected sibs compared to healthy co-twins. We predict that it could modify the expression of this gene.

505F

An epigenome-wide association study using monozygotic twins identifies putative loci associated with osteoporotic phenotypes. *V. Forgetta[1], W. Kirsten[2], P. Deloukas[2], T. Spector[2], J. Bell[2], JB. Richards[1].* 1) Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada; 2) Department of Twin Research, King's College London, London, United Kingdom.

Aim: To identify epigenetic loci associated with bone mineral density (BMD) phenotypes using monozygotic (MZ) twin data from EpiTwin (http://www.epitwin.eu/), a TwinsUK project (http://www.twinsuk.ac.uk/). **Methods:** As part of the EpiTwin project, a total of 172 TwinsUK samples were assayed using the Illumina 27k methylation kit, including 33 pairs of MZ twins, 43 pairs of dizygotic twins and 20 unrelated individuals. For the 33 pairs of MZ twins, we tested the Spearman correlation of the difference in methylation between each sib-pair to the difference in phenotype for total forearm (FA), total hip (TH) or total spine (SP) BMD. The phenotypes were standardized, adjusting for age, sex, weight (for total hip only), chip id, chip position, and chip batch. The phenotype value closest to the DNA extraction date was used. The significance threshold for the correlation test was set to $1\times10{-}4$. **Results:** Of the 33 pairs of MZ twins, 23 pairs had phenotype values for total forearm, hip and spine BMD. The methylation difference of a total of 7 genes had significant Spearman correlation ($p < 1\times10{-}4$) with differences in BMD phenotypes; 3 for TH (NTRK3, VSTM1 and ZNF438), 3 for FA (PIGC, PKP1, CBG), and 2 for SP (CBG and SLC6A11). The most significant correlation was obtained for corticosteroid-binding globulin (CBG) using the total forearm BMD phenotype (p-value $5.46\times10\text{-}6$), and was also significant, albeit to a lesser extent, for total spine BMD (p-value $1.67\times10\text{-}5$). The correlation for CGB is highly linear and positively correlated, where a positive difference in methylation leads to a positive difference in BMD phenotype between MZ sib-pairs. **Discussion:** This epigenome wide study has identified 7 putative loci correlated with BMD phenotypes. Among these loci is corticosteroid-binding globulin, a gene known to influence corticosteroid hormone levels, which in turn strongly influence bone physiology. Further validation is underway shortly, including replication of findings using a larger cohort of twins and unrelated individuals using the Illumina 450k methylation kit.

506T

Analysis of the epigenetic interplay between DNA methylation, histone modifications and gene expression in monozygotic twins discordant for psoriasis identified dysregulated disease-associated genes. *R. Lyle[1,2], K. Gervin[1,2], G.D. Gilfillan[1,2], M. Hammerø[1], H.S. Hjorthaug[1], A.O. Olsen[2,3], T. Hughes[1], J.R. Harris[4], D.E. Undlien[1,2].* 1) Med Gen, Oslo Univ Hosp Ulleval, Oslo, Norway; 2) University of Oslo, Oslo, Norway; 3) Department of Dermatology, Oslo University Hospital, Oslo, Norway; 4) Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway.

Psoriasis is a common, chronic inflammatory immune-mediated disease, which affects mainly the skin, but also the joints. The worldwide prevalence is reported to range between 1–11.8% depending on ethnicity and geographical area. Psoriasis has a strong genetic component with an estimated heritability of 66%. However, additional genetic, environmental and/or epigenetic factors are thought to be important, as concordance rates among MZ twins are only 35–72%. There is an increasing belief that epigenetic variants could explain some of the missing heritability in psoriasis. The aim of this study is to identify epigenetically dysregulated genes which contribute to the development of psoriasis. We performed comprehensive high-throughput bisulfite sequencing (RRBS), ChIP-seq and RNA-seq in CD4+ cells isolated from 20 MZ twin pairs discordant for psoriasis and explored DNA methylation at ~2 million CpG sites, histone modifications (H3K4me3 and H3K27me3) and gene expression, respectively. This approach enables an integrated analysis of the epigenetic interplay between DNA methylation, histone modifications and gene expression and identification of disease-associated epigenetic patterns and dysregulated genes. Preliminary findings identify potential susceptibility genes and point towards pathways containing epigenetically dysregulated genes in psoriasis. To our knowledge, this is the first study that attempts to systematically integrate DNA methylation, histone modifications and gene expression in MZ twins discordant for psoriasis in order to reveal and understand the epigenetic component in this disease.

507F

Genome wide DNA methylation profiling of monozygotic twins discordant for trisomy 21. *M.R SAILANI[1,2], F. Santoni[1], A. Letourneau[1,2], P. Makrythanasis[1], C. Borel[1], M. Guipponi[1], C. Gehrig[1], A. Vannier[1], S.E Antonarakis[1,2,3].* 1) Genetic MEDICINE and DEVELOPMENT, University of Geneva, Geneva, Geneva, Switzerland; 2) National Center of Competence in Research Frontiers in Genetics, University of Geneva, Switzerland; 3) iGE3 institute of Genetics and Genomics of Geneva, Switzerland.

DNA methylation is essential in mammalian development and has an effect on gene expression. We have hypothesized that methylation differences induced by trisomy 21 (T21) contribute to the phenotypic characteristics and heterogeneity in T21. In order to determine the methylation differences in T21 without the interference of the genomic variation, we have used fetal fibroblasts from monozygotic twins discordant for T21. We also used fibroblasts from monozygotic twins concordant for T21, normal monozygotic twins without T21, and unrelated normal and T21 individuals as controls. We applied Reduced Representation Bisulfite Sequencing (RRBS) to generate nucleotide resolution of DNA methylation based on high throughput sequencing (HiSeq 2X-100bp, One sample per lane) between each pair of twins. CpGs with at least 10X read coverage in two technical replicates were selected for the subsequent analysis. An initial analysis of MZ twins discordant for T21 identified 28,526 differentially methylated C nucleotides out of 1,589,507 interrogated nucleotides (DMCs) (FDR<0.001 and at least 50% methylation difference per C). The KEGG gene ontology analysis of genes harboring these DMCs, showed an enrichment for calcium signaling (FDR 3.1×10^{-7}), axon guidance (FDR 7.4×10^{-7}), and focal adhesion (FDR 5.1×10^{-5}) pathways. This preliminary study of methylation differences in monozygotic twins discordant for genomic abnormalities is a promising approach to understand the molecular pathophysiology of aneuploidies.

508T

DNA methylation analysis of the human X chromosome in multiple tissues. *A. Cotton[1,2], M. Jones[1,3], C. Chen[3,4], W. Wasserman[1,3], M. Kobor[1,3], C. Brown[1,2].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Molecular Epigenetics Group, Life Sciences Institute, Vancouver, BC, Canada; 3) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 4) Graduate Program in Bioinformatics, University of British Columbia, Vancouver, BC, Canada.

46, XX human females undergo X-chromosome inactivation (XCI) to achieve dosage compensation with 46, XY males. While the majority of X-linked genes are subject to XCI, over 15% of genes escape from XCI and are expressed from both the active X chromosome and the mostly silent inactive X chromosome. Previous studies have demonstrated that DNA methylation (DNAm) can be used to predict the XCI status of genes with CpG island promoters. Of the nearly 1800 X-linked islands on the Illumina Infinium HumanMethylation450 array, over half overlap a transcription start site and can be used to predict the XCI status of a gene. We examined X-linked DNAm in over 700 samples from human buccal cells, blood, brain and lung which allowed assessment of 331 genes for which no XCI status was previously known. The examination of over 50 individuals with matched tissue samples provided evidence for tissue-specific XCI. In agreement with previous reports, a subset of genes without island promoters demonstrated DNAm typical of an island associated with an XCI status. The sequence of subject genes was compared to escape genes to determine how sequence might influence the spread of XCI . A third of probes were located more than 1kb from a transcription start site but were used to examine DNAm within and between genes. Gene body DNAm levels were examined comparing genes features such as exons and introns in an effort to further our understanding on the complex relationship between DNAm and gene expression. Of the over 1500 probes found between genes, nearly 1000 were found in islands not associated with a gene. The vast majority of these islands demonstrated DNAm levels similar to that of subject genes. These non-promoter islands were divided into those enhancers predicted from the ENCODE projects to determine if DNAm associated with X-linked enhancers was unique. Non-island CpGs, including those in repetitive elements, were found to have significantly more DNAm in males compared to females suggesting hypermethylation on the active X. Many diseases occur at different rates in males and females. For most X-linked genes, XCI achieves dosage compensation between males and females however when a gene escapes from XCI females theoretically have a higher gene dose which may in turn contribute to difference in disease susceptibility. In addition to importance in modulating disease, the study of XCI provides new insights into epigenetics gene regulation and nuclear compartmentalization.

509F

Toward understanding the higher-order chromatin organization at the human inactive X chromosome. *E. Darrow, B. Chadwick.* Biological Science, Florida State University, Tallahassee, FL.

X-chromosome inactivation is the mammalian form of dosage compensation that balances X-linked gene expression between the sexes. This is achieved by repackaging the chosen inactive X chromosome (Xi) into facultative heterochromatin that ultimately shuts down most gene expression from the chromosome. The Xi is organized into two, non-overlapping types of heterochromatin, that at metaphase occupy multiple alternating chromatin bands, giving rise to a characteristic striped appearance. At interphase, heterochromatin of the same type aggregates together resulting in two distinct heterochromatin territories. Euchromatic marks are largely absent from the Xi with the exception of several specific regions that at metaphase reside at the intersection between the two heterochromatin types. We have determined the identity of the underlying DNA sequences for three of these euchromatin signals, and each consists of large tandem-repeat DNA, that are Xi-specific DNaseI hypersensitive sites and are bound by the epigenetic organizer protein, CCCTC-binding factor (CTCF). At interphase, these DNA sequences make Xi-specific, very long-range interactions that are significantly reduced when CTCF levels are depleted. Considering their location at the heterochromatin boundary and their frequent multi-megabase interactions, we propose that these repeats represent folding elements that contribute to the formation and maintenance of the Xi chromosome territory. To assess their role as chromosome folding elements we have developed active pairs of transcription activator-like effector nucleases (TALENs) that are capable of directly excising the two largest tandem repeats. These tools are being used to investigate whether these tandem repeats contribute to the three-dimensional conformation and maintenance of chromosome territories on the Xi.

510T

The p.V37I Exclusive Genotype Of GJB2: A Genetic Risk-Indicator of Postnatal Permanent Childhood Hearing Impairment. *l. li[1,2], J. Lu[1,2], Z. Tao[2,3], Q. Huang[1,2], Y. Chai[1,2], X. Li[1,2], Z. Huang[1,2], Y. Li[1,2], M. Xiang[1,2], J. Yang[1,2], G. Yao[4], Y. Wang[4], T. Yang[1,2], H. Wu[1,2].* 1) Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; 2) Ear Institute, Shanghai Jiaotong University, Shanghai, China; 3) Shanghai Children's Medical Center, Shanghai, China; 4) Shanghai Child Health Care Institute, Shanghai, China.

Background: Postnatal permanent childhood hearing impairment (PCHI) is frequent (0.25‰–0.99‰) and difficult to detect in the early stage, which may impede the speech, language and cognitive development of affected children. Genetic tests of common variants associated with postnatal PCHI in newborns may provide an efficient way to identify those at risk. We hypothesized that the p.V37I exclusive genotype of GJB2, a prevalent genetic variant in East Asians, is associated with postnatal PCHI and therefore serves as a genetic risk-indicator. Methods and Findings: We sequenced the GJB2 gene in a cohort of 45 Chinese subjects with postnatal PCHI, 1516 ethnically-matched control newborns with various newborn hearing screen (NHS) results, and 173 newborns who failed the NHS but passed the referral hearing diagnosis. The p.V37I exclusive genotype of GJB2 is present in a substantial percentage (20%) and is strongly associated with postnatal PCHI ($P = 1.4\times10{-}10$). The prevalence of this genotype also increases significantly in normal-hearing newborns who failed at least one NHS ($P = 0.024$ for those who failed only the initial NHS; $P = 1.7\times10{-}8$ for those who failed both the initial and the repeat NHS). Conclusion: The p.V37I exclusive genotype of GJB2 causes subclinical hearing impairment at birth and increases risk for postnatal PCHI. Genetic testing of GJB2 in East Asian newborns who fail at least one NHS will facilitate prompt detection and intervention of postnatal PCHI.

511F

Up-Regulates of angiogenesis-associated MicroRNAs in placenta from women with a low flow-mediated vasodilation. *L.M. RODRIGUEZ, A.F. DUQUE, A. SANCHEZ, A.C. AGUILAR, A. CASTILLO.* UNIVERSIDAD DEL VALLE, CALLE 4B 36-00, CALI - COLOMBIA.

Purpose: microRNAs profiling have emerged as a molecular tool to understand epigenetic regulation process and a possible impact in diseases prognosis or complications, such as in future mother at gestational period, during which endothelial function changes may be involved in differential expression profile of microRNAs associated with angiogenesis in placental vascularization. Methods: we evaluated FMD (flow-mediated vasodilation) during the first and third trimester of pregnancy in eleven primigravidae singletons as a vascular functional biomarker. Besides, at the delivery we took a placental sample to evaluate the microRNA expression associated with angiogenic pathway using miScript miRNA PCR array (Qiagen). Mann-Whitney and t-Student test, with p-value <0.05 were use to statistical analysis. Results: we found four upregulated miRNAs: hsa-miR-16 (p-valor: 0.044941), hsa-miR-130a (p-valor: 0.037793), hsa-miR-132 (p-valor: 0.045386) and hsa-miR-296 (p-valor: 0.043906) at women group that showed a negative change in FMD. Three of them (hsa-miR-130a, 132 and 296) are pro-angiogenic with a possible implication in a placental blood vessel increasing production and a decrease blood flow compensatory mechanism. Conclusion: Thereby, the impact in the microRNAs identification in the maternal-fetal health is important to prevent possible problem associated to placental vascularization.

512W

ChIP-seq ascertained occupancy of MEF2C, a GWAS-implicated osteoporosis locus, points to an inflammation-mediated role in bone density determination. *S.F.A. Grant[1,2], S. Deliard[1], F. Zhu[3], Q. Xia[1], A.D. Wells[4], K.D. Hankenson[3], M.E. Johnson[1].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA; 4) Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA.

Genome-wide association studies (GWAS) have demonstrated that genetic variation at the *MEF2C* (MADS box transcription enhancer factor 2, polypeptide C) locus is robustly associated with bone mineral density (BMD), primarily at the femoral neck. MEF2C is a transcription factor known to operate via the Wnt signaling pathway. Our hypothesis was that MEF2C regulates the expression of a set of molecular pathways critical to skeletal function. Drawing on our laboratory and bioinformatic experience with ChIP-seq, we analyzed ChIP-seq data for MEF2C available via the ENCODE project to gain insight in to its global genomic binding pattern. We aligned the ChIP-seq data generated for GM12878 (an established lymphoblastoid cell line). Using HOMER, a total of 17,611 binding sites corresponding to 8,118 known genes were observed. We then performed a pathway analysis of the gene list using Ingenuity. At 5kb, the gene list yielded 'EIF2 Signaling' as the most significant annotation, with a P-value of 5.01×10^{-26}. Moving further out, this category remained the top pathway at 50kb and 100kb, then dropped to just second place at 500kb and beyond by 'Molecular Mechanisms of Cancer'. Also consistently high on these lists at all distances was 'B Cell Receptor Signaling' and 'T Cell Receptor Signaling', plus at 50kb and beyond 'RANK Signaling in Osteoclasts' was a consistent feature and resonates with the main general finding from GWAS of bone density. We also observed that MEF2C binding sites were significantly enriched primarily near inflammation associated genes identified from GWAS; indeed, a similar enrichment for inflammation genes has been reported previously using a similar approach for the vitamin D receptor, an established key regulator of bone turnover. These results represent the first ever reported ChIP-seq derived genome wide map of MEF2C binding. Our analyses point to known connective tissue and skeletal processes but also provide novel insights in to networks involved in skeletal regulation. The fact that a specific GWAS category is enriched points to a possible role of inflammation through which it impacts BMD.

513T

Enhancer activity patterns across 100 epigenomes reveal novel human disease associations. *W. Meuleman, L.D. Ward, A. Kundaje, M. Kellis, the Roadmap Epigenomics Program.* MIT, Cambridge, MA.

The NIH Roadmap Epigenomics Program has generated a large resource of epigenomic marks including histone modification patterns in both primary human tissues and human cell lines, with the goal of creating global reference maps of regulatory elements and study their biological roles. We have used these datasets to generate chromatin state maps by learning combinations of histone modification patterns indicative of different functional classes. The chromatin state annotations reveal ~500,000 active and poised enhancer regions, and ~120,000 active and poised promoter regions across cell types, as well as strongly and weakly transcribed regions, repressed regions, and heterochromatic regions. We use these epigenomic maps to cluster regulatory regions into 55 enhancer modules and 70 promoter modules of coordinated activity across cell types. We find that the vast majority of enhancers and promoters are cell type restricted, and highly enriched in developmental processes. Surprisingly, only a small percentage of promoter regions are constitutively active, suggesting a higher similarity between enhancer and promoter regions than previously recognized. The identified enhancer and promoter modules were found to be associated with distinct GO categories among neighboring genes, which reflected their cell-restricted activity patterns. Because these activity patterns provide information about developmental dynamics beyond single-tissue annotations, we asked whether they provided additional insights into the noncoding genetic architecture of diseases and the strength of selective pressures on different regulatory programs. We first tested SNPs from the NHGRI GWAS catalog against individual cells' enhancer maps, revealing enhancer enrichments with a high degree of tissue specificity, such as adult liver enhancer SNPs associated with total cholesterol and adipose and pancreatic islet enhancer SNPs associated with fasting glucose levels. Interestingly, when we instead used enhancers clustered by activity, in some cases GWAS SNP enrichments were discovered for activity signatures and not for individual tissues; for example, coronary heart disease associated SNPs are enriched for a type of enhancer active in many non-immune tissues and enriched for proximal gene GO annotations of substrate adhesion-dependent cell spreading, apoptotic cell clearance, and response to fluid shear stress.

514F

Human-specific cytogenetic structures support the emergence of new regulatory elements. *G. Giannuzzi, A. Reymond.* Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

Human and chimpanzee genomes are 98.8% identical within comparable sequence. They however differ structurally in nine pericentric inversions, one fusion that originated human chromosome 2 (HSA2), content of their heterochromatin, and localization of segmental duplications. The possible functional consequences of these cytogenetic and structural differences are not fully understood and their possible involvement in speciation remains unclear. We analyzed the genome-wide distribution of histone modifications that decorate transcription start sites with human-specific enrichment compared to chimpanzee and macaque. We show that these histone modifications are not randomly distributed in the human genome, indeed they preferentially map to genomic regions with a human-specific organization, like subtelomeric, pericentromeric, and segmentally duplicated areas. Likewise, loci exclusive to the human lineage, such as the HSA2 fusion point and its ancestral pericentromeric region and the breakpoints of the HSA1 and HSA18 pericentromeric inversions, significantly accumulated human-specific transcription start sites. We find a similar enrichment of chimpanzee-specific H3K4me3 peaks in subtelomeric and segmentally duplicated regions of the chimpanzee genome, revealing a common propensity of new cytogenetic structure in acquiring potential novel transcription start sites. These concentrations of species-specific transcription start sites in genome structure unique to one species are not limited to a single tissue as we uncovered them in prefrontal cortex, neuronal cells, and lymphoblastoid cell line. Our findings support the existence of an evolutionary role of chromosomal rearrangements and allow inferring causality, i.e. chromatin reconfiguration follows genome reorganization. They suggest that evolutionary novelties derived from structural changes should be investigated not only for novel transcripts and gene expression differences but also for epigenetic and regulatory changes.

515W

Architectural proteins modulate the higher order chromatin structure of the *CFTR* locus. *N. Gosalia[1,2], A. Harris[1,2].* 1) Human Molecular Genetics Program, Ann and Robert H. Lurie Children's Hospital of Chicago Research Center, Chicago, IL; 2) Department of Pediatrics, Northwestern University, Feinberg School of Medicine, Chicago, IL.

Cystic fibrosis is an autosomal, recessive disease that results from mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, which encodes a chloride ion channel with a complex expression pattern. Tissue-specific *CFTR* expression is regulated by multiple *cis*-acting elements, including intronic enhancers, which interact with the promoter. The mechanisms that organize higher order chromatin structure to establish and maintain gene expression are not fully understood; however, data suggest they may involve CTCF and cohesin. These factors have important roles in the three-dimensional organization of loci and at insulators, which are critical barriers for preventing inappropriate activation or repression of genes. At the *CFTR* locus, depletion of CTCF or the cohesin component, RAD21, increases gene expression. ChIP for CTCF and the cohesin components, RAD21 and SMC1, after CTCF or RAD21 knockdown shows loss of all three factors at several sites across the region and retained occupancy at others. To determine if loss of CTCF and/or RAD21 altered the higher order organization of *CFTR*, quantitative chromosome conformation capture (q3C) was used. After RAD21 depletion, q3C data show partial loss of interactions across the locus between the gene promoter, known enhancers such as the one in intron 11, and CTCF/cohesin binding sites. In contrast, after CTCF knockdown no changes in enhancer-promoter interactions are observed, however, interactions between the -20.9kb insulator element and CTCF/cohesin binding elements 3' to *CFTR* are suppressed. These data suggest that though cohesin and CTCF are critical for some aspects of the chromatin organization at the locus they are not the major components of the complex that loops distal enhancers to the promoter. Moreover, the increase in *CFTR* expression after knockdown also suggests that CTCF and cohesin may inhibit *CFTR* expression by recruiting repressive complexes to the locus. Identifying these repressive factors may facilitate therapies to alleviate cystic fibrosis disease severity by increasing the levels of *CFTR* transcript and CFTR protein.

516T

Impact of three-dimensional organization of chromatin on long-range enhancers. *A. Pankov, J. Song.* Institute for Human Genetics, UCSF, San Francisco, CA.

In order to fully understand how genes and regulatory regions are organized and coordinated for transcriptional regulation, it is not sufficient to only consider the genomic locations of regulatory sites. In fact, the three-dimensional chromatin interactions have been shown to regulate transcriptional and epigenetic states demonstrating that long-range chromatin interactions act as a mechanism to regulate many important genes. Currently, methods to identify such long-range relationships have been technically challenging. High-throughput approaches such as DNase-Seq, ChIP-Seq, and Hi-C methods have been previously used to study the influence of long-range enhancers on regulating transcription, but these either have low resolution for human samples (Hi-C) or lack the important connectivity information to investigate how regulatory elements interact with their distal target gene promoters (ChIP-seq and DNase-seq). Through the use of ENCODE Chromatin Interaction Analysis Paired-End Tag sequencing (ChIA-PET) data as a high-resolution, cost-effective alternative, we are able to determine robust distal interactions and analyze their role in transcriptional regulation. As part of this project, we first remove background noise present in the data from the complexity of chromatin structures in nuclear space and the nature of proximity ligation. This involves creating an appropriate model for read counts to reduce the influence of confounding factors in the data; identifying the genomic location of regulatory elements; and removing random interactions between two regions as ones that occur infrequently relative to the overall signal of the two regions. Then, by modeling the data as a graph with regulatory sites as nodes and number of interactions as edge weights, we are able to use graph partitioning techniques to find densely connected substructures known as interaction hubs. Finally, we annotate each interaction hub with transcriptional information and match each node with its regulatory function through other ENCODE assays. By labeling each node, we are able to analyze how different structural and regulatory motifs contribute to transcriptional regulation. Through our findings, we are able to determine the functionally and three-dimensionally important structures that regulate gene transcription.

517F

Genetic determinants of population-level variation of chromatin modifications. *H. Kilpinen[1,2], S.M. Waszak[3], A.R. Gschwind[2,4], S.K. Raghav[3], R.M. Witwicki[4], A. Orioli[4], M. Wiederkehr[4], N. Panousis[1,2], N. Hernandez[4], A. Reymond[4], B. Deplancke[3], E.T. Dermitzakis[1,2].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology, Lausanne, Switzerland; 4) Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland.

The study of gene regulation through intermediate cellular phenotypes has lead to the identification of genomic loci that associate with quantitative changes in gene expression levels in human populations. While such regulatory variation is extremely widespread, it has been challenging to pinpoint the exact molecular processes that are affected by these variants before differential gene expression is manifested. To address this, we mapped quantitative trait loci (QTL) for genome-wide RNA polymerase II and PU.1 binding, as well as levels of histone modifications H3K4me1, H3K4me3, and H3K27ac in 52 unrelated individuals from the 1000 Genomes project. Chromatin immunoprecipitations of all assays were produced from a single growth of lymphoblastoid cell lines and sequenced to a high coverage (median > 48M high quality reads per individual). We identified between 591 and 2096 independent peaks with a cis-QTL for each assay at a false discovery rate of <5% (50 kb window centered on each peak), representing 2-4% of peaks genome-wide. Extending the window to 100 kb resulted in ~16% increase in the number of detected QTL peaks. The majority of variants affected only a single peak, and we observed on average 3-4-fold enrichment of the identified QTL at known expression QTL, as well as significant overlap of QTL among the different chromatin assays. We are currently exploring the biological properties and interactions of the identified chromatin QTL with the aim of understanding the regulatory mechanisms through which differential gene expression is achieved, as well as the proportion of gene expression variation in the population that can be explained by quantitative changes in chromatin and transcription factor binding.

518W

Molecular analyses of *cis*-interactions of the *SOX9* promoter with its sex determining long-range regulators using chromatin conformation capture-on-chip (4C). *P. Stankiewicz[1,2], M. Smyk[2], P. Szafranski[1].* 1) Dept. of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept. of Medical Genetics, Institute of Mother and Child, Warsaw, Poland.

Evolutionary conserved transcription factor *SOX9* is essential for development of testes and differentiation of chondrocytes. Heterozygous point mutations and genomic deletions involving *SOX9* lead to the skeletal malformation campomelic dysplasia (CD) often associated with male-to-female sex reversal. Chromosomal rearrangements with breakpoints mapping in the protein-gene desert regions up to 1.3 Mb both 5' and 3' to *SOX9* and likely disrupting *cis*-regulatory elements, have been described in patients with usually milder forms of CD. Breakpoint clustering 5̀ to *SOX9* allowed to define four non-overlapping intervals associated with different phenotypes: Pierre Robin sequence (~1.06-1.23 Mb), acampomelic CD (~789-932 kb), sex reversal (~517-595 kb), and moderate to severe CD (~50-375 kb). In addition, a different phenotype consistent with Cooks syndrome (brachydactyly-anonychia) has been reported as associated with genomic duplications ~1.2 Mb 5' to *SOX9*. We found that these regions overlap four gene clusters encoding long non-coding RNAs (lncRNAs): TCONS_00025479, TCONS_00025148, TCONS_00025195, and TCONS_00026151, suggesting that similar to other developmental genes, lncRNAs might contribute to long-range regulation of *SOX9* expression. The sex reversal region (RevSex) was proposed to harbor a putative testis-specific and sex determining enhancer. The other sex determining interval was mapped to a gene desert >1.3 Mb downstream to *SOX9*. We performed chromatin conformation capture-on-chip (4C) analysis in Sertoli cells and lymphoblasts to verify the already proposed long-range interactions and to identify potential novel regulatory elements responsible for sex reversal in patients with CD. We identified several novel *cis*-interacting regions both up- and downstream to *SOX9* with some of them overlapping lncRNAs preferentially expressed in testes. Our data further support the role of the two previously proposed regulatory regions in the control of *SOX9* expression during sex determination and point at lncRNAs as likely mediators of some of these interactions.

519T

Identification of genetic variants that affect histone modifications in human cells. *G. McVicker[1,2], B. van de Geijn[1,3], J.F. Degner[1,3], C.E. Cain[1], N.E. Banovich[1], N. Lewellen[2], M. Myrthil[2], Y. Gilad[1], J.K. Pritchard[1,2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Howard Hughes Medical Institute, University of Chicago, Chicago, IL; 3) Committee on Genetics, Genomics and Systems Biology, University of Chicago, Chicago, IL.

Histone modifications are important markers of function and chromatin state, yet the DNA elements that direct them to specific locations in the genome are poorly understood. Here we use the genetic variation in Yoruba lymphoblastoid cell lines as a natural experiment to identify genetic differences that affect histone marks and to better understand their relationship with transcriptional regulation. Across the genome, we identified hundreds of quantitative trait loci that impact histone modification or RNA polymerase (PolII) occupancy. In many cases the same variant is associated with quantitative changes in multiple histone marks and PolII, as well as in DNaseI sensitivity and nucleosome positioning, indicating that these molecular phenotypes often share a single underlying genetic cause. We find that polymorphisms in many transcription factor binding sites cause differences in local histone modification and identify specific transcription factors whose binding leads to histone modification in lymphoblastoid cells. Finally, we find that variants that impact chromatin at distal regulatory sites frequently also direct changes in chromatin and gene expression at associated promoters. In summary, the class of variants identified here generate coordinated changes in chromatin both locally and sometimes at distant locations, frequently drive changes in gene expression, and likely play an important role in the genetics of complex traits.

520F

Lethal lung developmental disorder due to a novel deep intronic deletion. *P. Szafranski[1], Y. Yang[1], M.U. Nelson[2], M.J. Bizzarro[2], R. Morotti[3], C. Langston[4], P. Stankiewicz[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dept. of Pediatrics; 3) Dept. of Pathology, Yale University School of Medicine, New Haven, CT; 4) Dept. of Pathology, Texas Children's Hospital, Houston, TX.

FOXF1 encodes a forkhead transcription factor mediating hedgehog signaling in tissues of mesodermal origin, primarily in endothelia and mesenchyme of the developing lungs and intestines. Haploinsufficiency of *FOXF1* causes an autosomal dominant neonatally lethal lung disorder, Alveolar Capillary Dysplasia with Misalignment of Pulmonary Veins (ACDMPV). We identified a 0.8 kb deletion within the single 1.4 kb intron of *FOXF1* in a deceased newborn with histopathologically verified ACDMPV. The deletion arose *de novo* on the maternal copy of chromosome 16q, supporting the notion that *FOXF1* is partially paternally imprinted in human lungs. It did not compromise the major splice sites and *FOXF1* splicing pattern was only moderately affected. However, the deletion reduced the level of *FOXF1* mRNA in the peripheral lung tissue by almost 40%. We found that the *FOXF1* intron exhibits transcriptional enhancer activity, and the deletion removed the major binding sites for CTCF and CEBPB, the key regulators of chromatin architecture and transcription. Interestingly, our *in vitro* reporter assay analyses suggest that the deletion not only disabled the enhancer but possibly transformed it into a transcriptional suppressor. Currently used diagnostic assays such as chromosomal microarray analysis or whole-exome sequencing do not detect deep intronic copy-number variation or simple nucleotide variation, thus underestimating their pathogenic significance. Our data further emphasize the importance of testing of the non-protein coding regions in the human genome.

521W

Rare Recurrent Homozygous Deletions in a Large Population Reveals High Penetrance Loss of Function Variants Underlying Rare Strong Phenotypes. *J. Glessner[1,2], J. Li[1], B. Chang[1], K. Thomas[1], R. Golhar[1], X. Chang[1], Y. Guo[1], D. Li[1], N. Abdel-Magid[1], C. Kao[1], Y. Li[1], J. Bradfield[1], C. Kim[1], C. Hou[1], F. Mentch[1], H. Qiu[1], C. Cardinale[1], M. Garris[1], L. Vazquez[1], D. Abrams[1], G. Otieno[1], T. Wang[1], R. Pellegrino da Silva[1], R. Chiavacci[1], J. Connolly[1], B. Keating[1,2], P. Sleiman[1], S. Grant[1,2], H. Hakonarson[1,2].* 1) The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 2) Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA.

We analyzed Illumina SNP microarray data for 68,028 high quality uniformly processed samples and called homozygous deletions (HZDs) using PennCNV with an intersection set of 520,017 SNPs. 263 homozygous deletion CNVRs were detected in multiple unrelated individuals. 146 directly overlapped gene exons. 57 show significant enrichment for a disease state. 631 HZDs were specific to one patient, of which 414 overlapped exons. Chip version, ethnicity population/evolutionary genetics, and sample sources (projects) were all data complexities explored for further insight into the genesis of the HZDs and to control for bias. HZDs are the most confident of all copy number states due to the strong drop in relative intensity (log r ratio) coupled to a uniformly random genotype (b allele frequency). Using a multi-normal curve fitting approach, all HZDs were confirmed. This high confidence calling and population based CNV quality assessment allows for a very small number of SNPs and length lower detection limit. HZDs closely parallel the extensive work in gene knockout studies to study phenotypes and allow for recapitulation of the phenotype in humans by naturally arising variations observed in our study. From a clinical perspective, HZDs provide the strongest possible genetic variant in complete ablation of a gene's function and may help to diagnose rare constellations of symptoms for early intervention. For example, we uncovered a HZD overlapping NF1 exclusively in an individual with multiple clinical features suggesting neurofibromatosis without the formal clinical diagnosis until now. Another child had HZD in PARK2, currently asymptomatic. By further evaluating deletions as an allelic genotype (CN=0->1 1, CN=1->1 2, other->2 2), we were able to detect deletions significantly deviating from Hardy-Weinberg equilibrium by lack of heterozygous deletions suggesting embryonic lethality of the heterozygous deletion variant a specific loci. ParseCNV and Plink were used for population analysis and association of the CNV calls. This large scale association of homozygous deletions to clinical phenotypes represents a major advancement in leveraging genomic variants in direct clinical application and my guide direction of effects for future therapeutic interventions.

522T

Identification of copy number variants (CNVs) using computational algorithms and array CGH technologies. *P.S. Samarakoon[1,2], H.S. Sorte[1,2], B.E. Kristiansen[2], T. Skodje[2], A. Stray-Pedersen[2], O.K. Rodningen[2], R. Lyle[2].* 1) Department of Medical Genetics, University of Oslo, Oslo, Norway; 2) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.

With advances in next generation sequencing technologies and genomic capture techniques, exome sequencing has become a cost effective approach for mutation detection in genetic diseases. However, computational CNV prediction algorithms using exome sequence data exhibit limitations due to the low sensitivity and specificity in predicting small CNVs (covering 1-4 exons). Additionally, Comparative Genomic Hybridization based microarrays (aCGH) used for genome wide high-resolution CNV detection also show restrictions in exonic CNV detection due to the low scoring probes over GC rich regions. Therefore, the goal of the research project was to develop a protocol to detect exonic CNVs (including 1-4 exons), combining computational algorithms and custom aCGH. In this study, we predicted CNVs of 30 exomes obtained from 1000 genomes project using seven computational programs, including an in-house developed algorithm. CNV calling of the proposed algorithm relies on GC% based read depth normalization followed by generation of the best-matched reference dataset for target exome. In parallel to the computational prediction, CNVs were also identified by using aCGH, which was designed to capture exonic regions in 1000 genomes exome and Gencode v.15. Next, true CNVs were identified by comparing computational predictions to the aCGH results. Results of each computational program showed a significant variation in predicted CNV counts (ranging 0-2500). Due to the drastic variation in resulted CNV counts, intersection operations were performed to identify CNVs predicted by multiple programs. As intersection operation showed a direct effect in decreasing the number of predicted CNVs to 0-250 from 0-2500, this was used as a filter for our CNV pipeline. Finally, CNVs resulted from complete protocol (true CNVs), which were not reported in database of genomic variants (DGV) were identified as novel CNVs. Application of the complete protocol on 9 individuals presented 77 novel CNVs including 26 that cover single exons. Additionally, protocol was then implemented on primary immunodeficiencies (PIDDs) patients at Oslo University Hospital and predicted PIDD causing CNVs from the computational approach were validated using the custom array. Thus, identifying PIDD causing CNVs while demonstrating the capability in detecting shorter exonic CNVs proved clinical importance of the proposed protocol.

523F

Genomic characterization and copy number analysis of human induced pluripotent stem cells (iPSCs) from 22q11.2 deletion syndrome patients. *C. Purmann[1,2], S. Pasca[3], H. Gai[3], A. Krawisz[3], X. Zhu[1,2], J. Rapoport[4], J. Berstein[5], J. Hallmayer[1], R. Dolmetsch[3], A.E. Urban[1,2].* 1) Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Neurobiology, Stanford University, Stanford, CA; 4) National Institutes of Health, Child Psychiatry Branch, NIMH, Bethesda, MD; 5) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA.

The 22q11.2 deletion syndrome, also known as DiGeorge Syndrome or Velocardiofacial Syndrome (VCFS), is a common genomic disorder in humans. The clinical phenotype is variable, but includes most notably neurodevelopmental abnormalities, craniofacial and cardiovascular malformations, as well as immune deficiencies. It is thought that large CNVs present critical aberrations during early brain development that manifest in childhood or later in life. The lack of understanding how exactly these CNVs exert their influence on a molecular level is exacerbated in neurogenetic diseases where it is difficult to access the relevant human tissues for detailed molecular studies. One very promising avenue around this obstacle comes from the recent development of induced pluripotent stem cells (iPSCs) whereby skin cells from patients can be reprogrammed into pluripotent stem cells which can then be differentiated into neuronal cells for controlled molecular study. Here, we report the genomic characterization of a panel of iPSCs from seven 22q11.2 deletion patients and seven matched controls. Multiple iPSC lines were created from each proband, and between one and three lines per proband were analyzed using SNP arrays, for a total of 26 iPSC lines analyzed. We used the new Illumina HumanOmni5Exome array which interrogates more than 4.5 million loci genome-wide. The resulting data can be used to detect duplications and deletions with high accuracy. For each iPSC line derived from a patient, we determined the exact extent of the main deletion. We also determined for all lines the overall complement of CNVs and SNPs, including in a small subset of the lines a few additional, smaller CNVs that would have been missed with karyotype analysis, but that are large enough that they should be taken into consideration while using the iPSCs as a model system. The iPSCs are showing full differentiation potential along the neuronal trajectory. For a subset of the iPSCs, we used RNA-Seq analysis and detected gene expression changes in a large number of the genes within the deletion region. These iPSC lines will be a highly valuable resource for the analysis of this important microdeletion syndrome, as well as in general for the elucidation of molecular effects of large copy number aberrations on the genomic control of cellular differentiation and functioning. Furthermore our analysis highlights that high-resolution genome analysis in iPSC based model systems should be standard practice.

524W

Common CNVs of Innate Immune Response Genes Defensin DEFA1/A3 And Complement C4A Are Medium Effect-Size Risk Factors for Human Autism Spectrum Disorders (ASD). *H. Wang[1,2], K. Linter[1], E. Hansen[1], B. Zhou[1], YL. Wu[1], K. Jones[1], A. Schwaderer[2], D. Hains[2], CY. Yu[1], G. Herman[1].* 1) Center for Molecular and Human Genetics the Research Institute at Nationwide Children's Hospital and Department of Pediatrics, The Ohio State University, Columbus, Ohio 43205; 2) Center for Translation Research the Research Institute at Nationwide Children's Hospital and Department of Pediatrics, The Ohio State University, Columbus, Ohio 43205.

Products of innate immune response genes are not only crucial in the defense against infections, but are also needed in nerve pruning during the formation of synapses in the central nervous system. Many innate immune response genes are characterized by frequent, inter-individual gene copy number variations (common CNVs). The role of common CNVs for immune-related genes in the pathogenesis of ASD is unknown. Our objective is to determine common CNVs of innate immune response genes for antimicrobial peptides α-defensins DEFA1 and DEFA3, and complement C4A and C4B in genetic risks of ASD. Our study population included 185 ASD patients (90.3% White, 9.7% other races) and >500 race-matched healthy controls from Ohio. Of the White ASD patients, 84% were male and 16% were female; the mean age (±SD) at disease diagnosis was 3.3±1.8 years old. Continuous CNVs for defensins DEFA1/A3 and complement C4 were determined by TaqI genomic restriction fragment length polymorphisms (RFLPs) for all control samples. The relative quantities of DEFA1 and DEFA3 were interrogated by 'hot-stop' PCR and HaeIII-RFLP, and those of C4A and C4B were elucidated by PshAI-PvuII RFLP. The CNV-defined samples facilitated development of sensitive TaqMan-based qPCR amplicons to decipher CNVs for DEFA1/A3 and C4A/C4B in ASD. The copy number for DEFA1/A3 in a diploid genome of White subjects varies from 4 to >15; and for C4, 2 to >6. The mean copy number (±SD) of DEFA1/A3 in ASD is 6.37±1.24, compared to 6.85±1.81 in controls (p=0.0013). On average, ASD patients have 0.5 copy of DEFA1/A3 less than healthy subjects. In particular, there is a marked decrease in the frequency of the high copy-number groups of (≥9 copies) in ASD (4.2% versus 13.0%; p=0.0006). The odds ratio (95% CI) for low copy number of DEFA1/A3 in ASD is 3.39 (1.52-7.56). DEFA1 and DEFA3 proteins differ by the A65D polymorphism. The absence of D65 (i.e., homozygous deficiency of DEFA3) has a frequency of 10.8% in ASD, compared to 18.1% in controls [OR=1.81 (1.02-3.21); p=0.035], suggesting the presence of DEFA3 is a risk factor for ASD. For complement C4, we observed a significantly lower copy number of C4A (ASD=1.93±0.74; controls= 2.10±0.76; p=0.008) but not C4B in ASD. The odds ratio for homozygous and heterozygous deficiency of C4A in ASD is 1.52 (1.02-2.29). In summary, common CNVs for defensins DEFA1/A3 and complement C4A are novel, medium effect-size genetic risk factors in White ASD.

525T

***TM4SF20* ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities.**
S.R. Lalani[1], W. Wiszniewski[1], J.V. Hunter[2], N.A. Hanchard[1], J.R. Willer[3], C. Shaw[1], S.W. Cheung[1], A. Patel[1], P. Hixson[1], C.A. Bacino[1], L. Potocki[1], R.L.P. Santos-Cortez[1], P.I. Bader[4], T.M. Morgan[5], S.A. Boyadjiev[6], T.E. Gallagher[7], A. Mutirangura[8], P. Stankiewicz[1], A.L. Beaudet[1], M. Maletic-Savatic[9], J.A. Rosenfeld[10], L.G. Shaffer[11], E.E. Davis[3], J.W. Belmont[1], C.C. Khor[12], S. Dunstan[13], P.E. Bonnen[1], S.M. Leal[1], N. Katsanis[3], J.R. Lupski[1]. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX; 3) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 4) Northeast Indiana Genetic Counseling Center, Fort Wayne, IN; 5) Division of Developmental Medicine, Vanderbilt University, Nashville, TN; 6) Section of Genetics, Department of Pediatrics, University of California Davis, Sacramento, CA; 7) Tripler Army Medical Center, Honolulu, HI; 8) Center of Excellence in Molecular Genetics of Cancer and Human Diseases, Chulalongkorn University, Bangkok, Thailand; 9) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Departments of Pediatrics and Neuroscience, Baylor College of Medicine, TX; 10) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 11) Paw Print Genetics, Genetic Veterinary Sciences, Inc., Spokane, WA; 12) Genome Institute of Singapore, Singapore; 13) The Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam.

White matter hyperintensities (WMH) of the brain are important markers of aging and small vessel disease. WMH are rare in healthy children and, when observed, often occur with comorbid neuroinflammatory or vasculitic processes. Utilizing 180k oligonucleotide-based, exon-focused array-comparative genomic hybridization in 15,493 children, we identified a 4-kb deletion co-segregating with WMH (penetrance ~70%) and early childhood language delay in multiple families, predominantly from Southeast Asia. The 15 unrelated carrier children, mostly of Vietnamese, Thai, Burmese, Filipino, Indonesian, and Micronesian descent were referred for the evaluation of early language delay, autism spectrum disorders, and/or brain imaging abnormalities. Formal speech and language assessment showed significant discrepancies between verbal and non-verbal skills in the deletion carriers. The premature brain aging phenotype with punctate to multifocal WMH was also observed in ~70% of young carrier parents. The complex 4-kb deletion in 2q36.3 removes the penultimate exon 3 of *TM4SF20*, a gene encoding a transmembrane protein of unknown function. Minigene analysis showed that the resultant net loss of an exon introduces a premature stop codon, which, in turn, leads to the generation of a stable protein that fails to target to the plasma membrane and accumulates in the cytoplasm. This deletion was found to be present in 46/2,018 (2.3%) umbilical cord blood samples from Vietnamese Kinh infants, indicating an allele frequency of about 1%, embedded in a common ancestral haplotype. Our study highlights the importance of understanding population-specific low-frequency highly penetrant alleles, and points to a likely toxic mechanism of TM4SF20 truncation that accounts for a strong effect on disease susceptibility related to familial early language delay and autism spectrum disorders in the Southeast Asian pediatric population.

526F

The use of MLPA-based strategy for discrete copy number genotyping of complex multi-allelic CNVs.
P. Kozlowski, M. Marcinkowska-Swojak. ECBiG, Poznan University of Technology, Poznan, Poland.

Copy number variation has recently been recognized as an important type of genetic variation that modifies human phenotypes. Copy number variants (CNVs) are being increasingly associated with various human phenotypes and diseases. However, the lack of an appropriate method that allows fast, inexpensive and, most importantly, accurate CNVs genotyping significantly hampers CNVs analysis. This limitation especially affects the analysis of multi-allelic CNVs that frequently modify various phenotypes. We present a multiplex ligation-dependent probe amplification (MLPA)-based strategy for multiplex and individual copy number genotyping of both simple bi-allelic and complex multi-allelic CNVs. We used this strategy for discrete genotyping of thee extensively studied CNVs, including: CNV-CCL3L1, CNV-DEFB and CNV-UGT2B17, which have been associated with risk of HIV infection, psoriasis and osteoporosis, respectively. Our experiments confirmed the high reproducibility and accuracy of the obtained genotyping results. Acknowledgements: NCN grant 2011/01/B/NZ5/02773.

527W

Co-evolutionary relationship between chemokines and chemokine receptor: An evolutionary perspective.
C. Tsui[1], Y. Hung[1], D. Liu[3], D. Chen[2]. 1) La Sierra University, Riverside, CA; 2) Dept Pathology, University of California, Irvine; 3) Center for Stem Cell and Translational Medicine, School of Life Sciences, Anhui University.

Chemokine are a family of small cytokines with the ability to induce chemotaxis in nearby responsive cells. The diversity of chemokine ligands and their receptors are generated via segmental duplication as evidence by genomic positions. Specifically, their genomic positions are tightly clustered within few genomic loci with evidence of segmental duplications. Here, we attempt to decipher the evolutionary history of chemokine receptors and their ligands by conduct cross species sequence analysis at the level of transcript sequences as well as the genomic sequences. Specifically, 27 annotated human chemokine receptor sequences as well as 35 annotated human chemokine ligand sequences are mapped to 24 available mammalian genomes. Reciprocal best hit alignments are obtained to generate a list of putative paralogous chemokine gene from each respective species. Each of these putative gene sequences are then used to determine potential orthologous gene within the respective genome. Using this method, we deduced generation of chemokine and its receptors diversity at various point in the mammalian evolution. In addition, there are clusters of chemokine receptors showing a reduction in the paralogous genes. The generation, as well as, the reduction of gene diversity likely hint the differences in evolutionary pressure each respective species face.

528T

High-resolution analysis of DNA copy number variations in Systemic Lupus Erythematosus patients.
F.B. Barbosa[1], M. Simioni[2], E.A. Donadi[3], V.L. Gil-da-Silva-Lopes[2], A.L. Simões[1]. 1) Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil; 2) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas (UNICAMP), Campinas, SP, Brazil; 3) Division of Clinical Immunology, Department of Medicine, School of Medicine of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil.

Advances in molecular-based techniques for DNA investigation enabled the detection of an important type of genomic variation named copy number variations (CNVs). CNVs are defined as genomic segments, usually greater than 1 kilobase (kb) in size, ranging in copy number when compared to a reference genome. They can contribute to risk variability among individuals in complex diseases etiology. Systemic Lupus Erythematosus (SLE) is an autoimmune disease with strong genetic component characterized by chronic inflammation and autoantibodies production. To date, several loci have been associated with SLE pathogenesis by genome-wide association studies (GWAS). However, there are few analyses about CNVs in SLE patients. The purpose of this study was to determine the role of CNVs in 23 SLE patients. To screen the CNVs, high-resolution array Genomic Hybridization Assay was performed using the Affymetrix® CytoScan™ HD platform. To calculate copy numbers, the data were normalized to baseline reference intensities using 366 samples (270 HapMap samples and 96 healthy normal individuals). Data was analyzed by *Chromosome Analysis Suite v.1.2.2* (ChAS) software, which includes the Hidden Markov Model (HMM) algorithm used to determine the copy number states. At total, 406 CNVs were identified (CNV average number per patient was 18), distributed across all chromosomes, except Y. Deletions were more frequent than duplications, 311 and 95, respectively. CNV profile showed 269 CNVs were overlapped by genes, 152 unique CNVs and 59 CNV regions (CNVRs). From all, 39 CNVs detected have not been described in the main database of healthy subjects, the Database of Genomic Variants (DGV). Nine of these have not been described in any structural variants databases. CNVs were found in ten genes previously related with autoimmunity: deletion-type CNVs in *STAT4*, *HLA-DPB2*, *CFHR4*, *CFHR5*, *SNTG1*, *IL3RA*, *UGT2B15*, *ADAM3A* and duplication-type CNVs in *MECP2* and *KIAA1267*. This is the first report of CNVs in these genes in SLE patients. The identification of CNV in SLE suggests a possible contribution of these variations to development of autoimmunity or the onset of the disease. This is consistent with the observed overlap between CNVs and genes implicated in the development of autoimmunity. The action of these genes in determining SLE or any other autoimmune disease should be investigated in future studies. Support: FAEPA, CAPES and FAPESP.

529F

Duplication-Normal-Duplication Rearrangements in Human Genomic Disorders. *C.R. Beck[1], P. Liu[1], J.R. Lupski[1,2,3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Inverted repeats are prevalent in the human genome, and are often implicated in genomic rearrangements such as inversions (see Human Molecular Genetics 2013 Mar 15 and Human Mutation 2013 Jan;34(1):210–20). Although copy number neutral events are plentiful between normal human genomes, inverted repeats may also play a role in the duplication of genomic locations that include dosage sensitive genes, leading to disease. Recently, duplication-triplication-duplication (DUP-TRP/INV-DUP) structures mediated by inverted low copy repeats (LCRs) have been described for both the *MeCP2* and the *PLP1* loci. Analysis of array data has also identified duplication-normal-duplication (DUP-NML-DUP) structures in multiple different patient cohorts. These structures can also be mediated by inverted repeat sequences, either through inversion and duplication or through complex mechanisms. To date, we have compiled 18 patients with DUP-NML-DUP rearrangements detectable by high-density array comparative genomic hybridization (aCGH). Discerning the breakpoint sequences and locations for a number of these patients has implicated both inverted LCRs and inverted *Alu* elements in the formation of these rearrangements. Interestingly, we have found that *Alu*-mediated inversion can precede tandem duplication in one patient family, and are investigating the frequencies of similar inversions in normal individuals. This implicates a potential two-step process in the generation of seemingly complex rearrangements, wherein a structural rearrangement that is copy number neutral can be present at least one generation before a non-recurrent tandem duplication leading to disease in a patient. This finding broadens the scope of the inverted repeat sequences important to genomic rearrangements and increases our understanding of how the use of a reference sequence to assess complexities in patient cohorts can confound analysis.

530W

High Resolution Copy Number Variation Analysis using Droplet Digital PCR. *J. Berman, J. Regan, N. Heredia, D. Skvortsov, N. Klitgord, S. Tzonev, E. Hefner.* Bio-Rad Digital Biology Center, Pleasanton, CA.

Copy number variations (CNVs) from single-gene to segmental duplications are increasingly recognized as critically dynamic features of the human genome. Altered copy number status is associated with several disease states, including autism spectrum disorder, schizophrenia, and multiple cancers. Technological hurdles have historically limited meaningful interrogation of copy number status. SNP-based microarrays, aCGH, and qPCR typically lack the sensitivity and fine quantitative discrimination required for resolution of higher copy number states. This is particularly true for copy number evaluation in heterogeneous samples, like somatic mosaicism or tumor biopsies, where only a small fraction of cells might have copy number alterations. Droplet digital PCR (ddPCR) enables accurate and reproducible copy number determination using a simple, cost-effective workflow amenable to high throughput. Using single-well ddPCR, consecutive copy number states can be distinguished between samples of 5 and 6 copy number at 95% confidence levels. Here we use validated Bio-Rad ddPCR CNV assays to discriminate copy number status of multiple genes with low to high copy number states. The evaluation of higher-order oncogene amplifications, such as MYC, MET, and FGFR2, will be demonstrated in FFPE and cell line matrices. The ability to resolve a 5% difference (1.05-fold) in copy number will be demonstrated using mosaic or admixed samples heterogeneous for copy number status. This has immediate implications for researchers interested in reproducibly resolving somatic mosaicism, tumor heterogeneity, or other applications where fine quantitative discrimination is essential.

531T

Accurate measurement of C4 gene copy number variation by chip-based digital PCR. *C. Chen, D. Do, K. Li, D. Keys.* Genetic Applications R & D, Life Technologies, 180 Oyster Point Blvd., South San Francisco, CA 94080.

The fourth component of human complement (C4) is an essential factor of the innate immunity. The C4 gene is represented as two isoforms, C4A and C4B, in the genome. Its copy number varies in human individuals. Lower copy number variation (CNV) of the C4 gene has been implicated with various diseases in human. Here, we have evaluated both chip-based digital PCR (dPCR) and conventional qPCR for measuring the copy numbers of C4 genes in human DNA samples from Coriell and patient samples. There is a high concordance in C4 copy numbers between two methods. However, dPCR offers a better precision which enables to measure a smaller change in DNA copy numbers for up to 8 copies. qPCR can determine the DNA copy numbers for 0- 4 copies. The chip-based dPCR system can accurately measure DNA copy numbers for at least 6 copies. Of 251 patients and 253 healthy controls examined, there is a strong association of the low C4A copy number with primary sclerosing xholangitis (PSC).

532F

Copy number variants associated with the risk of systemic lupus erythematosus: Design a lupus risk prediction system. *Y. Chung, S. Jung, J. Kim, S. Yim.* Integrated Res Ctr Genome Polymorphism, Catholic Univ Korea,Sch Medicine, Seoul, South Korea.

Copy number variation (CNV) is one of the major components of human genetic variations and it is thought to contribute to inter-individual differences in diverse phenotypes. Several CNVs have been identified to be associated with systemic lupus erythematosus (SLE) mostly by the target gene approach. However, genome-wide feature of CNVs and their roles in the risk of SLE remain unknown. We aimed to discover SLE-associated CNVs in Korean women. In this study, we performed genome-wide assessments of CNVs and replicated the significant candidate variants in 946 SLE cases and 702 controls. We found that three deletion-type CNVRs in 1q25.1, 8q23.3, and 10q21.3 were significantly associated with SLE. Of the three candidates, CNVRs in 1q25.1 (RABGAP1L) and 10q21.3 were successfully replicated (OR=1.30, P=0.038 and OR=1.90, P=3.6×10−5, respectively) and the associations were confirmed again by deletion-typing PCR. The CNVR in C4 gene, which showed a potential association in the discovery stage, was included in the replication analysis and found to be significantly associated with the risk of SLE (OR=1.88, P=0.01). Through deletion-typing PCR, the exact sizes and breakpoint sequences of the deletions were defined. Individuals with the deletions in all three loci (RABGAP1L, 10q21.3 and C4) had a much higher risk than those without any deletions in all three loci (OR=5.52, P=3.9×10−4) (Kim et al., Arthritis Rheum 2013 65:1055-63). Based on these findings, we develop a multiplex CNV analysis system for predicting the risk of SLE using MLPA-CE-SSCP method. This system contains six CNV targets including the three CNV loci described above. Here we present the efficacy of our MLPA-CE-SSCP based lupus risk prediction system.

533W

Genome-wide CNV association study of primary caries. *K.T. Cuenco[1], M. Lee[1], X. Zheng[1], E. Feingold[1], D.E. Weeks[1], R.J. Weyant[1], R.J. Crout[2], D.W. McNeil[2], M.L. Marazita[1].* 1) Univ Pittsburgh, Pittsburgh, PA; 2) West Virginia Univ, Morgantown, WV.

Background/Objective: Dental caries is a common childhood disease and highly heritable (30% to 60%) even after adjustments for shared environment. The heritability of primary tooth caries is estimated to be 59%. Genetic risk factors for dental caries in children have been investigated through the NIH Gene, Environment Association Studies Consortium (GENEVA) genome-wide association study (GWAS). Past GENEVA GWAS analyses of primary caries have not considered the contribution of genetic effects from copy number variation (CNV). A more comprehensive investigation of these genetic variants is needed. We attempt a genome-wide CNV association study to identify CNVs correlated with childhood dental caries using existing GWAS data. Methods: Participants were drawn from caries populations collected through University of Pittsburgh, Iowa Fluoride Study, and Iowa Head Start Study. Caries status and demographics were obtained from GENEVA children (n = 1245; ages 3–12 years). Subjects self-reported as Caucasian. Primary caries status was defined as decay and filling teeth (DFT) score ≥ 1 [n = 584] vs DFT = 0 [n=661]. These subjects were included in the GENEVA genotyping runs on a Illumina 610-Quad platform. Marker calls were generated by the Center for Inherited Disease Research and used for CNV calling. We used CNV calls generated using the GC model wave adjustment procedure in PennCNV software. Poor-quality samples are filtered out if the log R ratio standard deviation was greater than 0.3. A CNV call was considered acceptable if the call was based on > 3 consecutive markers. CNVs with copy number <2 were defined as deletions. Copy number >2 were defined as duplications. Caries association analyses focused on autosomes only, and adjusted for DNA sample source and age. Additional existing CNV probes were analyzed for association with caries. Results: We are currently assessing which CNVs associate with primary caries. Any CNVs that correlated with primary caries may give insight into additional genetic targets to pursue. Support: DE018903, DE014899, DE09551, DE12101, DE020127.

534T

Copy number variation profiling of patients with Oesophageal Atresia and VACTERL. *A. deKlein[1], E. Brosens[1, 2], H.P. Zaveri[3], E. de Jong[1, 2], D.A. Scott[3, 4], D. Tibboel[2].* 1) Clinical Gen, Erasmus MC, Rotterdam, Netherlands; 2) Pediatric Surgery, Erasmus Medical Centre - Sophia Children's Hospital, Rotterdam, Netherlands; 3) Molecular & Human Genetics , Baylor College of Medicine, Houston, Texas, USA; 4) Molecular Physiology & Biophysics, Baylor College of Medicine, Houston, Texas, USA.

Esophageal Atresia (EA) with or without Tracheal-Esophageal Fistula (TEF) are common congenital anomalies whose cause is unknown in over 90% of affected patients. EA/TEF can be present either as an isolated defect or in association with other developmental defects: e.g. as one of the core features of the VACTERL (Vertebral, Anal, Cardiac, TEF, Renal and Limb anomalies) association. The hypothesis that genetic defects contribute to both EA/TEF and VACTERL etiology is supported by the fact that EA/TEF is a variable feature of several known monogenetic syndromes. Among these possible defects are Copy Number Variations (CNVs). As de novo CNVs can help to identify causal genes or affected biological pathways, the recurrence of unique and rare inherited CNVs may, in combination with other factors, predispose for the development of EA/TEF and other features of the VACTERL association. We therefore profiled 255 affected individuals with SNP-arrays. All had one or more large CNV, most of them were known polymorphisms. We observed six unique loci with a de novo CNV: 4p15, 4q35, 5q11, 6q23, 7p22 and 8q13. We also identified over 300 inherited CNVs which were either absent or uncommon in published control cohorts and our in-house database. Interestingly, 45 of these inherited variants were recurrent in our patient cohort. Ingenuity Pathway Analysis revealed enrichment of several biological functions including embryonic and digestive tract development. Using our genome-wide approach we identified several loci that may impact biological pathways disturbed in EA/TEF or VACTERL association patients.

535F

Human Endogenous Retroviral Elements (HERVs) Mediate Multiple Large Deletions and Reciprocal Duplications Suggestive of NAHR. *P. Dittwald[1, 2, 6], I.M. Campbell[3, 6], A. Shuvarikov[4], C.R. Beck[3], P. Hixson[3], T. Gambin[3], C.A. Shaw[3], A. Gambin[2, 5], J.A. Rosenfeld[4], P. Stankiewicz[3].* 1) College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, Poland; 2) Institute of Informatics, University of Warsaw, Warsaw, Poland; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 5) Mossakowski Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland; 6) Equal contribution.

The vast majority of recurrent deletions, duplications, inversions, and translocations, resulting in genomic disorders or variation, are mediated by nonallelic homologous recombination (NAHR) between low-copy repeats (LCRs) or segmental duplications distributed throughout the human genome. However, recurrent genomic rearrangements are also mediated by non-LCR structures: AT-rich cruciforms [e.g. t(11;22)] and human endogenous retroviruses (HERVs) [AZFa deletions in Yq12 and t(4;18)]. Most recently, using chromosomal microarray analysis, we identified nine de novo, identically-sized 3.3 Mb deletions in 3q13.2q13.31, which were likely mediated by NAHR between two directly oriented HERV elements 4,954 and 5,684 bp in length sharing 4,839 bp of 95.71% DNA sequence identity. We hypothesized that other HERV pairs throughout the human genome may serve as substrates for NAHR and lead to genomic instability. Based on the HERV properties at the 3q13.2q13.31 and Yq12 loci, we developed features of HERV elements mediating NAHR. We identified 204 HERV pairs matching our criteria and thus potentially predisposing ~ 11.8% of the human genome to HERV-mediated recurrent rearrangements. Cross-referencing these predicted susceptibility regions with copy-number variants (CNVs) in our chromosomal microarray databases revealed deletions and apparently reciprocal duplications, ranging in size between 189 kb and 1.36 Mb and mapping to LCR-free chromosomal regions 1q41, 2p12 (two loci), and 11q24.3 in nine unrelated patients. DNA sequencing of long-range PCR-amplified junction fragments showed that all CNVs had different breakpoints mapping within HERVs 5,711–6,136 bp in size and sharing stretches of 93-95% DNA sequence identity with their partner. The presence of HERV-mediated deletions and reciprocal duplications indicates NAHR as a causative mechanism, even though the length and the sequence homology of the HERV elements are less than that currently thought to be required for NAHR. Our data demonstrate an underappreciated role of HERV elements in human genome instability. We propose that in addition to HERVs, other repetitive elements such as LINEs may also be responsible for the formation of recurrent CNVs via NAHR.

536W

A comprehensive high resolution map of copy number variants shows unique disease risks in a consanguineous Arab population. *K. Fakhro[1], J.L. Rodriguez-Flores[2], N. Yousri[1], A. Robay[1], J.G. Mezey[2,3], R.G. Crystal[2].* 1) Department of Genetic Medicine, Weill Cornell Medical College - Qatar, Doha, Qatar; 2) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY; 3) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Genetic studies in Arab populations suffer a lack of databases containing population-level background genetic variation, making it difficult to assess the impact of newly discovered mutations in Arab patient cohorts. We present here the first comprehensive copy-number variation (CNV) map of 108 native Qataris, comprising individuals from all 3 Arab subpopulations (Q1 - Bedouin, Q2 - Persian and Q3 - African). We used QuantiSNP and CNVPartition to make CNV calls from Illumina 2M SNP array data (mean call rate >99.8%) [1624 high quality CNV Regions (CNVRs) - 1201 deletions, 423 duplications; mean size 49.9 kb; range 59 bp-1.97 Mb; total coverage 21 Mb]. Of 1624 CNVRs, 445 affected 2209 unique genes, including 78 severe Mendelian disease genes. Of interest, 459 of 1624 CNVs (28%) were novel to Qataris; 91 of these affected 243 genes, including 14 severe OMIM genes. In order to also assess CNVs below the detection resolution of arrays, we used CNV calls from whole genome sequencing (WGS) data (>50X coverage) in the same 108 individuals. 400,239 CNVs were detected (average >3,800/individual; size range 50bp-1.2Mb) affecting >246Mb of genomic content - the majority of which were non-genic (>80%) and novel to Qataris (52%). Surprisingly, only a minority (~30%) of array-CNVRs overlapped WGS-derived CNVRs; thus, despite a >10-fold increase in number of CNVRs called, WGS-algorithms may still miss a significant number of potentially real, large CNVRs in a population. In total, CNVRs in this population affected >9000 unique genes, with ontological enrichment of >10 KEGG pathways relevant to public health (e.g., cancer, cardiovascular function, type II diabetes) and 78 severe Mendelian disease genes (e.g., ciliary disorders, mental retardation, and congenital defects). The distribution of CNVs was skewed across the three major subpopulations, with a higher number of CNVs as well as a larger percentage of the genome affected per individual in Q1 *vs* Q2 or Q3. Together, this data suggests that CNVs could be important to public health of the Qatari population, that public databases lack a significant amount of CNVs from understudied global populations, and that SNP-array CNVs may still be useful rather than redundant if the aim is to comprehensively uncover all CNVRs within a population.

537T

Post-zygotic Structural Changes in the Nuclear Genome of Human Blood Cells. *L.A. Forsberg[1], C. Rasi[1], D. Absher[2], L. Lannfelt[3], A. Morris[4], C. Lindgren[4], E. Ingelsson[5], L. Lind[6], D. Dumanski[1].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; 3) Department of Public Health and Caring Sciences, Division of Molecular Geriatrics, Uppsala University, Uppsala, Sweden; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom; 5) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 6) Department of Medical Sciences and Science for Life Laboratory, Uppsala University.

Post-zygotic, acquired with age mutations in normal cells represent an understudied aspect of genome biology. Human post-zygotic mosaicism has been studied in embryos, aborted fetuses, children with developmental defects and in cancer. However, little is still known about the type and frequency of acquired genetic aberrations in normal cells from subjects in the general population, especially from larger cohorts that are well stratified by age. Our recent analyses suggest that such mosaicism is surprisingly common, with the highest estimate of 3.5% for elderly/old subjects being affected by various large-scale aberrations. Our analyses are now extended to include 1153 elderly men from a Swedish population-based cohort ULSAM (Uppsala Longitudinal Study of Adult Men) that has been followed clinically for more than 40 years. We aim to i) better describe to frequency and genomic distribution of acquired structural genetic variants in normally aging subjects, ii) further investigate longitudinally collected samples and describe how the frequencies of variant-clones varies with time within individuals and their relation to onset of disease, iii) to determine which cellular sub-compartments of blood usually display different types of genetic structural aberrations and to find phenotypic correlations. So far, peripheral blood sampled at ages ranging from 70.7-83.6 years has been analyzed using the Illumina 2.5MHumanOmni-chip with strict selection of genotyping quality. The size thresholds for scoring the structural variants were 250 kb for gains and 10 kb for deletions and CNNLOH. Next generation whole genome sequencing was used for validation. Remarkably, at least one structural change was observed in 42.7% of the genotyped men with gains being the most common type. These results illustrate the high frequency and importance of post-zygotic mosaicism in normal cells, which should be studied further for associations with various diseases.

538F

Creating a haplotype map for multi-allelic forms of copy number variation in humans. *R. Handsaker[1,2], S. McCarroll[1,2].* 1) Harvard Medical School, Department of Genetics, Boston MA; 2) Broad Institute, Program in Medical and Population Genetics, Cambridge, MA.

An important need in human genetics is to begin to understand multi-allelic copy number variants (mCNVs) - polymorphisms involving genomic segments that segregate in 4, 10, 20 or more different potential copy-number levels in human populations, generally due to tandem or dispersed duplications. Such variants have been refractory to almost all earlier molecular technologies and have therefore not been understood at the levels of alleles, allele frequencies, and haplotypes.

To accurately measure the copy number of mCNV loci using whole-genome sequence data, we first developed new computational approaches building upon Genome STRiP (Handsaker, 2011), a method for discovering and genotyping deletion polymorphisms. Extending our approach to high copy-number and mCNVs, we have ascertained over 10,000 CNVs (including 1600 mCNVs) in 849 individuals using low-coverage (4x) sequencing data from Phase 1 of the 1000 Genomes Project and we have made precise measurements of integer copy number in each individual. Using intensity data from SNP arrays, we estimate the false discovery rate of these CNVs to be less than 5%. The subset of these CNVs that were 'genotypeable' in an earlier, array based study (Conrad et al.) show genotype concordance greater than 99% with our results.

To understand how CNV alleles relate to SNPs and haplotypes, we needed to determine not just diploid copy number in each individual, but the contribution of each chromosome and haplotype to diploid copy number. We developed an approach that combines accurate copy number measurements with dense SNP data, building upon innovations in the Beagle software, to infer CNV alleles, phase them with nearby SNPs, and utilize haplotypes to predict the copy-number contribution of each chromosome in each individual. This work has allowed us to begin to evaluate the limits of genotype imputation for multi-allelic, recurrently mutating forms of copy number variation. We describe new ways to phase and impute the states of multi-allelic, high-copy-number CNVs, and their application to GWAS data. We find that these methods are highly successful at many mCNV loci, including loci with many common alleles.

539W

Quantitative Analysis of Mosaic CNVs in Human Fibroblasts and iPSC By Digital Droplet PCR. *M. Haney[1,2], A. Abyzov[3,4,5], Y. Zhang[3,8], J. Ferrandino[3,6,7], J. Mariani[3,6], D. Palejev[3,6], L. Tomasini[3,6], L. Belmaker[3,6], A. Szekely[3,8], M. Wilson[3,4,5], A. Kocabas[3,6], N. Calixto[3,6], E. Grigorenko[3,6,9,10], A. Huttner[3,12], S. Weissman[3,8], M. Gerstein[3,4,5,11], F. Vaccarino[3,6,7], A. Urban[1,2].* 1) Department of Genetics, Stanford University, Palo Alto, CA; 2) Department of Psychiatry and Behavioral Sciences,Stanford University, Palo Alto, CA; 3) Program in Neurodevelopment and RegenerationYale University, New Haven CT; 4) rogram in Computation Biology and Bioinformatics, Yale University, New Haven CT; 5) Department of Molecular Biophysics and BiochemistryYale University, New Haven CT; 6) Child Study Center,Yale University, New Haven CT; 7) Department of Neurobiology,Yale University, New Haven CT; 8) Department of Genetics,Yale University, New Haven CT; 9) Department of Psychology,Yale University, New Haven CT; 10) Department of Epidemiology and Public Health,Yale University, New Haven CT; 11) Department of Computer Science,Yale University, New Haven CT; 12) Department of Pathology,Yale University, New Haven CT.

Digital Droplet PCR (ddPCR) has recently emerged as an ultra-sensitive method of analyzing specific nucleic acid sequences in a complex mixture of sequence fragments. We have developed an approach to employ this technology platform to detect and quantify human genomic somatic mosaic Copy Number Variants (CNVs) that are present in as little as .01% of the native cell population. This unprecedented degree of sensitivity was enabled by the novel approach of 'junction probe placement.' In this we used fluorescent probes that uniquely bind to the junction sequences of mosaic CNVs. The result is a binary fluorescent readout for each ddPCR reaction droplet, each of which contains only picograms of genomic DNA, yielding high sensitivity while maintaining a low false positive rate. Using this method we were able to confirm and quantitate a previously unreported abundance of somatic genome variation, being present in the form of mosaic lineage-manifested CNVs (LM-CNVs) in human fibroblast tissue. We defined the term LM-CNV to describe CNVs detected by genome-wide analyses in an iPSC line but not in the fibroblast culture from which the given iPSC line was derived -- without making a statement as to the nature of the CNV-forming event.These LM-CNVs had become unmasked in iPSC lines derived in a clonal fashion from the fibroblast tissue of origin. The LM-CNVs had originally been detected by low-coverage whole-genome sequencing in 7 fibroblast samples and 20 corresponding induced pluripotent stem cell lines obtained from two families [Abyzov et al., Nature. 2012 Dec 20;492(7429):438–42]. We found that on average an iPSC line has two LM-CNVs. After detecting LM-CNVs by sequencing based analysis in the iPSC lines we investigated the masked, mosaic presence of the same CNVs in the fibroblast tissue of origin. Using standard PCR across the predicted sequence breakpoints we determined that more than half of the LM-CNVs detected in iPSC lines were already present as low allele frequency, mosaic somatic CNVs in the fibroblasts and that up to 40% of fibroblast cells carry such medium-sized to large somatic CNVs. We then used ddPCR and custom designed junction placement probes to determine the degree of mosaicism of the LM-CNVs. We found that in the samples analyzed the range of mosaicism for LM-CNVs was from 0.3%–14% allelic frequency.

540T

Copy number variation studies by single cell sequencing. *J. He[1], P. Liu[2], Y. Man[3].* 1) Department of Biology, South University of Science and Technology of China, Shenzhen, guangdong, China; 2) Department of Chemistry Boston University, Boston, MA 02215; 3) Department of Biology Boston University, Boston, MA 02215.

Studying complex biological systems such as a tumor, a developing embryo, or a brain often involves understanding the behavior and heterogeneity of the individual cells that constitute the system and their interactions. Genomic analysis by single cell sequencing has provided important insights; however current single cell amplification methods and the bioinformatics tools are at the early developing stage. In particular, current single cell sequencing technology has strong amplification bias which makes the CNV analysis challenging. We have therefore been working to optimize methods for CNV analysis of single cells. We performed comprehensive comparisons of MALBAC, GenomePlex WGA4 and MDA methods for single cell amplification. We also developed CNV detecting bioinformatics tools that are good for single cells sequencing data with amplification bias and relatively low coverage. We will describe the application of single cell sequencing in profiling CNVs in neurons. There is a major unanswered question in neuroscience that whether there exists genomic variation between individual neurons of the brain. This variation may contribute to functional diversity or to an unexplained burden of neurological disease. To address this question, we developed a method to amplify genome of single neurons from rat. Single-neuron sequencing allows systematic assessment of genomic diversity in normal brain and shed light into understanding the role of copy number variations in recognition and memory.

541F

Measurement of Cyclin D1 copy number variation at the single cell level using droplet digital PCR. *E. Hefner, Y. Jouvenot, N. Klitgord, K. Hamby.* Bio-Rad Laboratories, Pleasanton, CA.

Amplification of the oncogene CCND1 is linked to poor prognosis in a variety of cancers including melanoma. Accurate copy number variation (CNV) measurement at the single cell level could shed light on the spectrum of CCND1 amplifications within the tumor and ultimately refine prognosis. Current methods for single cell CNV include qPCR, sequencing and array CGH. All of these methods require targeted or whole genome enrichment strategies followed by discrete sample measurements. However, the pre-amplification solution creates a different set of issues including increased cost, time and most importantly the potential for skewed results due to bias in the enrichment procedure. Here, we present a method for determining CCND1 copy number state in single cells without the need for pre-amplification. Samples containing subpopulations of cells with various levels of CCND1 amplification were subjected to single cell ddPCR and the results compared to bulk PCR measurements. The results clearly demonstrate the resolving power of ddPCR for the detection of CCND1 copy number state at the single cell level.

542W

Comparison of copy number variation (CNV) calling performance in large numbers of technical replicate SNP array data using three different, widely-used CNV calling algorithms. *A. Hofmann[1], S. Herms[1,2], F. Degenhardt[1], T.W. Mühleisen[3], M.M. Nöthen[1], S. Cichon[1,2,3], P. Hoffmann[1,2,3].* 1) Institute of Human Genetics Department of Genomics Life & Brain Center University of Bonn; 2) Genomics Research Group Division of Medical Genetics University hospital Basel; 3) Institute of Neuroscience and Medicine Genomic Imaging Research Center Juelich.

Copy number variants (CNVs) have been shown to explain part of the heritability in various multifactorial diseases. Many of these findings are derived from SNP-array data generated in the course of large genome-wide association studies (GWAS). This is not without challenges, however: SNP-arrays contain an ever increasing density of probes which results in a decreased signal-to-noise ratio. The latter causes problems for automated CNV calling algorithms and is a major cause for the unambiguous calling of smaller (<350kbp) and/or low frequency CNVs. Recent studies have therefore focused on the much more reliable calling of larger CNVs (number of consecutive marker or length) and often considering the easier to detect deletion events only. This study compares the performance and differences in CNV calling using three widely-used CNV calling algorithms: CNVPartition, QuantiSNP2 (v2.2), and PennCNV. As SNP array data a large number of technical replicates (n>500) all genotyped at the University of Bonn on Illumina's HumanOmniExpress- and HumanOmni1M arrays were used. We observed an unexpectedly high fluctuation in the prediction of cnv events throughout the three algorithms. All gave comparable findings for larger findings (> 1Mbp) but suffered to give consensus results for smaller variants. Since the replicates were typed on the same array type, this allowed to evaluate effects of the chemistry or the operator in the lab. Results will be shown and based on that parameters will be presented that allow for a better evaluation of the quality of CNV callings from SNP array data.

543T

Ultraconserved Elements: often disrupted in disease-specific copy number variation, almost never involved in benign CNVs. *R.B. McCole, C.Y. Fonseka, C.-T. Wu.* Department of Genetics, Harvard Medical School, Boston, MA.

Ultraconserved elements (UCEs) are genomic regions showing exceptionally strong and unexplained levels of DNA sequence conservation between related species. We have hypothesized that UCEs function as 'copy counters,' helping genomes to maintain exactly the correct number of chromosomes and the right amount of genetic information. This model predicts that perturbations in the copy number of UCEs will be highly deleterious to the cell and, ultimately, to the individual. Early tests of the model examined the co-occurrence (overlap) of UCEs with copy number variants (CNVs) and segmental duplications (SDs) in the human genome. These studies demonstrated that the profiles of human CNVs and SDs are highly depleted for UCEs, thereby lending support to our model. We have now extended these studies to address the following questions: How quickly do UCEs become depleted from the profiles of CNVs? When does this depletion occur? Does it require one or more rounds of human reproduction? Using new sets of UCEs and the most recent datasets of CNVs, we find that depletion is rapid and does not necessarily involve passage through the germline. We also find that depletion is absent from CNVs that arise specifically in cancer, suggesting that while UCEs are refractory to deletion or duplication in healthy cells, in the disease state they are often disrupted, which may be an important but hitherto overlooked aspect of cancer initiation or progression.

544F

Copy Number Variation Analysis for Whole-Exome and Targeted Sequencing using NextGENe® Software Version 2.3.4. *J. McGuigan, J. Wu, C.S. Liu.* SoftGenetics, LLC., State College, PA.

NextGENe version 2.3.4 includes a sophisticated new algorithm for copy-number variation (CNV) detection from a wide variety of projects, including whole-exome and targeted sequencing panels. This algorithm is based on fitting a beta-binomial model to the coverage ratio. This fitting process results in the amount of noise in the data being measured automatically. The confidence of CNV calls is adjusted based on the amount of noise.

Regions are defined for the aligned data in a 'sample' project and a 'control' project. Whole-exome sequencing can use CDS locations, while targeted sequencing can define regions as the location of amplicons. Each region has a total RPKM coverage (sample plus control) and a coverage ratio (sample divided by total). The fitted equation returns a 'dispersion' value for any level of coverage and this value is used to generate beta-binomial distributions for 3 cases- heterozygous deletion, normal (no CNV), and heterozygous insertion (increased copy number). Normalized likelihoods are calculated from these distributions and used in a Hidden Markov Model (HMM) to make the final CNV calls. Each call is given a phred-scaled probability score for insertion and deletion.

The final report contains the calls, quality scores, annotation, and analysis results (dispersion values and likelihoods). Results can also be viewed in graphical form, showing the coverage ratio and call for each region. In this analysis both whole-exome and targeted sequencing data was analyzed.

545W

Identification of deleterious CNVs in a low SES African American Population. *K.B. Mercer[1], L.M. Almli[1], K.J. Ressler[2], J.G. Mulle[1].* 1) Department of Psychiatry, Emory School of Medicine, Atlanta, GA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA.

Genome-wide association studies (GWAS) and copy number variant (CNV) discovery have become the standard approaches in attempts to identify genetic variants that result in heritable disorders. The discovery of CNVs has advanced our understanding of genetic variants associated with disorders such as autism and schizophrenia, and will continue to reveal variants that are responsible for yet unknown genetic risk. However, most surveys of CNVs have utilized self-selected case-control populations rather than population-based samples, which may lead to study bias and flawed estimates of CNV prevalence. In the current study, we aim to estimate the prevalence of CNVs associated with genomic disorders in an underserved, impoverished, at-risk African American population. Methods: The Grady Trauma Project is actively recruiting study participants from Grady Memorial Hospital, (Atlanta, GA) which offers healthcare to low income individuals. Study participants are recruited from waiting rooms of either primary care or OB/GYN offices. Willing participants are asked to donate DNA to be used in genetic studies, and to complete questionnaires assessing demographics, trauma exposure and various health outcomes, particularly those related to Depression and PTSD. We used Illumina Omni-Quad genome-wide array data (1M SNPs) derived from 2,927 unrelated African Americans and the PennCNV detection program to identify large CNVs (>100kb) that have previously been found to significantly associate with disease risk. We compared the frequency of disease associated CNVs ($p<0.05$) reported by Cooper et al in Nature Genetics 2011 (healthy controls only) to the frequency of these CNVs in our study sample cohort. Results: We find an unusually high prevalence of genomic-disorder associated CNVs: 2% of this population carries at least 1 CNV associated with a genomic disorder (n = 64), compared with 1% of controls (n = 88 out of 8,329; p-value =1.971e-06). The odds ratio (2.22; 95% CI: 1.59-3.10) reveals an increased risk of deleterious CNVs in this population. Notably, we find 2 previously undiagnosed individuals with the 22q11.2 (VCFS) deletion, and 2 individuals with the RCAD (renal cysts and diabetes disorder; 17q12) deletion. This Z fold excess implies that this population may be underserved and undertreated with respect to detection of genomic disorders.

546T

Determining the utility of MitoExome targeted array CGH in the diagnosis of OXPHOS disorders. *H.S. Mountford[1,2], E.J. Tucker[1], A.G. Compton[1,2], N.J. Lake[1,2], S.G. Hershman[3,4,5], S.E. Calvo[3,4,5], V.K. Mootha[3,4,5], D.R. Thorburn[1,2,6].* 1) Murdoch Childrens Research Institute, Royal Children's Hospital, Flemington Road, Parkville, VIC, 3052, Australia; 2) Department of Paediatrics, University of Melbourne, Melbourne, VIC, 3052, Australia; 3) Center for Human Genetic Research and Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Sixth Floor, Boston, MA 02114, USA; 4) Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA; 5) Broad Institute of Harvard and Massachusetts Institute of Technology, 7 Cambridge Centre, Cambridge, MA 02141, USA; 6) Victorian Clinical Genetics Services, Royal Children's Hospital, Flemington Road, Parkville, 3052, Australia.

Mitochondrial oxidative phosphorylation (OXPHOS) disorders are the most common inborn error of metabolism, affecting at least 1 in 5000 live births. OXPHOS diseases are notoriously difficult to diagnose, as they show extreme clinical and genetic heterogeneity, comprising over 150 monogenic disorders with many more 'disease genes' yet to be discovered. For about 50% of patients and families, the genetic basis of their OXPHOS disorder remains elusive. Previously, we developed a targeted DNA capture and massively parallel sequencing method to detect variants within the mitochondrial genome and exons of 1034 nuclear genes encoding the mitochondrial proteome or MitoExome [1]. Our MitoExome study investigated 45 unrelated Australasian patients with an enzyme and clinical diagnosis of an OXPHOS disorder, none of whom had a previous molecular diagnosis. Our studies have now provided a molecular diagnosis in 15 of these patients with mutations in 11 mitochondrial OXPHOS-related disease genes [1, 2] and identified variants in another 5 candidate genes (9 patients) we now regard as new disease genes. The remaining 21 patients (~46%) had no recessive-type likely pathogenic mutations identified. To further investigate these unsolved patients we designed a Roche NimbleGen 12x135 CGH array to target the MitoExome genes. The design covered all 14,053 exons; 51,380 probes falling within an exon and at 50bp intervals; 63,937 probes falling within introns at 900bp intervals; and 22,285 probes designed at 3,600bp intervals designed upstream, downstream and across each MitoExome gene to form a 'backbone'. Currently, we have applied this technology to 12 patients and identified an average of 4.1 novel copy number changes per patient; 2.4 duplications and 1.6 deletions. Current investigations aim to validate copy number changes found in these patients by long-range PCR before investigating the pathogenicity of these alterations in patient tissue and fibroblast cell lines. To date, at least 2 appear to contain pathogenic copy number changes. Our study is investigating the utility of using a combination of MPS and aCGH to enable a greater diagnostic yield than targeted or Exome MPS approaches alone would enable. The sensitivity and specificity of the CGH array make it very useful to follow up patients without a molecular diagnosis from MPS. 1.S. E. Calvo et al., Sci Transl Med 4, 118ra10 (2012) 2.E. J. Tucker et al., Cell Metab 14, 428 (2011).

547F

European-specific selection at the CCL3L1 locus. *H.T. NGUYEN[1, 2], T.R. MERRIMAN[1], M.A. BLACK[1].* 1) Biochemistry Department, University of Otago, 710 Cumberland street, Dunedin 9016, New Zealand; 2) Mathematics and Statistics Department, University of Otago, 710 Cumberland street, Dunedin 9016, New Zealand.

Infectious disease has been a prominent selective agent in Europe for over one thousand years. The chemokine CCL3L1 binds to the CCR5 receptor, and a 32-bp-deletion mutant (delta32) of CCR5 confers resistance against human immunodeficiency virus (HIV) infection. CCL3L1 copy number (CN) variation has been associated with rheumatoid arthritis and susceptibility to HIV infection, supporting the hypothesis that genetic variants in CCL3L1-CCR5 have previously been subject to selection. We therefore examined selection at the CCL3L1 locus in East Asian, West African, European and American populations using 3 megabases of 1000 Genomes Project data in this region. Two tests based on extended haplotype homozygosity were applied: Integrated Haplotype Score (iHS) and Cross Population Extended Haplotype Homozygosity (RSB). The strongest iHS-signal values were seen in the European population between CCL4 and CCL3L3. Selection specific to European was supported by the RSB signal between European and the three other populations. We then measured CCL3L1 gene CN and identified tag-SNPs for CCL3L1 CN using a newly-developed tool, CNVrd2, in order to identify the relationship between CCL3L1 CN and the selected haplotype identified from the iHS and RHB tests. This haplotype was common in individuals with CN=2, but less common in other groups ($p = 0.00045$, 61%, 34% and 45.6% in CN=2, CN<2 and CN>2 sample sets respectively). Another haplotype was identified specifically for CN<2 ($p = 1.68e{-}08$, $r = -0.3$) and tag-SNPs for deletion (CN<2) were identified only in European (maximum $r2 = 0.42$, $p < 0.05$). We conclude that there is evidence of selective pressure specific to populations of European ancestry, consistent with a role for the CCL3L1 locus in resistance to infectious disease in Europe.

548W

Integration of copy number and structural variation across families can provide unique insight into disease pathology: *Osteogenesis Imperfecta* and Autism cases explored. *A. O'Hara, L. Culot, S. Verma, Z. Che, S. Shams.* BioDiscovery Inc., Hawthorne, CA.

Rapid identification of candidate genomic aberrations responsible for disease phenotype is important in both research and clinical settings. Increasingly, structural variant information must also be combined with sequencing data in order to uncover pathogenic events and then be further resolved through comparison with related genomes. Here we explore two constitutional cases, *Osteogenesis Imperfecta* in a proband with consanguineous parents, explained through comparison with sibling, and an autism case with a suspected *de novo SHANK2* deletion and family trio data. An approach that can integrate structural and sequence variation across related samples will be presented.

549T

SMN1 and SMN2 conversion rates and their influence on the identification of spinal muscular atrophy carriers. *J. Regan, T. Legler, D. Shelton, D. Maar, K. Hamby.* Bio-Rad Laboratories, Pleasanton, CA.

Spinal muscular atrophy (SMA) is the second leading cause of death in children behind cystic fibrosis, affecting 1 in 6000 newborns. SMA is an autosomal recessive disease attributed to deleted or mutated survival motor neuron gene 1 (SMN1). This gene is located within a 500 kb region that is copied and inverted to the centromeric side of SMN1. The copied SMN gene is called SMN2 and it has a single nucleotide change from SMN1 that causes alternative splicing of transcripts, which leads to a less stable protein than protein derived from full-length SMN1 transcripts. The repetitive nature of the inverted region makes it prone to homologous recombination and copy number variation. Individuals with no functional copies of SMN1 due to deletion or mutation have SMA, the severity of which is determined by the number of SMN2 genes. Accurately quantifying the number of SMN1 and SMN2 genes is difficult as just a single nucleotide difference distinguishes the two genes. Afflicted individuals with just 1 - 2 copies of SMN2 have onset of muscle weakness within months of birth followed by death before the age of two, whereas individuals with higher numbers of SMN2 can have symptom onset after the third decade of life and no shortened life expectancy. Histone deacetylases activate SMN transcription, but the window for applying therapies to newborn babies is very small as >95% of motor neurons are lost within 6 months. The promise of reducing the severity of the disease enhances the importance of early diagnosis and intervention before irreversible motor neuron damage occurs. The autosomal recessive nature of the disease provides the possibility that through careful screening the incidence of the disease could be significantly reduced. Traditional carriers (1,0) can be identified using qPCR, however this methodology fails to identify carriers that have cis-configured copies of SMN1 (2,0). Here, we use droplet digital PCR (ddPCR) to accurately quantify the number of SMN1 and SMN2 genes in 384 HapMap samples from four different ethnic backgrounds. Furthermore, we employ digital linkage analysis to determine the frequency of SMN1-to-SMN2 conversion and vice versa. This information is useful for understanding the factors that influence the frequency of carriers, which is estimated to be 1/50. Lastly, we explore the possibility and limitations of using linkage analysis for identifying cis-configured carriers, which are missed by all other methodologies.

550F

Frequency of gene usage and copy number variation within the rearranged Immunoglobin Heavy-Chain Variable locus based on immune repertoire sequencing. *M.J. Rieder[1], D. Williamson[1], A. Sherwood[1], R. Emerson[1], C. Desmarais[1], M. Chung[1], H. Robins[1,2], C. Carlson[1,2].* 1) Adaptive Biotechnologies, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA.

The human adaptive immune system is composed of both B and T cells that undergo somatic recombination at specific loci to create rearrangements of Variable (V), Diversity (D) and Joining (J) gene segments. For the B-cell immunoglobin receptor heavy-chain (IGH), the CDR3 regions are defined by the VDJ gene segments and nucleotide insertions/deletions at these junctions that create the vast sequence diversity of the IGH repertoire. Characterizing the germline DNA in these regions is impeded by the high sequence similarity between gene segments, mutation and copy-number variation (i.e. large insertions/deletions). Currently, there is a fundamental lack of information about the baseline IGH immune repertoire V gene usage and diversity within healthy human controls. To provide an estimate of this, we sequenced functionally recombined gene segments to infer the underlying gene structure. From a set of 132 healthy controls we sorted C19+/CD27+ B-cells from whole blood and amplified genomic DNA using a highly multiplexed PCR assay that targeted the rearranged IGH receptor locus. Following DNA sequencing and data processing to assign V, D and J gene families and names, we examined the usage frequency of IGHV gene segments across all individuals. We found that of the 98 V gene segments only 56 (57%) were used at a frequency > 0.1%, and ~10 showed little to no usage (present in <1% of individuals). This data also allowed us to identify two IGHV genes currently annotated as orphons (pseudogenes assigned to an alternate chromosomal location) that had unambiguous functional usage (IGHV4/OR15-8; IGHV3/OR16-09) and therefore must reside at the IGH locus on chromosome 14. Finally, by taking this functional approach we were able to screen all V gene segments for germline copy-number variation (e.g. large insertion/deletion events encompassing individual genes) by looking for an excess of deletion events or modal changes in gene usage. We confirmed that existence of 12 of 15 previously identified deleted IGHV gene segments. Strong deletion evidence was observed for an additional six IGHV genes (IGHV3-NL1, IGHV3-33, IGHV1-24, IGHV4-04, IGHV3-41, IGHV3-35) and ten with highly likely germline deletion events. These data suggest that functional immune profiling of rearranged immune receptors provides a more robust method of identifying individual structural variation and provides insight into the immune repertoire of healthy controls.

551W

Copy Number Variants near SLC2A9 are Associated with Hyperuricemia. *R.B. Scharpf[1], L. Mireles[2], E. Halper-Stromberg[3], A. Tin[2], A. Chakravarti[4], E. Boerwinkle[5], J. Coresh[2], W.H.L. Kao[2].* 1) Oncology, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD; 3) Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 5) IMM Center for Human Genetics, University of Texas School of Public Health, Houston, Texas.

Hyperuricemia is associated with multiple diseases, including gout, cardiovascular disease, and renal disease. Serum urate is highly heritable suggesting a strong genetic component, yet genome-wide association studies of single nucleotide polymorphisms (SNPs) and serum uric acid concentrations explain only a small fraction of the heritability. Whether common copy number variants (CNVs) contribute to uric acid levels is not known. Here, we use high-throughput genotyping arrays to assess DNA copy number on a genome-wide scale among 9,738 individuals of European ancestry who participated in the Atherosclerosis Risk in Communities (ARIC) study. Loss of DNA copy number at genomic coordinates 10,002,252-10,009,766bp on chromosome 4p16.1 is associated with a 5.39 percent increase of uric acid concentrations among women (95% CI: 4.32-6.47, p = 4.4e−24) and a 1.39 percent increase among men (95% CI: 0.30-2.49, p = 0.012). The CNV locus is approximately 200kb from several SNPs in the urate transporter SLC2A9 that have been previously associated with uric acid concentrations in ARIC, including SNP rs7675964. Among women with the same rs7675964 genotype, loss of copy number is associated with a 3.01 percentage increase (95% CI: 1.96-4.07) of uric acid concentrations (p = 4.84e-33). In addition to variation of DNA sequence, loss of DNA copy number may contribute to the genetic predisposition of serum uric acid concentrations, particularly among women, explaining some of the missing heritability from standard genome-wide analyses with SNPs.

552T

Comprehensive comparison of copy number variations detection using Illumina Omni 2.5M and Affymetrix CytoScan® arrays. *C. TANG[1,2], E. WONG[1], H. GUI[1], S. CHERNY[1,4], P. SHAM[1,2,4,5], P. TAM[3,5], M. GARCIA-BARCELÓ[3,5].* 1) Department of Psychiatry; 2) Centre for Genomic Sciences; 3) Department of Surgery; 4) State Key Laboratory of Brain and Cognitive Sciences; 5) Centre for Reproduction, Development and Growth, the University of Hong Kong, Pokfulam, Hong Kong.

Structural variation has been recognized as a genetic risk factor contributing to human diseases, and in particular, congenital disorders. Smaller scale copy number variations (CNVs) have also been linked to a number of neurodevelopmental phenotypes, including intellectual disability as well as autism spectrum disorders. The precise detection of CNVs is therefore necessary for understanding disease pathogenesis. Recently, the new generation of SNP-based arrays, Affymetrix CytoScan® and Illumina Omni 2.5M offer an unique opportunity for improved discovery of CNVs with their special design. We explored the performance of these new platforms by genotyping in duplicate on each platform, 4 samples from patients diagnosed with a congenital disease. Performance of the CNV calling was assessed on the basis of sensitivity and specificity, both within and across platforms using various CNV detection software. Similar to previous generations of SNP-based genotyping arrays, the concordance of CNVs was found to be moderate and dependent on the calling software. In general, Cytoscan offered higher sensitivity whereas more specific calls were achieved using Omni. To conclude, multiple CNV calling methods should be employed for reliable CNV calling.

553F

Characterisation of the *RNU2* CNV, a bulky neighbour for *BRCA1*. *C. Tessereau[1, 2], N. Monnet[1], M. Imbert[1], M. Buisson[1], L. Barjhoux[1], C. Cuenin[7], C. Schluth-Bolard[3, 4], D. Sanlaville[3, 4], Z. Herceg[7], E. Conseiller[2], M. Ceppi[2], L. Duret[5], OM. Sinilnikova[1, 6], S. Mazoyer[1].* 1) Genetics of Breast Cancer, Cancer Research Center of Lyon, CNRS UMR5286/Inserm U1052/Université Lyon 1, Lyon, France; 2) Genomic Vision, Bagneux, France; 3) Service de Génétique, Laboratoire de Cytogénétique Constitutionnelle, Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, France; 4) INSERM U1028, CNRS UMR5292, Université Claude Bernard Lyon 1, Equipe TIGER, 69000 Lyon, France; 5) Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Université Lyon 1, CNRS, INRIA, UMR5558, Villeurbanne, France; 6) Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon, Centre Léon Bérard, Lyon, France; 7) Epigenetics Team, Molecular Carcinogenesis and Biomarkers Group, International Agency for Research on Cancer, F-69008, Lyon, France.

The question of the implication of multiallelic CNVs in complex traits remains largely open as most of them cannot be genotyped by array technology. In this work, we focused on the *RNU2* locus, a variable number of tandem repeats that contains the gene coding for the snRNA U2, an essential element of the splicing machinery. *RNU2* was shown many years ago to reside close to the breast cancer susceptibility gene *BRCA1* but is still missing from the latest human genome assembly and cannot therefore be investigated by recent genomic approaches. Using unassembled contigs, we precisely located *RNU2* within the chromosome 17 reference assembly, 124 kb telomeric of *BRCA1*. By FISH analyses on combed DNA (Molecular Combing), we determined more precisely the exact allelic number of repeats than with the previously used Pulse Field Gel Electrophoresis technique and found a range of 6-82 and a level of heterozygosity of 98% in 41 individuals. We used the 1,000 Genome Project data for analysing the variability of this macrosatellite by mapping sequence with unlocalized human genomic contigs and confirmed its high degree of polymorphism suggesting that depth-of-coverage calculation is a very useful tool for accurate multi-allelic CNV characterization. We found 24 frequent SNPs within the *RNU2* basic unit, and the genotype data for 1,106 individuals confirmed previous results showing a concerted evolution of this CNV. Thanks to our precise location, we were able to confirm that the *RNU2* array is within the *BRCA1* linkage disequilibrium block, which allowed us to study the *RNU2* array transmission over a large number of generations. A surprising resulting observation is that a highly variable locus can nevertheless be highly stable. Given the high level of polymorphism of this locus, we also measured the expression of U2 snRNA by qRT-PCR in 16 individuals carrying 36 to 110 repeats. Although the U2 level varied up to 6.6 times between individuals, it is not linked to the *RNU2* CNV copy number. By pyrosequencing, we found a higher level of methylation of the *RNU2* gene enhancer sequences in individuals with the highest copy number, suggesting that methylation could be involved in dosage compensation. These findings extend our knowledge of a recently neglected CNV that could be valuable for evaluating the potential role of structural variations in disease due to its location next to a major cancer susceptibility gene.

554W

Testing rare coding deletions identified using dense exome chip array data for contribution to type 2 diabetes. *M. Thurner[1], A. Mahajan[1], N. Robertson[1], A. Kumar[1], W. Rayner[1], F. Karpe[2], C. Palmer[3], T. Spector[4], M. McCarthy[1], K. Gaulton[1], GoT2D consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK; 3) Diabetes Research Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK; 4) Department of Twin Research, King's College London, London, UK.

Loss of function (LoF) of gene activity is a major contributor to phenotypic variability and contributes to both Mendelian and complex disease. The extent to which LoF events might influence susceptibility to type 2 diabetes (T2D), however, is currently unknown. In particular, partial or complete deletions of coding sequence leading to gene LoF have not been explored on a large scale. We thus investigated exonic deletions that could be discovered using the Illumnia HumanExome SNP array (consisting of >240,000 primarily exonic variants). We developed an analytical pipeline for deletion discovery from this array that applies several existing structural variant calling algorithms (PennCNV and QuantiSNP) and merges the resulting calls together to produce one set of calls per sample. We first applied this pipeline to exome chip data from 650 samples whole genome sequenced as part of the GoT2D project and for which genome-wide SV calls have been made using GenomeSTRIP. After filtering merged calls using strict quality thresholds removing samples with large variability and periodicity in signal intensity we detected 20 deletions. 15 of 20 calls (75%) were also detected in low pass sequencing with 100% specificity in genotyping at all but one call. We then applied the same procedure to exome chip data for 11,686 UK T2D case and control samples. A total of 208 deletions were identified 100% of which were low frequency (MAF <0.01) and 48% (100) were singletons. We tested these 208 variants for individual association to T2D at each single SV, and identified large deletions on chromosome 7 overlapping *RARRES2*, *REPIN1*, *LRRC61*, *C7orf29* and *ZNF775* (66,394bp, P=3.0×10−5) and on chromosome 15 overlapping *TSPAN3* and *PSTPIP1* (37,580bp ,P= 1.1×10−4) with significant T2D association after correction for multiple testing. Both are present in the Database of Genomic Variants (DGV), supportive of these being true deletion events. We further characterized 26 and 101 very rare deletions (allele count < 10) present only in cases and controls, respectively, and identified several in potential genes of interest such as *SLC30A8*. These results strongly suggest dense exome arrays can accurately identify gene deletions that can be assessed for contribution to disease. Continued refinement to calling and analysis procedures, coupled with validation and deeper phenotypic investigation of specific deletions of interest, will help clarify the role of coding deletions in T2D susceptibility.

555T

Multiplex emulsion haplotype fusion PCR to determine haplotypes at structurally complex regions. *J. Tyson, H.A. Black, J.A.L. Armour.* School of Life Sciences, University of Nottingham, Nottingham, United Kingdom.

Current human genome sequencing studies produce a vast amount of data that is generally phase insensitive. The importance of haplotypes is well documented, yet the reconstruction of maternal and paternal haplotypes over distances longer than a few kilobases is not trivial. Despite the progress made in both statistical and experimental determination of phase, problems remain in assembling (and reconstructing linear haplotypes in) regions of structural variation. Copy number variable regions (CNVRs), such as the *DEFA1A3* locus on chromosome 8p23.1, pose a particular challenge with respect to the determination of haplotypes. Regions with multiple copies of a highly similar sequence tend to be collapsed into a single copy on the genome assembly. As is the case for many CNVRs, multiple levels of complexity exist at the *DEFA1A3* region in the form of copy number variation, gene identity and sequence variation and, as such, determination of the true sequence haplotype in these regions is problematic. For many CNVRs, whilst the diploid copy number can be straightforwardly measured, haplotype copy number and positional information regarding the order of genes with the phase of SNPs either side of the CNVR is often overlooked. We have developed a multiplex emulsion haplotype fusion PCR (EHFPCR) approach to determine structural haplotypes and provide positional information about the location of the genes and associated sequence variants across the *DEFA1A3* CNVR. Amplicons of up to 1kb in length were designed both centromeric and telomeric to the CNVR, to fuse to amplicons within the CNVR. EHFPCRs were carried out in a multiplex reaction, with allele-specific PCR and sequencing used to detect the phase. Integration of sequence data led to the construction of a structural haplotype of approximately 59kb across the *DEFA1A3* CNVR in HapMap individuals, in which the relative positions of different gene sequence variants were defined. The spatial arrangement of genes within any CNVR is valuable in studying the relationship between gene copy number and gene expression, and the relationship between haplotype structure and expression. This work has enabled the reconstruction of haplotypes across the *DEFA1A3* CNVR, providing a more complete view of the haplotype structures at this locus. In addition, EHFPCR is applicable to other regions of the genome, when a more focused approach to assembling haplotypes may be required.

556F

Study of 455 molecularly unsolved LWD and ISS cases: identification of two deletions and the first duplication upstream of SHOX. *H. Verdin[1], L. Borms[1], E. Debals[1], B. D'haene[1], G. Matthijs[2], E. Maris[3], S. Depoorter[4], E. De Baere[1].* 1) Center for Medical Genetics, Ghent University, Ghent, Belgium; 2) Center for Human Genetics, University of Leuven, Leuven, Belgium; 3) Department of Pediatrics, Ghent University Hospital, Ghent, Belgium; 4) Department of Pediatrics, AZ Sint-Jan, Bruges, Belgium.

Short stature homeobox-containing gene (*SHOX*) is located in the telomeric pseudo-autosomal region (PAR1) on the short arms of the X- and Y-chromosome. This PAR1 region also contains seven enhancer elements, three located upstream and four downstream of *SHOX*. A heterozygous defect of *SHOX* or its enhancer elements can lead to Léri-Weill dyschondrosteosis (LWD) or idiopathic short stature (ISS). LWD is a dominant skeletal malformation syndrome characterised by disproportionate short stature, mesomelic limb shortening and the Madelung deformity. ISS is defined as a height below -2 SDS in the absence of known causative disorders. Mutations in *SHOX* or its downstream enhancer elements are found in 50 to 100 % of LWD cases, and in about 2 % of ISS cases. Only recently, the first upstream deletion in a patient with ISS has been described encompassing two of the three upstream enhancer elements. The objective of this study is to assess the contribution of upstream copy-number variations in a large population of molecularly unsolved LWD- and ISS-patients who have been pre-screened for *SHOX* and the downstream PAR1-region. The copy number of the three upstream (CNE-5, CNE-3 and CNE-2) enhancer elements was assessed for 455 patients using real-time quantitative polymerase chain reaction (qPCR) with in-house developed assays for each enhancer element. In 424 patients a normal copy-number was found for the three enhancers. An upstream deletion was found in two ISS patients. In the first patient a deletion of CNE-3 and CNE-2 was found while in the second patient CNE-5 and CNE-3 were deleted. Interestingly, the recently published upstream deletion comprises CNE-5 and CNE-3 which means that the shortest region of overlap for these three deletions only includes CNE-3. Even more remarkable is that we also found a duplication encompassing CNE-5, CNE-3 and CNE-2 in a patient with ISS. These observations therefore point to a possible crucial rule for CNE-3 as enhancer element in the regulation of *SHOX*. In the past, another upstream deletion of *SHOX* comprising the three enhancer elements has been described in a family with brachymesomelic dysplasia and Peters anomaly of the eyes. As our patients do not present with this severe phenotype it remains to be determined why a deletion of the same enhancer elements leads to a different phenotype. In conclusion, we report two upstream deletions and the first upstream duplication in ISS patients.

557W

Diversity of the human LILRB3/A6 locus encoding a myeloid inhibitory and activating receptor pair. *N. Vince[1,2], A. Bashirova[1,2], R. Apps[1], Y. Mochalova[3], X. Yu[2], M. Carrington[1,2].* 1) Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Ragon Institute of MGH, MIT and Harvard, Boston, MA; 3) University of Maryland, Baltimore County, Baltimore, MD.

Leukocyte Ig-like receptor B3 (LILRB3) and LILRA6 represent a pair of inhibitory/activating receptors with identical extracellular domains and unknown ligands. LILRB3 can mediate inhibitory signaling via ITIMs in its cytoplasmic tail whereas LILRA6 can signal through association with an activating adaptor molecule, FcRγ, which bears a cytoplasmic tail with an ITAM. The receptors are encoded by two highly polymorphic neighboring genes within the Leukocyte Receptor Complex (LRC) on human chromosome 19. Here we report that the two genes display similar levels of single nucleotide polymorphisms with the majority of polymorphic sites being identical. In addition, the *LILRA6* gene exhibits copy number variation (CNV) whereas *LILRB3* does not. A screen of healthy Caucasians indicated that 32% of the subjects possessed more than 2 copies of *LILRA6*, whereas 4% have only one copy of the gene per diploid genome. Analysis of mRNA expression in the major fractions of PBMCs showed that *LILRA6* is primarily expressed in monocytes, similarly to *LILRB3*, and its expression level correlates with copy number of the gene. We suggest that the *LILRA6* CNV may influence the level of the activating receptor on the cell surface, potentially affecting signaling upon LILRB3/A6 ligation.

558T

Highly variable tandem repeat genes: hotspots for primate evolution and human disease susceptibility. *C.T. Watson[1,2], D. Ho[2], M. Brahmachary[2], A. Guimatre[2], C. Borel[2], P. Warburton[2], A.J. Sharp[2].* 1) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Tandemly repeated (TR) genes can exhibit extreme variation in copy number and account for a significant portion of sequence variation between individuals, populations, and species. Targeted studies of TR genes have implicated roles for specific loci in adaptation and human disease risk; however, due to their multi-copy and structurally complex nature, TR gene polymorphisms are poorly captured by genome-wide technologies, and as a result remain understudied. Using a basic computational approach, we identified >180 RefSeq genes in the human genome with >1 copy at a non-overlapping interval on the same chromosome. These genes were enriched in poorly assembled regions, 33% mapping within 50 Kbp of an assembly gap, compared to only 1.6% of all RefSeq genes. To assess the extent of TR gene copy number variation in humans and related species, we designed custom locus-specific Nanostring assays for digital genotyping of copy number at each TR gene in 165 HapMap individuals (YRI;CHB;CEU) and five non-human primate (NHP) species. Secondary probes designed for several loci confirmed the precision of our assays, revealing high concordance in copy number estimates between probes (R^2>0.75). From this screen, we found that the majority of TR genes (~70%) were copy number variable in humans. Inter-individual variation was as high as 5-10 fold, and many genes exhibited significant stratification between populations (V_{st}>0.2). Similarly, we observed a high degree of variation within NHPs, and in many instances, the degree of copy number polymorphism showed species-specific patterns; for example, *REXO1L2P* has 100-250 copies in humans, whereas greater than 900 copies were observed in Gorilla. In contrast, the *GAGE* gene was in single copy in NHPs, but ranged between 15-60 copies in humans. Furthermore, TR genes had significantly higher rates of non-synonymous amino acid changes (P=3.3×10^{-7}). We also uncovered features of TR genes that are important in the context of human disease, in particular, most show very low levels of linkage disequilibrium with flanking SNPs, with only 2 TR genes tagged by neighboring SNPs at r^2≥0.8. Together, these findings demonstrate that previous GWAS have had limited power to assess the impact of TR gene variation in disease susceptibility. To address these shortcomings, we are currently screening 26 of the most polymorphic TR genes in populations with various autoimmune diseases for which a majority of disease variance remains uncharacterized.

559F

Detection of CNV gains and losses with Affymetrix® Axiom® arrays. *T. Webster, H. Zuzan, J. Gollub, J. Schmidt, A. Roter.* Algorithms & Data Analysis, Affymetrix, Santa Clara, CA.

Copy number variations (CNVs) polymorphisms are common in phenotypically normal individuals and some have been shown to increase risk for human diseases, for example schizophrenia and other neuropsychiatric disorders. The ability to detect single nucleotide polymorphisms (SNPs) as well as copy number variation in the same genome screen is efficient and advantageous. We show that Affymetrix® Axiom® arrays, designed to detect genome-wide associations with SNPs and InDels, also detect copy number variations. The CNV detection method computes log2(ratios) for each marker, where for each marker site the sum of allele intensities from an individual is normalized by the intensity produced by taking the average of several genomes from phenotypically normal individuals, assumed to represent the normal copy number state. While genotypes are determined by analyzing a large set of samples at a single location, achieving an accuracies of greater than 99.5%, CNVs are evaluated from the same data set by analyzing a contiguous set of markers along the genome in a single sample. The method was applied to whole genome data from a set of HapMap individuals. We present evidence of CNV gains and losses in HapMap individuals that were genotyped with Axiom® arrays and by literature evidence and established CNV detection technologies, such as the Affymetrix CytoScan® HD microarray.

560W

Fusion genes resulting from complex duplications in chromosome Xq28. *L.W. Zuccherato[1], B. Alleva[1], C.M.B. Carvalho[1], J.R. Lupski[1,2,3].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Changes in gene dosage likely are responsible for many genomic disorders. However, advances in technology are revealing that complex genomic rearrangements (CGRs) can lead to the formation of new genes that have potential effects in the disease phenotype. Chromosomal rearrangements can be produced by recombination and replication mechanisms and can convey diverse phenotypic effects. Gene fusions have been observed in somatic alterations such as cancer, and despite the importance of these fusion genes, the underlying mechanism(s) that result in their formation in constitutional chromosome alterations and other rearrangements are poorly understood. Data from the breakpoint junction sequencing of two individuals carrying duplications involving the Xq28 region enabled us to predict a complex pattern of rearrangements that include the fusion of exons from genes *F8/CSAG1* and *BCAP31/TEX28*. RT-PCR experiments confirmed the expression of the new fusion genes in transformed lymphoblastoid cell lines from these patients. The rearrangement found in the *F8/CSAG1* produced an inversion of the 5'UTR of the *F8* gene, leading to a transcript in the same orientation as transcript *CSAG1*. Moreover, the newly expressed *F8/CSAG1* transcripts included a partial sequence of ERVL-MaLR intronic repetitive element. This provides evidence that the 'exonization' of repetitive elements can occur in the formation of a fusion gene and may be triggered by changing the genomic context in which a gene is transcribed. Future experiments will address the functional relevance of these changes at the protein level and the further impacts of structural variation on patient phenotypes. Thus, complex rearrangements mediated by the replication mechanism FoSTeS/MMBIR may contribute to exon shuffling processes and diversify the repertoire of expressed transcripts, and therefore CGRs may have a role in the evolution of both individual genes as well as the human genome.

561T

Mapping of a human genome with a single molecule nanochannel array platform for genome-wide structural variation analysis and de novo sequence assembly of next-generation sequence reads. *Y.Y.Y. Lai[1], A.C.Y. Mak[1], E.T. Lam[2], J. Silbert[3], T.P. Kwok[4], J.W. Li[4], A.K.Y. Leung[4], J.J.K. Wu[5], A.K.Y. Yim[4], A. Poon[1], C. Chu[1], C. Lin[1], M. Requa[2], A. Hastie[2], T. Anantharaman[2], H. VanSteenhouse[2], H. Dai[2], F. Trintchouk[2], M. Saghbini[2], M. Austin[2], K. Haden[2], H. Cao[2], S.M. Yiu[5], K.Y. Yip[4], T.F. Chan[4], M. Xiao[3], P.Y. Kwok[1].* 1) University of California, San Francisco, San Francisco, CA, United States; 2) BioNano Genomics, San Diego, CA, United States; 3) Drexel University, Philadelphia, PA, United States; 4) Chinese University of Hong Kong, Shatin, Hong Kong; 5) University of Hong Kong, Hong Kong, Hong Kong.

Genetic variation results in human population diversity and differential disease susceptibility. Different methods, such as quantitative PCR, array comparative genomic hybridization (aCGH), Sanger or next-generation sequencing have been used for identification of these polymorphic loci, however they have either low sensitivity in repetitive regions, are too labor-intensive and time consuming, or do not provide positional information about the structural variants. To overcome the above disadvantages from current methods, we use a single-molecule mapping approach to identify structural variation as well as to provide scaffolds for de novo assembly of a diploid human genome, NA12878, from 'short reads' sequences. Genome mapping utilizes highly parallel nanochannel arrays in which thousands of very long single DNA molecules are linearized and imaged. This novel approach is automated on the Irys System, which can scan the entire genome rapidly to generate physical maps that provide a more comprehensive view of the genome. NA12878, the daughter in a CEPH-CEU trio, was used as this sample has been genotyped and sequenced extensively as part of the International HapMap and 1000 Genomes Projects. We nicked and fluorescently labeled DNA fragments, size ranging from 100 kb - 500 kb, at Nt. BspQI (GCTCTTCN/) sites. To date, we generated over 50X coverage data and constructed de novo assembled genome maps that cover more than 90% of the genome using an automated assembly pipeline. We identified many structural variants (indels) including those previously known in this sample. Overall, this genome mapping approach is simple and can be performed in any modern molecular genetics laboratory.

562F

CNP imputation using 1000 Genome Project data as reference panel and intensity based analysis of copy number variation in African Americans.
Y. Meng[1], J. Nemesh[2], D. Altshuler[3], EJ. Benjamin[4], E. Boerwinkle[5], D. Bowden[6], CWK. Chiang[7], M. Fornage[8], J. Glessner[9], A. Kutlar[10], G. Lettre[11], M. Li[12], S. Musani[13], G. Papanicolaou[14], S. Redline[15], A. Reiner[16], S. Rich[17], D. Siscovick[18], X. Zhu[19], H. Hakonarson[20], JG. Wilson[13], B. Keating[21], JN. Hirschhorn[1,3,22], SA. McCarroll[3,22]. 1) Metabolic Disease Initiative, Broad Inst, Cambridge, MA; 2) Computational Biology and Bioinformatics Program, Broad Inst, Cambridge, MA; 3) Medical and Population Genetics Program, Broad Inst, Cambridge, MA; 4) School of Medicine, Boston University, Boston, MA; 5) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 6) Wake Forest University School of Medicine, Winston-Salem, NC; 7) Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 8) Institute of Molecular Medicine, University of Texas, Houston, TX; 9) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 10) Department of Medicine, Medical College of Georgia, Augusta, GA; 11) Department of Medicine, Université de Montréal, Montreal Heart Institute, Quebec, Canada; 12) Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA; 13) University of Mississippi Medical Center, University of Mississippi Medical Center, Jackson MS; 14) Division of Prevention and Population Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD; 15) Department of Medicine and Center for Clinical Investigation, Case Western Reserve University, Cleveland, OH; 16) Department of Epidemiology, University of Washington Seattle, WA; 17) Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 18) Division of Cardiology, Department of Medicine, University of Washington, Seattle, WA; 19) Department of Biostatistics and Epidemiology, Case Western Reserve University, Cleveland, OH; 20) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA; 21) Department of Cardiology, University of Pennsylvania, Philadelphia, PA; 22) Department of Genetics, Harvard Medical School, Boston, MA.

Copy number variants (CNVs) are important source of variation in the human genome, but can be challenging to assay directly with current chip-based genotyping methods used for genome-wide association studies (GWAS). Therefore, there is a need to improve our ability to assess this class of variation. The release of Phase I of the 1000 Genomes Project (1KGP) has catalogued a large number of both SNPs and CNVs. One of the major uses of the SNP data has been imputation of genotypes at locations not directly genotyped in GWAS conducted on conventional SNP array. Because common copy number polymorphisms (CNPs) are often in linkage disequilibrium with neighboring SNPs, it is possible to impute common CNPs using SNP data. We evaluated the feasibility of imputation of 1KGP deletion CNVs in a large panel of 1845 African Americans from Jackson Heart Study (JHS) within CARe (Candidate Gene Association Resource) project. We used the IKGP Phase 1 data as a reference panel, and cross-validated imputation with directly called CNPs, comparing imputed genotypes with those directly measured on the array using Birdsuite software (Canary for known common CNPs and Birdseye for rare CNVs). We evaluated several metrics for imputation. First, reachability: how many actual 1KGP CNVs for a known sample (CEU HapMap) were imputed. Second, stability: how often a CNV was assigned the same copy number across 24 sample replicates. Third, accuracy of CNP imputation in JHS samples. The reachability analysis showed that imputation had highest, and Canary intermediate, and Birdseye the lowest reachability. The stability analysis showed that the imputation was the most stable among the 24 replicates, with almost 100% stability. The accuracy analysis showed that CNPs were imputed with accuracy similar to that of SNPs at MAF≥0.05, with less accuracy at MAF<0.05 (dosage r^2 of 0.65 vs. 0.8 with beagle r^2>0.4). Moving forward, we are cataloguing CNVs in the full CARe data set, comprising more than 8000 African American individuals from the Atherosclerosis Risk in Communities (ARIC) study, Cleveland Family Study, Coronary Artery Risk Development in Young Adults (CARDIA) study, JHS, and the Multi-Ethnic Study of Atherosclerosis (MESA). We will also generate a large catalogue of directly typed CNV genotypes in African Americans who also have SNP genotype data. This catalogue will be useful for imputation in west African-ancestry samples and for testing association to phenotypes.

563W

Haplotype imbalance reveals subtle genomic mosaicism in blood and cell line samples.
S. Vattathil[1,2], X. Xiao[2], P. Scheet[2,1]. 1) Human and Molecular Genetics Program, The University of Texas at Houston Graduate School of Biomedical Sciences; 2) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston.

Genomic mosaicism resulting from proliferation of somatically-mutated cells is known to occur in ostensibly normal individuals, however the prevalence and consequences of mosaicism remain unknown. The challenge is that, since somatic mosaicism is infrequent in the population and clonal mutations may be rare within any mosaic individual, methods must be both efficient enough to apply to many samples and sensitive enough to detect minor cell populations. Recent large-scale analyses (Laurie ***et al.***, *Nat Gen* **44**:642; Jacobs ***et al.***, *Nat Gen* **44**:651) have used SNP microarray data to detect somatic mosaicism and outline its relationship to age and cancer incidence, but acknowledge limited power of detection due to difficulties created by noisy data, especially when the clonal population is small. We have previously described a method for combining individual haplotype estimates and B allele frequencies from SNP microarrays to localize allelic imbalance existing in minor cell populations. The advantage of the haplotype-based approach is that it is more robust to stochastic noise inherent to array data, and therefore more sensitive to mutations existing in very low proportions (<10%) of the sampled cells. We are applying the method to a large set of publicly available samples collected by the GENEVA Consortium. Preliminary results from 4,121 blood and blood-derived cell line samples genotyped on the Illumina Human1M microarray identify 1003 mosaic events, including recurrent somatic variation on chromosome 22q in 0.2% (7/2,801) of blood samples and 11.7% (155/1,320) of cell line samples that corresponds to recent findings of recurrent deletion in the region but indicates higher recurrence rate than previously reported. While these events have strong phase concordance signal from our method, some of them may have been missed in other analyses because stringent filtering was required to compensate for low noise tolerance of the detection methods used.

564T

Genome-wide bioinformatic analysis demonstrates distinct *Alu* elements predispose specific loci to pathogenic structural rearrangements.
I.M. Campbell[1], C.R. Beck[1], P.M. Boone[1], C.A. Shaw[1], J.R. Lupski[1,2,3]. 1) Molecular & Human Genetics, Baylor College Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Although the genomic features associated with large recurrent copy number variations (CNVs) of the human genome are well studied, the sequence characteristics that predispose to smaller rearrangements remain poorly understood. *Alu* elements are non-autonomous retrotransposons that have accumulated during primate evolution to make up approximately 8% of the reference human genome. *Alu*s are associated with genome instability and are enriched at the breakpoints of genomic deletions and duplications. A distinct subset of generally small CNVs are mediated by directly-oriented *Alu* elements aligned at the CNV breakpoints. Such CNVs are a frequent cause of some human diseases, including inactivating mutations of LDLR, SPAST and VHL. Although the *Alu* elements present at the breakpoints of these CNVs are aligned, they are much shorter and less similar than low-copy repeats that mediate non-allelic homologous recombination. Thus, the mechanism and genomic features that promote this type of CNV remain largely unknown. We hypothesized that sequence features and distribution of *Alu* elements predispose some loci to *Alu-Alu* mediated CNVs. To test this, we analyzed *Alu* elements participating in such CNVs using breakpoints determined in our laboratory at the SPAST locus as well as from the literature. We determined that elements participating in *Alu-Alu* CNVs are longer, have more conserved Pol III binding sites, have longer poly-A tails, and are richer in GC content than the genome wide average. Analyses suggest these differences cannot be explained by element length alone. We applied knowledge of the characteristics of *Alu*s that participate in CNVs genome wide to identify loci with large numbers of directly oriented exon-flanking pairs matching the characteristics of *Alu-Alu* mediated CNVs. We cross-referenced this list of loci with potentially clinically relevant CNVs identified in over 25,000 patients from our diagnostic laboratory. We discovered a significant enrichment in small (<250 kb) genomic deletions involving the genes with the largest numbers of these distinct *Alu-Alu* pairs. Notably, a number of the genes identified as both enriched in the distinct *Alu-Alu* pairs and with small CNVs in our database were previously reported to undergo *Alu-Alu* recombination. These data help define repetitive elements participating in copy number variation, elucidate mechanisms underlying these rearrangements, and may allow prediction of disease causing CNVs at a genomic level.

565F

Comprehensive analysis of polymorphic numt insertions in Human and Primate populations. *G. Dayama[1], S.B. Emery[2], J.M. Kidd[1,2], R.E. Mills[1,2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

The transfer of mitochondrial genetic material into the nuclear genomes of eukaryotes is a well-established phenomenon. Many studies over the past decade have utilized reference genome sequences of numerous species to characterize its prevalence and also its role in human diseases. The recent advancement of high throughput sequencing technologies has enabled the interrogation of genomic variation from whole genome sequence data at a much finer scale, resulting in high quality maps of SNP, INDEL, and copy number variation. In contrast, there has been little focus on the diversity of nuclear mitochondrial insertions (numts) in human populations. To address this deficiency, we have developed an approach to discover and genotype previously undiscovered numt insertions using whole genome, paired-end sequencing data. We have identified more than 150 novel sites of numt insertions by applying our method to over a thousand individuals in twenty populations from the 1000 Genomes Project and other datasets. This expands our current knowledge of existing numt locations in the human genome by 20% and represents a substantial increase from the 14 polymorphic numt loci previously reported. Over 90% of the newly identified numts were found in less than 1% of the samples we examined to date, suggesting that they occur infrequently in nature or have been rapidly removed by purifying selection. We further extended our analysis to the whole genomes of over sixty non-human primates belonging to four genera and collectively examined these variants for biases in sequence context at insertion sites and their potential for functional effects. We believe this research will help clarify to what extent that numts play a role in hominoid diversification and phenotypic variability.

566W

An Estimate of Effective Population Size from Individuals of Various Populations. *H.R. Johnston IV.* Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA.

The advent of whole genome sequencing has allowed for population genetic analyses, previously only theory, to be applied to the data they were designed for. In particular, genome sequences from multiple populations across the globe allow for inferences to be made about historical population events. Utilizing a method originally proposed by RC Griffiths(1), and later leveraged by Johnston and Cutler(2), the effective population size of the population an individual is drawn from can be calculated. In this approach, the heterozygous SNPs of an individual, being a natural sample of size two, are binned over short distances, with the resultant distribution used to infer effective population size. Individuals from different populations are then compared to determine whether any patterns in effective population size can be discerned. Preliminary data, based on a single individual per population, shows that the effective population size for an African population is 8000, while a European population is 7625, a Chinese population is 6004 and a Korean population is 5362. These results need to be fleshed out using multiple individuals to limit the inherent error in these estimates, but the current result is as expected. The effective population size in African population is larger than those of other populations. A next step will involve analyzing sequences of individuals within multiple different populations from each continent, allowing for a finer analysis of demographic history. 1.Griffiths R (1981) Neutral two-locus multiple allele models with recombination. Theoretical Population Biology. 2.Johnston Henry R & Cutler David J (2012) Population Demographic History Can Cause the Appearance of Recombination Hotspots. The American Journal of Human Genetics 90(5):774-783.

567T

Genome Wide Associated Variants in Migraine Susceptibility: A Replication Study from North India. *B. Mittal[1], J. Ghosh[1], S. Pradhan[2].* 1) Genetics, SGPGI, Lucknow, UP, India; 2) Neurology, SGPGI, Lucknow, UP, India.

Objective: Genome wide association studies (GWAS) have identified various migraine susceptibility variants. We aim to replicate five GWAS associated polymorphisms (rs1835740, LRP1 rs11172113, TRPM8 rs10166942, PRDM16 rs2651899 and TGFBR2 rs7640453) in North Indian population. Furthermore, we checked the SNPs in strong linkage disequilibrium (LD) with the selected variants. We also undertook to predict the functional effect (in silico) of the variants. Design: The study included 340 migraineurs and 200 controls. Genotyping was performed by PCR-RFLP, ARMS PCR and Taqman. Logistic regression was used for association analysis. LD plot was prepared using genotyping data retreived from ENCODE and HapMart. Functional effect was predicted by F-SNP and FastSNP. Results: We did not observe any significant effect of the variant genotype or allele of the first migraine GWAS associated marker, rs1835740. However, significance was observed in case of heterozygous genotype for total migraineurs and in few subgroups: migraine without aura (MO) and females. We suggest potential protective effect of LRP1 rs11172113 polymorphism in migraine susceptibility especially in males. A marginal association was observed for TRPM8 rs10166942 variant in total as well as female migraineurs. PRDM16 rs2651899 variant genotype and allele showed a protective effect on migraine and MO susceptibility On the other hand, TGFBR2 rs7640543 variant did not have significant influence on migraine susceptibility in North Indian population. Most of the selected SNPs (except LRP1 rs11172113) and some of the SNPs in strong LD were predicted to affect transcriptional regulation. Functional effect of LRP1 rs11172113 could not be found but SNP in strong LD with it was found to affect transcription factor binding sites. Conclusion: We present the first replication study of GWAS associated polymorphisms in a population other than European.

568F

Truncating mutations in protocadherin 15 exon 33 which encodes the cytoplasmic domain are unlikely to be disease-causing. *C.L. Perreault-Micale, N. Chennagiri, C.J. Kennedy, A. Frieden, D. Neitzel, N. Faulker, S. Hallam, V. Greger.* Good Start Genetics®, Inc., Cambridge, MA, USA.

The protocadherin 15 gene (PCDH15) encodes a calcium-dependent cell adhesion protein important in mechanosensory transduction. Loss of function mutations in PCDH15 cause Usher syndrome type 1F, an autosomal recessive disease characterized by profound congenital hearing loss, vestibular dysfunction and retinitis pigmentosa. Usher syndrome 1F is one of the disorders that show increased incidence in the Ashkenazi Jewish (AJ) population, with the founder mutation R245X accounting for 75% of alleles. In order to determine the mutation spectrum in a pan-ethnic population, we sequenced the coding region and intron-exon borders of PCDH15 using next generation DNA sequencing technology in 8456 individuals referred by IVF clinics across the US. All mutations were confirmed using Sanger sequencing. We identified 5 carriers of 3 previously described pathogenic mutations (R245X, R643X, and R929X). We also found 11 novel truncating variants in 28 individuals. The novel mutation R1106X is located in exon 25 (transcript variant C, NM_033056.3), upstream from the most 3' known pathogenic variant (c.4257delA) in exon 32 and is predicted to be pathogenic since loss of this region of PCDH15 has already been associated with disease. The remaining 10 novel truncating mutations are clustered in exon 33, which encodes most of the C-terminal cytoplasmic domain. One of those variants, c.4831_4834dupAACA, was observed 16 times. 12 carriers of this duplication were self-reported African Americans (out of a total of 599 tested), resulting in a minor allele frequency of 1% in this population. The high incidence of truncating variants observed in exon 33 in our population indicates that exon 33 variants are unlikely to cause the severe phenotype associated with Usher syndrome type 1F. PCDH15 has a large number of transcript variants, with at least 3 alternate exons coding for the cytoplasmic domain. The truncating exon 33 variants are probably tolerated because differentially spliced cytoplasmic domain isoforms can function redundantly. It is widely accepted that truncating variants disrupt gene function, and thus are expected to be disease-causing in genes underlying recessive (loss of function) disorders. However, this example provides an exception to this rule. Deep sequencing of pan-ethnic populations has enabled us to address these key assumptions regarding the functional effects of some PCDH15 truncating mutations.

569W

Rapid Genome Wide Mapping at Single Molecule Level Using Nano-Channel Array for Structural Variation Analysis and *De Novo* Assembly.
H.C. Cao[1], A.H. Hastie[1], E.L. Lam[1], H.D. Dai[1], T.A. Anantharaman[1], M.X. Xiao[3], P-Y.K. Kwok[2]. 1) BioNano Genomics Inc, San Diego, CA; 2) UCSF, San Francisco, CA; 3) Drexel University, Philadelphia, PA.

Despite continued cost reduction in raw base generation, improvement in base-calling accuracy, and recent advances in read length, complete *de novo* assembly with accurate genome wide structural variant (SV) analysis of an individual large complex genome remains expensive and challenging. In particular, many disease relevant SVs up to hundreds of kilobases long in the human genome are severely underestimated due to a lack of effective tools. We present a rapid genome-wide analysis method based on new NanoChannel Array technology (IrysTM System) that temporarily confines and linearizes extremely long DNA molecules for direct image analysis at tens of gigabases per run. This high-throughput platform automates the imaging of genomic DNA hundreds to thousands of kilobases in length at single-molecule level, retaining long-range haplotypes. High-resolution genome maps assembled *de novo* via unique sequence motif labeling preserves native large and small structural variation information (especially highly repetitive regions), which are intractable with current short read NGS platforms. This information is collected independently of sequencing methods and is very valuable to identify structural variants as well as to validate and finish sequencing assemblies. Here we report the complete *de novo* assembly and analysis of complex regions and whole genomes of several human samples (including a cancer genome) with this approach. Unlike inference from mate-pair library sequencing approaches, hundreds of large structural variants were uncovered without apparent bias (e.g., size or insertion vs deletion) due to its more direct visualization and measurement. We have corrected errors in previous assemblies, spanned and sized many of the remaining gaps, identified known and novel structural variants and phased haplotype blocks, including in the highly variable complex regions related to human immune system functions. We have also discovered abundant previously unknown highly complex large repetitive patterns (greater than 2kb and inverted) spanning large regions of genome and pinpointed foreign genomic components inserted within the host human genome, important for understanding disease and oncogenesis. Widespread use of this technology will continue to enable new genome discoveries, expand our view of genome architecture, and improve understanding of functional regions.

570T

Describing translocations using HGVS sequence variation nomenclature, suggested extensions. *P. Taschner, J. den Dunnen.* Dept Human Genetics, Leiden Univ Medical Ctr, Leiden, Netherlands.

Translocation breakpoints are traditionally described using ISCN nomenclature based on chromosomal banding patterns (1). Due to the limited number of translocation breakpoint sequences identified the sequence variation nomenclature guidelines of Human Genome Variation Society (HGVS, http://www.hgvs.org/mutnomen), which are mainly focused on simple variants, did not require specific rules for detailed description of genetic rearrangements. This changed with the introduction of new technologies allowing rapid discovery of breakpoint sequences from complex structural rearrangements including translocations. The description of such complex variants challenges the existing guidelines. Here, we suggest extending the HGVS nomenclature guidelines to facilitate unambiguous description of translocations and their breakpoints. A main feature of the description is that the precise chromosomal breakpoint can be derived easily. The suggested format should provide sufficient flexibility and consistency limiting alternative interpretations and ambiguous descriptions. The new translocation rules can be combined with those proposed previously for complex changes, which included: i) nesting to support description of changes within inversions and duplications, ii) composite changes to support concatenation of inserted sequences (2). We have applied the rules in practice by describing all translocation breakpoints involving the DMD gene (see DMD gene variant database, http://www.lovd.nl/DMD). The specifications should allow easy implementation in sequence variant nomenclature checkers (e.g. Mutalyzer, https://Mutalyzer.nl). We are planning to extend the functionality of Mutalyzer to incorporate the latest version of the HGVS sequence variation nomenclature guidelines as part of the development of curational tools for gene variant databases (Locus-Specific DataBases, LSDBs). We believe that the HGVS translocation description format is easier to understand and more effective in practice than the equivalent descriptions in the VCF4.1 format, which was designed to describe variant data from next generation sequencing experiments (3). 1) ISCN (2013). 2013. An International System for Human Cytogenetics Nomenclature. Shaffer LG, McGowan-Jordan J, Schmid M (eds). Basel: Karger. 2) Taschner PE, den Dunnen JT. Hum Mutat. 32:507–511 (2011). 3) http://www.1000genomes.org/wiki/Analysis/Variant%20Call%20Format/vcf-variant-call-format-version-41.

571F

Common variation in the Melanocortin 4 Receptor gene (MC4R) is associated with increased food intake and obesity in American Indians.
Y.L. Muller, R.L Hanson, M. Thearle, D. Hoffman, B. Gene, K. Huang, S. Kobes, S. Votruba, J. Krakoff, W.C. Knowler, C. Bogardus, L.J. Baier. Diabetes Molecular Genetics Section, National Institutes of Health, Phoenix, AZ.

We previously identified 10 rare non-synonymous mutations in the Melanocortin 4 Receptor (MC4R) gene by sequencing 7900 American Indians living in the Gila River Indian Community of Arizona, where most of the residents are of Pima Indian heritage. The overall prevalence of MC4R functional coding mutations in this population was 2.4%. In the current study, we investigate whether common non-coding variants in MC4R could also contribute to risk of obesity in American Indians. Single nucleotide polymorphisms (SNPs; N=63) spanning a ~414 kb region encompassing MC4R were genotyped in 5880 American Indians who had longitudinal measures of body mass index (BMI). Most of these individuals had a measure of maximum BMI from a non-diabetic exam during childhood (maximum z-score, age and sex adjusted, between ages of 5 and 20 years) and had a measure of maximum BMI from a non-diabetic exam during adulthood (maximum BMI after the age of 15 years). A subset of these individuals had been studied as inpatients in our Clinical Research Center and had measures of body composition as assessed by dual x-ray absorptiometry (N=415) and ad libitum food intake measured over 3 days by an automated vending machine system (N= 203). A promoter variant (rs8097783) with a risk allele frequency (RAF) of 0.48 was associated with BMI in childhood (p=0.01 for age and sex adjusted Z score) and adulthood (p=0.0007, adjusted for age, sex, birth year and heritage). Another promoter SNP (rs1187299; RAF= 0.94) was also associated with BMI in childhood (adjusted p=0.005) and adulthood (adjusted p=0.06), where the risk allele was further associated with a higher percentage of body fat (p=0.01, adjusted for age, sex and family membership) and an increase of 640 kcal/day in total food intake (p= 0.003, adjusted for age, sex, percent body fat, family membership and heritage). The rs11872992 promoter variant was predicted to cause a loss of function by the Ingenuity Variant Analysis. *In vitro* luciferase assay of rs11872992 found that the risk allele had a modest decrease in promoter activity (p=0.005). We conclude that both common and rare variation in MC4R can influence food intake and contribute to risk of childhood and adulthood obesity in American Indians.

572W

Next-generation sequencing of complete mitochondrial genomes of Slovenian Lebers's Hereditary Optic Neuropathy patients revealed one novel mutation and several probable synergistic variants. *D. Glavac[1], M. Tajnik[1], M. Jarc-Vidmar[2], M. Hawlina[2].* 1) Department of Molecular Genetics, Faculty of Medicine, University of Ljubljana, Korytkova 2, SI-1000 Ljubljana; 2) Eye Hospital, University Medical Centre, Grabloviěeva 46, SI-1000 Ljubljana.

Purpose. Leber hereditary optic neuropathy (LHON) is maternally inherited eye disorder. It results from point mutations in highly polymorphic mitochondrial DNA (mtDNA). Although the LHON in the majority of the patients is a result of one of the three most common mutations, in some cases the genetic background is not clear. We investigated nine Slovenian patients diagnosed with LHON, from which only two were positive for one of the most common mutations. In order to find novel pathogenic variants, we performed deep sequencing of whole mtDNA. Methods. DNA of nine LHON patients and 2 controls was extracted from whole blood samples. Patients were first screened for most common LHON mutations using Sanger sequencing. In addition, for 6 patients and 2 controls we performed whole mtDNA amplification and deep sequencing using Ion Torrent technology. Results. Two patients were positive for T14484C and G3460A mutations. In the other patients, whole mtDNA deep sequencing detected novel homo and heteroplasmic variations. We identified 25 non-synonymous and 36 synonymous substitutions in mtDNA protein-coding regions. Their impact on protein structure and function was determined using bioinformatic prediction tools. We found 16 novel non-synonymous LHON-associated variants, from which 11 were homoplasmic and 5 were heteroplasmic. Two patients had significant visual improvement, one with concomitant improvement of VEP amplitudes. Conclusions. In early stage of LHON, before optic disc pallor develops, VEP P100 was abnormal in all, while PERG N95 was attenuated only in some patients. The prevalence of most common LHON mutations in the Slovenian patients is lower than in other parts in Europe. Using deep sequencing approach, we identified new potentially pathogenic mtDNA variations in Slovenian LHON patients, which do not harbour one of the three most common mutations.

573T

Genetic diversity in black South Africans from Soweto. *M. Ramsay[1, 2], S. Hazelhurst[3], Y. Li[4], M. Waldvogel[4], J. Eichenberger[4], S.A. Norris[5], M. Govind[5], M. Tikly[6], C. Hon[4], K.J. Johnson[4], N. Hartmann[4], F. Staedtler[4], A. May[1, 2].* 1) Division of Human Genetics, University of the Witwatersrand, Faculty of Health Sciences, Johannesburg, South Africa; 2) National Health Laboratory Service, Johannesburg, South Africa; 3) Wits Bioinformatics, University of the Witwatersrand, Johannesburg, South Africa; 4) Novartis Institutes for Biomedical Research (NIBR), Human Genetics and Genomics, Basel, Switzerland or Cambridge, MA, USA; 5) MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 6) Division of Rheumatology, Chris Hani Baragwanath Hospital and the School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Background: Due to the unparalleled genetic diversity of its peoples, Africa is attracting growing research attention. Several African populations have been assessed in global initiatives such as the International HapMap and 1000 Genomes Projects. Notably excluded, however, is the southern Africa region, which is inhabited predominantly by south eastern-Bantu speakers, currently suffering under the dual burden of infectious and non-communicable diseases. Limited reference data for these individuals hampers medical research and prevents thorough understanding of the underlying population substructure. Here, we present the most detailed exploration, to date, of genetic diversity in 94 unrelated south eastern Bantu-speaking South Africans, resident in urban Soweto (Johannesburg). Results: Participants were typed for ~4.3 million SNPs using the Illumina Omni5 beadchip. PCA and ADMIXTURE plots were used to compare the observed variation with that seen in selected populations worldwide. Results indicated that Sowetans, and other south eastern Bantu-speakers, are a clearly distinct group from other African populations previously investigated, reflecting a unique genetic history with small, but significant contributions from diverse sources. To assess the suitability of our sample as representative of all Sowetans, we compared our results to a larger control and case sample selected for a rheumatoid arthritis study. The control group showed good clustering with our sample, but among the cases were individuals who demonstrated notable admixture of possible clinical relevance. Conclusions: Our data represent a suitable reference set for south eastern Bantu-speakers, on par with a HapMap type reference population, and constitute a prelude to the Southern African Human Genome Programme.

574F

The African Genome Variation Project Phase II: Detailed characterisation of genetic variation in Sub-Saharan Africa. *D. Gurdasani[1,2], T. Carstensen[1,2], African Genome Variation Project Phase II Investigators.* 1) Public Health & Primary Care, Univ Cambridge, Cambridgeshire, United Kingdom; 2) Wellcome Trust Sanger Inst, Cambridge, Cambridgeshire, United Kingdom.

Background: The African Genome Variation Project (AGVP) is an international collaboration that aims to produce a comprehensive catalogue of human genetic variation in Sub-Saharan Africa (SSA) in order to extend our understanding of population history, genetic diversity among populations in SSA, as well as provide a global resource to help design, implement and interpret genomic studies in SSA. Phase 1 of the AGVP involved genotyping of 100 unrelated individuals from each of 16 ethno-linguistic groups from SSA with the 2.5 million (M) Illumina genotype arrays. Here, we present preliminary data from phase II of the AGVP. This project specifically aims to extend AGVP phase I by whole genome sequencing (WGS) of 2000 individuals from genetically diverse populations within Africa in order to develop a resource that provides a comprehensive catalogue of genetic variation in SSA along with core phenotypes accessible to the global scientific community. Methods: AGV phase II has completed low coverage (4x) WGS of 3 geographically distinct populations from Ethiopia, South Africa and Uganda. By October 2013, we aim to extend sequencing to include 1000 individuals from distinct populations including Khoe-San groups in Namibia. WGS was carried out using Illumina Hiseq 2000. Read alignment, recalibration and genotype calling was carried out using GATK best practice recommendations. Results: We identified 19M, 19.2M and 18.8M variants across the whole genome among Baganda (100 samples), Ethiopia (120 samples) and Zulu (100 samples) population groups. Of these, 10%, 18.1% and 9% of variants were novel, respectively. Consistent with previous literature, non-synonymous variants were enriched in the lower allele frequency spectrum, with the ratio of non-synonymous to synonymous SNPs being reversed, indicating enrichment for functional variants. The overlap of variants between populations was modest, with 9.5%, 19.4%, and 11% of variants being private in these populations, respectively. Genetic differentiation measured by Fst was high among populations (mean pairwise Fst= 0.036), with differentiation being greatest between Ethiopian and other populations (Fst=0.05). Conclusions: We identified a large number of novel and potentially functional rare variants among the 3 populations sequenced, with a high proportion of private variants in each population. These findings argue for more detailed characterisation of genetic variation in SSA populations using WGS approaches.

575W

The completion of the 1000 Genomes Project. *A. Auton, The 1000 Genomes Project Consortium.* Albert Einstein College of Medicine, Bronx, NY.

The 1000 Genomes Project has expanded the scope of known human genomic variation through the detection or validation of over 40 million SNPs, indels, and larger structural variants. As such, it is expected that over 95% of variants with an allele frequency of at least 1% within the human population will have been characterized by the project. In the final phase of the project, data have been collected from an additional 12 populations, with a particular expansion in sampling from populations with African or South Asian ancestry. In total, low coverage (~5X) whole genome sequence data have been combined with high coverage (~80X) exome data for over 2,500 individuals from 26 global populations.

Through these efforts, nearly all the short sequence variants detected in a typical human genome sequenced today will have been previously identified. To demonstrate the utility of the 1000 Genomes resource, we have obtained deep whole genome sequence data from 138 diverse trios, allowing direct comparison of the high and low depth sequencing approaches, as well as investigation of integrated calling methods. We show that even for high coverage genomes, utilizing 1000 Genomes data can improve the accuracy of detected variation and aid downstream interpretation.

In addition, the project has also been working towards ascertaining additional types of variation. While earlier phases of the project focused on simple types of genetic variation in easily accessible parts of the genome, the final phase is using novel analytic methods to identify more complex types of variation such as multi-allelic polymorphisms, complex structural variants, and short tandem repeats (STRs). By integrating multiple detection approaches that leverage information from read mapping, locally reassembly, and full-scale de novo assembly, we are able to generate a more complete picture of human genetic variation.

576T

Quality control metrics for whole-exome sequence data analysis and detailed genetic variation from 6,000 disease participants within the UK10K project. *J. Floyd on behalf of the UK10K Consortium: exomes.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The UK10K project is a large-scale next generation sequencing project that has sequenced nearly 4,000 low-coverage whole genomes (6x depth) and over 5,000 high-coverage whole exomes (50x depth). Very few projects of this magnitude have been undertaken to date. Therefore, detailed characterization of the developed quality control (QC) and variation observed with comparison to other large resources such as the 1000 Genomes Project and the NHLBI EVS is of benefit to the broader community. Genome-wide association studies have previously demonstrated the importance of high quality data prior to analysis and these issues are even more likely to influence analysis of next-generation sequence data. The sample-level QC undertaken included identifying sample contamination (~3% of samples sequenced), cryptic population relatedness (0–3% depending on subgroup), genuine familial (but sometimes inaccurately reported) relatedness, and ethnic outliers (0–4% depending on subgroup). To detect possible effects of changing sequencing chemistry and/or variant calling we also report metrics such as the Ti/Tv ratio for SNVs and 3n/non-3n ratio for indels, mean depth and mean number of variants by minor allele frequency bins. After sample QC, 5233 samples remained for analysis. Ethnic outliers and related samples were removed or controlled for within the analyses. For instance, non-UK families were intentionally gathered and analyzed for some of the rare diseases. On average, nearly 70,000 variants passed variant QC per sample of which ~98% were SNVs. Variant QC included VQSR and a per-subject variant filter based on genotype quality. After sample and variant QC, we report a list of highly mutable genes where we see a higher than expected number of novel variants after controlling for gene length across several case phenotypes and sub groups within UK10K and NHLBI. We also evaluated the differences in multi- and single-sample variant calling and their respective abilities to accurately call singleton variants. The majority (~90%) of SNVs called and passed by single-sample are also called and passed by multi-sample with high genotype concordance (>99.5%). We are currently pursuing a validation of a subset of these variants using Sequenom plexes to estimate the proportion of variants missed by single or multi-sample calling alone. The QC steps and evaluations described here provide important information to the exome-sequencing community as ever more sequencing efforts proceed.

577F

Detailed annotated whole genome sequences of a Tibetan Trio family revealed high-altitude genetics and demographic insights. *M. He[1], A. Asan[1], X. Jin[1], E. Huerta-Sanchez[2], R. Wang[3], Z. Cuo[4], Y. Shan[1], Y. Shi[1], M. Yang[2], X. Xie[1], K. Harris[2], D. Cao[1], I. Song[1], J. Zhao[1], Z. Su[1], J. Zhang[1], Y. Chang[1], C. Yu[1], H. Huasang[1], J. Luosang[1], X. Yi[1], Y. Liang[1], R. Nielsen[2], Jun. Wang[1], W. Wang[1], Jian. Wang[1].* 1) BGI-ShenZhen, Shenzhen, Guangdong, China; 2) Statistics and Integrative Biology, UC Berkeley, Berkeley, CA; 3) Naqu local Peoplès Hospital, Lhasa, China; 4) The Second People's Hospital of the Tibet Autonomous Region, Naqu, China.

Unlike most other population groups in the world, the Tibetan population is unique in that it thrives at altitudes higher than 4000 meters, having experienced extreme cold and extreme altitude conditions for many generations. Understanding how Tibetans are adapted on a genetic level to live in low oxygen environments may elucidate the molecular pathways underlying the negative health effects of oxygen deprivation. Thus, the whole-genome sequences of the Tibetan trio will be a useful resource because these genomes will (1) contribute to the diversity of whole-genomes that have been sequenced, (2) be a good reference for studies on identifying the genetic basis of high altitude adaptation, (3) aid studies of gene-environment interactions for those disease variants and (4) be valuable for understanding human genetic history and demography. Here, we sequenced and analyzed the first whole genomes of a Tibetan trio family (father-mother-son). We identify 3,947,651 SNVs, 730,709 InDel and 3,480 structure variants from the two genomes(dad and mom), including 278,689 novel variants (7.06%) might be specific to the Tibetan population, and measured their relevance for individual health, high altitude adaptation and demographic history. We also inferred the effective population sizes of the Tibetan population from the two diploid genomes(dad and mom) of the trio and indicated Tibetans split from Han around 22,000 years ago. This is inconsistent with previous exome based estimation, partly caused by few recombination events were left in present sequence more recently than 20,000 years ago. We found a very rare variants (low frequency in other population) from moms disease prediction may be associated with her high-altitude heart disease. The discovery of a large proportion of population specific variants and the other insights we obtained in the study demonstrate the necessity of population-scale whole genome sequencing for genetic health profiling and more accurate demographic inference of the Tibetan population in the future.

578W

Genetic Structure and Diversity of 38 Singapore Indians from Deep whole-genome sequencing. *L.P. Wong, H.T. Ong, K.H. Lai, W.Y. Saw, X. Liu, E.P.N. Pillai, Y.Y. Teo.* Saw Swee Hock School of Public Health, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260.

Deep whole genome sequencing across multiple samples provides an opportunity to elucidate the genetic structure and diversity of a population at finer resolution. We conducted deep sequencing of 38 cosmopolitan Indians in Singapore to understand the genetic structure of Asian Indians which is unique in relation to its historical background and genetic admixture. Genetic diversity of Asian Indians is a powerful avenue to establish disease loci map. Furthermore, this population which contributes close to one fifth of the world's population is absent in the Phase 1 release of the 1000 Genomes project. Therefore, our aim of sequencing 38 Asian Indians is to analyze their genetic structure and fully characterize their genetic profiles. On top of that, the formal study of their genetic diversity would reveal many implications in medical research and serve as an invaluable resource. Given the availability of whole genome sequences of archaic hominin genomes (Denisovan and Neanderthal) and 15 other populations across the globe, we examined the relationship of modern human genomes with archaic hominin genomes. The genetic makeup of our samples was more closely related to Southern Indians than Northern Indians. We then demonstrated that 38 Indian genomes has high genetic sharing with Denisovan, evidence of potential admixture between Denisovan and modern humans who are ancestors to Asian Indians. Moreover, we discovered genetic ancestry of mitochondrion and Y chromosome that deviates from the autosomes.

579T

Detection of variations and their frequencies of the *CCR5* gene and its promoter region in Japanese and Okinawan population by NGS analysis using pooled DNAs. *T. Kaname[1], K. Yanagi[1], M. Higa[1], S. Song[2], K. Naritomi[1].* 1) Dept Med Gen, Univ Ryukyus, Nishihara, Japan; 2) Roche Diagnostics K.K., Tokyo, Japan.

Discoveries of the GWAS based on common SNPs could explain 2-15% of heritable variation in disease risk. It was speculated that there are considerable 'missing heritability' existing. Since rare variations would be next resources to elucidate such 'missing heritability', it is important to find novel rare variations in a targeted region in a population. We developed a method for survey analysis of low frequent variations in a targeted region in a specific population by next-generation sequencing (NGS) analysis using pooled DNAs.

Here, we show an example to trace variations and to estimate each frequency in the *CCR5* gene including promoter region in 100 Japanese and 100 Okinawan people. Genomic DNAs of 100 individuals were mixed in equal amount or genomic DNA in each individual was used to amplify CCR5 genomic region by LA-PCR. Then, pooled targeted region of CCR5 for 100 individuals was analysed using a NGS platform, GS Junior (Roche). After mapping reads to the reference, SNPs and indels were called and the frequency was estimated by count rate of the reads. Next, we confirmed each variation and calculated each frequency in the population by PCR-RFLP, allele specific PCR or direct sequencing in each individual.

Allelic frequency estimated by the NGS analysis using pooled DNAs almost correlated the real frequency calculated by the individual analysis. Finally, eight novel variations including non-synonymous substitution were found in 200 individuals. For sensitivity to low allelic frequency, 0.5% of allele in the population could be detected at least.

580F

Functional characterization of rare variants in human Dopamine receptor D4 gene by genotype - phenotype correlations. *A. MichealRaj[1], N. Jatana[2], Md. Jafurulla[3], L. Narayanan[2], A. Chattopadhyay[3], B.K. Thelma[1].* 1) Department of Genetics, University of Delhi South Campus, New Delhi, Delhi, India; 2) Bioinformatics Center, Sri Venkateswara College, New Delhi, India; 3) Membrane and Receptor Biology Group, Centre for Cellular & Molecular Biology, Hyderabad, India.

Objective: Rapid population growth and weak purifying selection in human populations have been suggested to be causal for the origin of rare variants, many of which could be deleterious and have a major effect of relevance to disease risk. Several rare variants have been reported in dopamine receptor D4 (DRD4) gene, belonging to the GPCR family and associated with many neuropsychiatric conditions. DRD4 is also an important drug target in antipsychotic medication. Therefore, the aim of this study was to identify rare variants of therapeutic relevance in *DRD4* using a combination of in silico and in vitro tools. **Methods:** Cells stably expressing the cDNAs of four non-synonymous coding rare variants of *DRD4* (rs1800443, V194G; rs4991150, R237L; rs3889692, A281P; rs34662058, S284G) were generated independently. Ligand binding using [3H]spiperone, G-protein mediated signal transduction (Inhibition of forskolin stimulated adenylate cyclase acivity and phosphorylation of extra cellular receptor kinase) and molecular dynamics-simulation studies were performed and analysed to identify functionally important variants. **Results:** Of the four variants, A281P and S284G ensued to be functionally similar to wild type (WT). V194G variant protein was not able to inhibit forskolin stimulated adenylate cyclase activity and also failed to phosphorylate the extra cellular receptor kinase due to its insensitive nature towards both dopamine and quinpirole. Further, ligand binding studies showed significant reduction in binding affinity (Kd = 2.16 nM; p<0.001) and total number of binding sites (~66%; p<0.001) compared to WT. Potency of dopamine and quinpirole reduced to ~ 6-fold (p<0.01) and 3 fold (p<0.01) respectively with R237L when compared to WT protein and ligand binding studies showed reduction only in total number of binding sites (~40%; p<0.01). Ligand docking studies with these two variants revealed that binding of both dopamine (agonist) and spiperone (antagonist) with V194G variant seems to be superficial due to the structural rearrangement while variant R237L was structurally similar to WT. **Conclusion:** Valine (at V194G) and arginine (at R237L) residues seem to be important for the activity of DRD4.V194G variant seems to be structurally altered and thus rendered non-functional while R237L seems to be functionally active but with altered expression level. This novel finding is of potential value for lead molecule development for dopamine dysregulated disorders.

581W

Determining the properties of loss-of-function variation in candidate essential human genes. *A.R. Rao, S.F. Nelson, E. Vilain, W.W. Grody, H. Lee.* David Geffen School of Medicine, University of California, Los Angeles, CA.

Rare genetic variants that severely disrupt protein-coding genes are known as loss-of-function (LoF) variants, and they are of considerable scientific and clinical interest, due to their presumed high probability of being deleterious and having a causal role in severe Mendelian disorders. A previous study surveying exomic variation in imputed low-coverage data and seven high-coverage exomes has estimated that a typical human genome contains approximately 100 genuine LoF variants with ~20 genes completely inactivated. We replicated this study using clinically generated, high-coverage (~100X mean coverage) exome data from over 200 individuals. In addition, we surveyed the frequency of individuals that are heterozygous or homozygous for LoF variants (nonsense mutations, splice site disrupting SNVs, frameshift indels, or larger deletions) in each gene, with further analysis on 1088 candidate 'essential' genes: orthologs of mouse (Mus musculus) genes with lethal knockout phenotypes. Only high-quality variants with a frequency of less than 1% based on publicly available exome data were considered. An individual genome contains ~280 candidate high quality LoF variants after initial filtering, and 1.4% of genes harbor LoF variants, on average. The proportion of genes mutated is not significantly different for essential versus non-essential genes, but the properties of LoF variants in the two categories differ. Nonsense mutations in essential genes are biased in their location and tend to be close to the start or the end of protein transcripts relative to nonsense mutations observed in non-essential gene list. Additionally, ~15% of the frameshift mutations found in non-essential genes affect all known transcripts of a gene. In contrast, this value is ~0 in essential genes. Thus, inspection of the genomic context of LoF variants is an important aspect of variant pathogenicity filtering of rare SNVs. In conclusion, the distribution of LoF variants in genes may suggest which regions of the studied genes are non-essential, or whether protein products may be rescued by alternative splicing, alternate pathways or redundant genes. Importantly, a measure of the expected LoF variation in a gene enables the estimation of the probability that a gene is causal for a severe Mendelian disorder, if a novel mutation were to be observed in that gene, and incorporating this information during variant interpretation would have a great impact on clinical sequencing practice.

582T

Functional assessment of genetic variants associated with Lp(a) levels on chromosome 6q25-26. *K. Chen[1], W. Lu[1], Y. Cheng[1,2], K. Ma[1], G.S. Gerhard[3], C.D. Still[3], X. Chu[3], R. Yang[1], A. Gorden[1], M.J. Quon[1], B.D. Mitchell[1], A.R. Shuldiner[1,2], M. Fu[1].* 1) University of Maryland School of Medicine, Baltimore, MD; 2) Veterans Affairs Maryland Health Care System, Baltimore, MD; 3) Geisinger Obesity Institute, Geisinger Clinic, Danville. PA.

An increased level of lipoprotein(a) [Lp(a)] is a risk factor for atherosclerosis and aortic valve calcification. Through a genome-wide association study for Lp(a) in 1200 Old Order Amish subjects, we identified 137 common variants ($P = 5 \times 10^{-8}$ to 3.91×10^{-19}) spanning ~5.3 Mb region on chromosome 6q25-26 that were within or flanking 26 genes including *LPA*. To further investigate the effect of *LPA* variants on Lp(a) levels, we sequenced 40 exons, all intron-exon boundaries and 2 kb of the promoter region of the *LPA* gene in 24 Amish subjects. We identified 23 variants including 6 missense variants in exons 26, 32, 37, 39 and 40. Among these, two rare variants (rs3798220, MAF=0.009; rs10455872, MAF=0.022) showed the most significant associations with Lp(a) levels (P value: 1.07×10^{-14}; 1.85×10^{-12} respectively). We also measured copy numbers of kringle IV-2 (KIV-2), a known variant associated with Lp(a) levels, using qPCR. The number KIV-2 was significantly associated with Lp(a) levels ($P = 2.28 \times 10^{-9}$). Conditional analyses revealed that rs3798220 and rs10455872 were associated with Lp(a) levels independent of one another, and KIV-2 numbers were partially dependent on the two SNPs. Finally, we performed conditional analysis of SNPs on chromosome 6q25-26 region adjusted for the three variants in LPA gene. The magnitude of the association in the region of 6q25-q26 with Lp(a) levels was largely reduced (the most significant $P = 3.01 \times 10^{-6}$). To investigate the function of rs3798220 and rs10455872, we examined LPA gene expression by Quantitative PCR (qPCR) in 70 liver samples according to genotypes, matched by age and gender. LPA mRNA levels were significantly higher in carriers than noncarriers of rs10455872 (P=0.0002). We didn't detect any difference in LPA mRNA levels between rs3798220 carriers and noncarriers (p>0.05). These findings suggest rs10455872 within *LPA* may play a key role influence mRNA levels of *LPA* (transcription of stability), while rs3798220 may influence Lp(a) levels through effects on translation or protein stability. Further studies are underway to precisely elucidate the complex genetic architecture and regulation of Lp(a) and the role of genetic variation as a risk factor for cardiovascular diseases.

583F

Identification of nine mutations in the COL1A1 gene in Czech patients with osteogenesis imperfecta. *L. Šormová[1], I. Fialkowski[2], W. Van Hul[2], G. Mortier[2], I. Maík[3], I. Mazura[1].* 1) Charles University in Prague, 1st Medical Faculty, Prague, Czech Republic; 2) Antwerp University and University Hospital, Centre for Medical Genetics, Antwerp, Belgium; 3) Ambulant Centre for Defects of Locomotor Apparatus, Prague, Czech Republic.

Type I collagen is an abundant protein in connective tissues, especially in bone. The type I collagen molecule is a heterotrimer composed of two copies of the alpha 1(I) procollagen chain (encoded by the COL1A1 gene) and one copy of the alpha 2(I) procollagen chain (encoded by the COL1A2 gene). With this study we aimed to identify the genetic defect in a series of Czech patients with osteogenesis imperfecta (OI) (types IA, II/III, III, IVA and IVB). We elected to restrict the study to the analysis of the COL1A1 gene since 60% of OI mutations are present in this gene. We used gDNA isolated from whole peripheral blood. This research is the first study of the whole coding sequence of the COL1A1 gene of Czech patients affected by OI. The obtained genomic data were analysed using the Ensembl database. Unreported DNA variations in the coding and intronic consensus sequences were checked using the PC program Alamut (algorithms SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer and Human Splicing Finder). We were able to identify the causal mutation in eight patients diagnosed with OI. In three cases with OI IA, we identified nonsense mutations p.Tyr47X at exon 2 (light blue sclera, mild bone deformities, joint hyperlaxity), p.Arg131X at exon 5 (shortened upper body segment, slightly barrel chest) and p.Glut1341X at exon 50 (blue sclera, light bone deformities, barrel chest, higher fractures frequency). In two patients we observed heterozygosity for a missense mutation p.Cys61Phe at exon 2 (OI type III - bone deformities, osteoporosis, barrel chest) and p.Pro1186Ala at exon 48 (OI type IA - blue sclera, mild bone deformities, lower BMD). In the latter patient also a heterozygous nonsense mutation p.Arg415X was found at exon 19. Changes of noncoding sequences were identified in three cases: c.1057-1G>T (intron 16, OI type IA/IVA - blue sclera, hearing loss, bone deformities, osteoporosis), c.1300-1G>A (intron 19, unclassified OI type, clinical data currently not completed), c.1353+38C>G (intron 20, OI type IA - blue sclera, face asymmetry, joint hyperlaxity, bone deformities, osteoporosis).

584W

Findings from the third Critical Assessment of Genome Interpretation, CAGI 2013, a community experiment to evaluate phenotype prediction. *S.E. Brenner[1], D. Barsky[1], J. Moult[2], CAGI Participants.* 1) University of California, Berkeley, CA; 2) IBBR, University of Maryland, Rockville, MD.

The Critical Assessment of Genome Interpretation (CAGI, 'k-jē) is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. In the experiment, participants are provided genetic variants and make predictions of resulting phenotype. These predictions are evaluated against experimental characterizations by independent assessors. A long-term goal for CAGI is to improve the accuracy of phenotype and disease predictions in clinical settings.

The third CAGI experiment (2012 - 2013) consisted of ten diverse challenges. At time of abstract submission, assessment was underway and results will be known in July 2013. CAGI deliberately extends challenges from previous years, with the continuity allowing measurement of progress. For example, in the second CAGI, in a challenge to predict Crohn's disease from exomes, one group was able to identify 80% of affected individuals before the first false positive healthy person. In the third CAGI experiment, this challenge used an improved dataset, and several groups performed remarkably well, with one group achieving a ROC AUC of 0.94 in initial assessment. Another expanded challenge involved using Personal Genome Project genome data to predict phenotypes and match health records; this year, several groups were able to successfully map some genomes.

In the expanded challenge to predict benign versus deleterious variants in DNA double-strand break repair MRN genes~Rad50 (from last year), Mre11, and Nbs1~as determined by those that appear in a breast cancer case versus healthy control, predictions show how methods differ sharply in their effectiveness even amongst proteins in the same complex. A new challenge this year was to use exomes from families with lipid metabolism disorders, Familial Combined Hyperlipidemia (FCH) and Hypoalphalipoproteinemia (HA), to predict lipid profiles and a causative variant. Assessment of this challenge revealed a twist wherein real-world data differed sharply from theoretical models.

Other CAGI challenges include predicting: cancer impact of BRCA variants; splicing impact of p53 gene variants; and cell proliferation impact of p16 variants. Complete information about CAGI may be found at http://genomeinterpretation.org.

585T

Whole genome sequencing of rhesus macaques reveals substantial functional variation and justifies a `reverse genetics' approach to identify new models of human disease. *M. Raveendran[1], D. Rio Deiros[1], G.L. Fawcett[1,2], Z. Johnson[3], N.H. Kalin[4], R.W. Wiseman[5], B. Ferguson[6], E. Vallender[7], S. Kanthaswamy[8], H. Doddapaneni[1], S. Jhangiani[1], D.M. Muzny[1,2], R.A. Gibbs[1,2], J. Rogers[1,2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Yerkes National Primate Research Center, Emory University, Atlanta, GA; 4) HealthEmotions Research Institute and Dept. of Psychiatry, University of Wisconsin, Madison, WI; 5) Wisconsin National Primate Research Center, Madison, WI; 6) Oregon National Primate Research Center, Oregon Health & Science Univ., Beaverton, OR; 7) New England Primate Research Center, Harvard Medical School, Southborough, MA; 8) California National Primate Research Center, University of California-Davis, Davis, CA.

Rhesus macaques (Macaca mulatta) are the most widely used nonhuman primates in biomedical research, serving as models in multiple areas, including infectious disease, neurobiology, endocrinology and metabolic diseases among others. Traditionally, macaque models of human disease were developed through expensive surveys to identify animals with appropriate phenotypes. To explore an alternative strategy, we generated whole genome sequences (25x coverage) for 51 unrelated Indian-origin rhesus macaques from five NIH primate research centers. Identification of potentially functional genetic variation in this species will create novel opportunities to study targeted genetic models of specific disease mechanisms or genetic pathways in animals much more similar to humans than rodents or other mammalian model organisms. Sequence reads (100bp, paired-end) were mapped to the rhesus genome assembly (rhemac2) using BWA, and variation called using SAMtools mpileup. Variant calls were filtered with a minimum q-score of 20, and only variants observed in two or more animals were retained. This identified >27.3 million SNPs (9.5 SNPs per kb) and >4.2 million small indels (1.5 indels per kb), more variation than would be expected in a similar sized survey of humans. Variants were annotated using ENSEMBL to identify those with possible functional effects. We found 48,485 non-synonymous variants, 756 stop codons gained or lost, 2,901 frameshift indels in coding genes, 1,753 codon insertions or deletions, and other annotation classes. The list of genes affected includes many known or suspected to influence risk of human disease. Using the same read data but less stringent filtering (not requiring that variants be observed in 2 animals) yields a total variant count of >37 million. The allele frequency spectrum for this distribution is heavily weighted to low frequency variation, similar to humans. More than 95% of this total variant list has minor allele frequency <0.25 and more than 45% have MAF <0.05. We will report results concerning coding and noncoding variation in rhesus macaques and discuss the significance of this reverse genetics approach. These results show that a significant amount of potentially functional variation is segregating in macaque research colonies, and opens new opportunities for analysis of specific phenotypic effects of functional mutations in genes known or suspected to influence human disease.

586F

Expression Minigenes Reveal the Limitations of Algorithms Predicting the Consequences of Putative Splice Site Mutations. *N. Sharma[1], P. Sosnay[1,2], A. Franca[1], C. Douville[3], A. Ramalho[4], L.B. Gottschalk[1], K.R. Siklosi[1], M. Amaral[4], R. Karchin[3], G.R. Cutting[1,5].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD; 4) University of Lisboa, Faculty of Sciences, BioFIG - Centre for Biodiversity, Functional and Integrative Genomics, Lisboa, Portugal; 5) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Splice sites flanking each exon direct the generation of mature mRNA from heteronuclear RNA. Algorithms that utilize sequence conservation in splice sites provide reasonably accurate predictions as to whether DNA variants will affect the splicing process; however, the ability of algorithms to predict expression of full-length, truncated or no protein is untested. This is a critical issue as the nature and relative abundance of synthesized protein is generally the major determinant of the clinical phenotype caused by a mutation. To address this issue, we developed a novel expression minigene (EMG) system containing flanking intron sequence (~200bp) from multiple adjacent exons cloned into a full-length cDNA in a mammalian expression vector. Presence of the CMV promoter and EMG enables RNA splicing and translation to be tested concurrently. As a proof of concept, two EMGs including introns 11 and 12 and introns 14 to18 of the CFTR gene were created. Splicing was analyzed by RT-PCR followed by Sanger sequencing and protein translation by Western blot in two cell lines: Human Embryonic Kidney (HEK) 293 cells and CF Bronchial Epithelial (CFBE41o-) cells. A normal pattern of RNA splicing and robust protein translation was confirmed for the two wild-type EMGs. Five mutations associated with cystic fibrosis (c.1585-1G>A, c.1585-2A>G, c.1585-3T>G, c.1585-8G>A, c.1585-9T>A) were introduced into intron 11, three mutations (c.2657+3delG, c.2657+5G>A, c.2657+2_2657+3insA) into intron 16, one in intron 18 (c.2988+1G>A) and one in exon 18 (c.2988G>A). Loss of protein synthesis was observed for 9 of the 10 mutations due to nonsense mediated mRNA decay caused by a frameshift or an in-frame premature stop codon introduced by aberrant splicing. One mutation (c.2657+2_2657+3insA) generated fully processed CFTR protein. We compared our results with two currently available algorithms (ASSEDA and HSF) used to predict the transcript isoforms resulting from these mutations. Overall, HSF correctly predicted the aberrant transcript isoforms for half of the mutations while ASSEDA correctly predicted the abnormal splicing patterns for 2 of the 10 mutations. Neither of these programs predicted whether protein would be synthesized. Despite the limitations of current algorithms, experimental data generated by EMGs should enable the generation of algorithms that accurately predict the effect of splicing mutations upon protein production.

587W

Linear Decay of Retrotransposon Antisense Bias across Genes is Contingent upon Tissue Specificity. *S. Linker[1], D. Hedges[2].* 1) HIHG, University of Miami, Miami, FL; 2) Division of Human Genetics, The Ohio State University, Columbus, Ohio.

Retrotransposons (RT), which make up approximately half of the human genome, are gaining attention due to new techniques which allow polymorphic RT insertions to be identified from next-generation sequencing data. The ability to move forward and infer the effects of these polymorphisms will require a substantial increase in the understanding of the various effects that retrotransposons have on a coincident gene. Previous work in this area has highlighted the non-uniform distribution of retrotransposons across the genome leading researchers to propose that there may be local effects that modify their retention in various genomic contexts. Our work furthers this effort by determining predictable trends of retrotransposon accumulation as well as through identifying links between the presence of retrotransposons and gene expression. Our primary findings show that retrotransposons which are in the antisense orientation with respect to a gene, exhibit a linear decay in frequency across the length of the gene. Conversely, retrotransposons in the sense orientation maintain a low-level frequency independent of location within a gene. This linear trend of antisense RTs is inversely correlated with exon frequency, which we show has a linear increase in frequency across the length of most genes. Interestingly, this correlation is dependent upon the gene type (ie: housekeeping versus tissue-specific). Moreover, we have found that transcription factor binding motifs which are contained within RTs also exhibit a pattern of accumulation in certain regions of the gene. We used retrotransposons which are polymorphic in the human population to determine the potential for effect on expression of these inserts. Indeed we identified a subset of polymorphic RTs that were significantly correlated with altered gene expression after correction with FDR. These findings begin to map out predictable variables for the retention of RTs in the genome, and may aid researchers in inferring the effects of novel polymorphic RTs on local gene expression.

588T

Nicotine causes genome-wide microsatellite instability in normal epithelial cells. *J. Bavarva, L. McIver, T. Hongseok, H. Garner.* Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA.

Background: Nicotine has been associated with cancer development. However, there is no direct evidence of nicotine on genome-wide instability, including microsatellites, a highly unstable component of the genome. We hypothesized that nicotine can cause genome-wide microsatellite variability that can have significant biological consequences. Methods: We used a combined approach of customized microsatellite specific aCGH and low coverage whole genome sequencing to quantitate the microsatellites variability in MCF-10A normal human breast epithelial cells upon stressful exposure of nicotine. Results: Our customized microsatellite specific oligonucleotide array quantitates the content of microsatellites including all-possible repeats (1- to 7-mer) in a genome wide scale, which provides us an ability to quantitate microsatellites changes in mass. Global microsatellites content (GMC) exhibited hyper variability (17.8%) upon nicotine exposure. GC rich microsatellites showed a slight bias as they tend to loose GMC (54%) more than AT rich (46%) under nicotine stress. Effects were dose dependent and a stress recovery test indicated that GMC instability did not revert upon removal of stress. We next analyzed effect of oxidative stress on genome-wide microsatellites and discovered that pure oxidative stress is a very powerful microsatellite instability inducer (30.9%). The possible mechanism of nicotine induced microsatellite variability may therefore be linked to nicotine's ability to cause oxidative stress. Next, we used unbiased next-generation sequencing to perform low coverage whole genome sequencing of nicotine stressed cells and controls, and to re-validate our array based findings. We discovered that 10,502 microsatellite loci vary between nicotine stressed and control cells. We cross-examined all significant microsatellite motif families that were revealed through aCGH analysis and found general agreement between WGS and aCGH findings. We identified five variable microsatellite loci in coding regions (MOG, RPL14, ALLC, FAM157B and MAGEF1) and number of the loci upstream of coding sequences that may possibly influence the gene transcription and regulation. Conclusion: The results indicate that nicotine induces genome-wide microsatellite changes, may promote genomic instability and inhibit DNA-damage repair through oxidative stress that can facilitate cancer genesis.

589F

Probing genes for hyperphagia in rare-obesity related disorders. *M.G. Butler[1], J.D. Marshall[2], J. Rethmeyer[1], K. Wang[1], A.M. Manzardo[1].* 1) Psychiatry & Behavioral Sciences, University of Kansas Medical Center, Kansas City, KS; 2) The Jackson Laboratory, Bar Harbor, ME.

Hyperphagia and obesity are key features of several rare genetic obesity-related disorders including Prader-Willi syndrome (PWS) and Alström syndrome (ALMS). A better understanding of genetic causes through the study of obesity-related disorders should provide a more comprehensive picture of mechanisms that control food intake and energy balance related to the development of obesity. Comparison of gene expression patterns associated with rare mutations should provide insight into commonly disturbed gene pathways involved in appetite control and body weight regulation with application to obesity in the general population. We compared coding and non-coding gene expression in PWS, ALMS and nonsyndromic obesity relative to non-obese adult male controls using both lymphoblasts and brain tissue with the Affymetrix Human Exon 1.0 ST and GeneChip miRNA 2.0 arrays from total RNA isolated from PWS (N=7 males; mean age = 25yr; mean BMI=39), ALMS (N=6 males; mean age = 30yr; mean BMI=42), obese (N= 7 males; mean age = 32yr; mean BMI=51) and non-obese (N=7 males; mean age = 26yr; mean BMI=22) controls. In relationship to lymphoblasts from non-obese males, we found disturbances with 231 genes up-regulated in ALMS but no genes up-regulated in the obese or PWS males (FDR ≤ 0.2, fold ≥ 1.5). Of 124 down-regulated genes in ALMS, only the metallothionein gene (involved in heavy metal toxicity and oxidative stress) was significantly down-regulated in common with obese males (FDR ≤ 0.2, fold ≤ -1.5). Only SNRPN was down-regulated in PWS. Fifty-two miRNAs were up-regulated in lymphoblasts from ALMS and four were in common with obese but none with PWS (FDR ≤ 0.05, fold ≥ 2.0). Four separate target genes (EFR3B, CSF2RB, CDX5R1, TBC1D8) were significantly down-regulated in association with five up-regulated miRNAs (miR339-5p, miR-92a-1*, miR-93*, miR-25*, miR-92b-1*) in ALMS. Six separate target genes (CD68, MXI1, FAM102A, TP53INP1, MYO1D, ZRANB1) were significantly down-regulated in association with seven disturbed miRNAs (miR-93*, miR-373*, miR-29b-2*, miR-30c-1*, miR27a*, miR27b*, miR-149). Up-regulated genes in ALMS impacted development, cell cycle and transcription while down-regulated genes were involved with metabolic processes, immune responses and cell signaling transduction. Similar disruption of gene expression was found in brain tissue. The high number of gene disturbances observed may reflect the broad, multi-organ involvement of ALMS pathophysiology.

590W

High sensitivity, single-cell expression analysis with the QX100 Droplet Digital PCR system. *G. Karlin-Neumann, S. Wang, Y. Jouvenot, E. Heffner.* Digital Biology Center, Bio-Rad Laboratories, Pleasanton, CA.

Over the last decade, it has become increasingly evident that gene expression profiles can vary from cell to cell, even within an apparently homogenous population. The analysis of this heterogeneity has become a focus of interest in various fields of biology, especially in stem cell research. The main obstacles to analysis of gene expression at the single-cell level are the low amount of starting material and the low abundance of many transcripts of interest. This requires a high level of confidence in results obtained from unique samples, making it difficult to be done accurately by traditional quantification methods such as qPCR. Droplet digital PCR (ddPCR™) provides absolute quantification of individual molecules with high precision, and without the requirement for standard curves or pre-amplification steps. Using Bio-Rad's QX100 ddPCR system, we developed a method that measures single-cell gene expression in multiplexed assays with high sensitivity and reproducibility, thus enabling us to simultaneously analyze expression of different targets in cDNA from the same cell. In order to perform this, we carefully evaluated various cell lysis and cDNA synthesis methods and developed a protocol with flow-sorted Jurkat cells that is fully compatible with ddPCR, easy to use and capable of analyzing gene expression in single cells without pre-amplification. We show that high, medium and low abundance transcripts (< 20 copies/cell) can be reproducibly measured with ddPCR. We further used this protocol to measure cell cycle-specific genes and revealed distinctive gene expression patterns in populations of single cells. This method allows us to easily and quickly measure the expression of multiple genes of interest in single cells, minimizing the stochastic effect of sampling and empowering us to accurately and sensitively detect and quantify low-expressing genes in single cells.

591T

Individual variation in the rate of retrotransposition in iPS cells and its effect on genomic instability with regard to medical utility. *T.T. Doucet[1,2], C. Smith[1,2,4], A. Ewing[3], K. Burns[1,5], L. Cheng[1,4], H.H. Kazazian[1].* 1) Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD USA; 2) Pre-doctoral training program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; 3) Center for Biomolecular Science and Engineering, University of California at Santa Cruz, Santa Cruz, California 95064, USA; 4) Stem Cell Program, Institute for Cell Engineering, Division of Hematology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 5) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Genomic instability can be caused by repetitive sequences in the human genome (retrotransposons) that can copy and paste themselves into new sites. We seek to understand how retrotransposition plays a role in genomic instability during the reprogramming process as well as the differentiation process. For iPS cells to be utilized as a medical treatment, their genomes must be stable. We are evaluating iPS and parental cell lines, the cells from which the iPS cells were derived, from three individuals utilizing next generation sequencing, RT qPCR, and immunohistochemistry. To ensure we find all the putative somatic insertions in the iPS lines, we have used a LINE-1 enrichment technique, L1-seq, prior to next generation sequencing. To validate putative somatic insertions in the iPS cells, we will use site specific PCR, a LINE-1 specific primer and a genome specific primer designed for each individual insertion, and Sanger sequencing. The presence of LINE-1 mRNA and LINE-1 associated proteins such as ORF1p and ORF2p are positive indicators of retrotransposon activity. Using the RNA isolated from both the parental lines and the iPS lines we have demonstrated an ~17-fold increase in LINE-1 RNA in the iPS lines evaluated compared to the expression in the parental lines. We have also seen ORF1 protein expression in the iPS cells, which is not present in the parental cells. To evaluate the differences in the rate of retrotransposition between individuals' iPS lines, we will utilize the retrotransposition assay in cell culture in which a full length LINE-1 is transfected into the cells and the rate of retrotransposition is characterized based on the protein expression of GFP using FACS. In the future, we will differentiate the iPS cells into several types of cells such as neuronal precursor cells and gastrointestinal cells to determine whether or not the differentiation process induces retrotransposition as well. By evaluating the rate of retrotransposition in iPS cells and their differentiated counterparts from several individuals, we hope to characterize retrotransposition to determine whether or not iPS cells derived from patients are sufficiently stable to be utilized as a medical therapy.

592F

High-resolution mapping of break locations in human common fragile sites FRA3B, FRA7H, and FRAXB. *C.L. Kaminski, W. Fitzsimmons, K. Dziuba, A. Layer, A.M. Casper.* Department of Biology, Eastern Michigan University, Ypsilanti, MI.

Common fragile sites are chromosomal loci that tend to break when DNA replication is partially impeded. These sites were first identified in humans and orthologs were later identified in many other mammals; the yeast Saccharomyces cerevisiae also contains loci susceptible to break formation under replication stress. Fragile site breaks can result in chromosomal alterations that drive early tumorogeneisis. For example, deletions at FRA3B and FRA16D cause loss of function at the tumor suppressor genes FHIT and WWOX, respectively, and FRA7G instability drives amplification of the MET oncogene. Several hypotheses have been proposed to explain the mechanism leading to breaks at fragile sites. One hypothesis proposes that a lack of origins in fragile site regions leads to incomplete replication when replication is slowed, causing breaks in un-replicated regions during anaphase. A second hypothesis proposes that fragile sites contain AT-rich sequences that are considered to be flexible and prone to forming secondary structures during replication, which can cause further replication stalling and replication fork collapse. We are evaluating these hypotheses using yeast artificial chromosomes (YACs) containing inserts of human DNA from fragile sites FRA3B, FRA7H, and FRAXB. The YAC insert sizes are 1.3 Mb, 730 Kb, and 362 Kb, respectively. We have placed yeast under replication stress, caused by low levels of polymerase alpha, to induce fragile site breaks. We have mapped 30 break locations in each YAC to a 6 Kb resolution. In all three YACs, there is a hotspot for breaks near the distal end of the human DNA fragile site insert. We found that 60% of FRA3B breaks are clustered within a 54 Kb hotspot, 60% of FRA7H breaks are within a 45 Kb hotspot, and 60% of FRAXB breaks are within a 55 Kb hotspot. We analyzed break locations in FRA3B and FRA7H relative to the locations of flexible AT-rich sequences and potential origins. There is no significant association between fragile site break locations and the AT-rich sequence locations. However, the average distance from break locations to potential origins is significantly larger than expected from the distribution of control, non-break areas ($p<0.01$). Analyses of break locations in FRAXB are in progress. These data support the origin-related hypothesis for fragile site breaks. To further test this hypothesis, we are generating a YAC with an additional replication origin inserted near the break hotspot in FRA3B.

593W

Functional characterization of DcR3 in EBV transformed cell lines from IBD patients of different allelic background and role in disease pathogenesis. *R. Pandey[1], C. Cardinale[1], K. Kachelries[2], S.F.A. Grant[1,3], R. Baldassano[2,3], H. Hakonarson[1,3].* 1) Centre for Applied Genomics, Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia; 3) Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania.

Aim: Decoy receptor3 (DcR3), member of the TNFR superfamily is a soluble receptor for FasL and plays significant role in immune suppression and tumor progression by neutralizing the FasL mediated apoptosis signal. Here we investigate the possible immuno-modulation mediated by DcR3 and NF-κB in EBV transformed control and patient derived cell lines with and without risk variants in the TNF Receptor Superfamily 6B gene, TNFRSF6B captured by the tagging SNP, rs2315008. **Methods:** Expression of DcR3 and its kinetics were examined by immunoblot analysis in whole cell lysates from EBV transformed control and patient derived cell lines of different genotype background for rs2315008 (AA, AT, TT). DcR3 induced rapid activation of nuclear factor κB (NF-κB) monitored by immnunoblot analysis of IkBα. Expression pattern of NF-κB complexes and kinetics in non-secretors, control and patient-derived EBV transformed cell lines was examined by immunoblot analysis in whole cell lysates, cytoplasmic and nuclear extracts. DcR3 knockdown was performed using specific DcR3 siRNA. Cell proliferation and cell death was measured by MMT assay. Caspase8, caspase9, caspase3, and Bcl2 expression were analyzed by western blots. **Results:** EBV transformed cell lines derived from IBD patients harboring risk variants in TNFRSF6B (A allele) exhibit differential pattern of DcR3 expression and NF-κB activation kinetics in comparison with wild type. siRNA mediated knockdown post 24hrs of nucleofection results in decreased DcR3 expression, increased cell death and decreased cell proliferation, effects that were also genotype-dependent. **Conclusion:** EBV cell lines from IBD patients harboring risk variants in the TNFRSF6B gene exhibit differential pattern of DcR3 expression and NF-κB activation. We propose that pathogenic inflammation in CD may be in part be the result of non-canonical developmental signals impinging on a NF-κB signaling module with an altered homeostasis of IκB proteins. Therefore, therapeutic intervention targeting the NF-κB activation pathways represents a promising opportunity for future therapy of IBD.

594T

Functional characterization of the role that components of the retrograde transport machinery play in early HIV events. *S. Liu[1, 2, 4], M. Dominska[1, 2, 4], D.M. Dykxhoorn[1, 2, 3, 4].* 1) John P. Hussman Institute for Human Genomics; 2) Dr. John T. Macdonald Foundation Department of Human Genetics; 3) Department of Microbiology and Immunology; 4) University of Miami MIller School of Medicine, Miami, FL.

Human Immunodeficiency Virus (HIV) is a highly lethal lentivirus which is responsible for a tremendous amount of suffering and death in the United States and throughout the world. By attacking the host's immune system, HIV leaves the infected individual susceptible to a wide variety of opportunistic infections. HIV, like all viruses, relies on host cell factors for successful infection, replication and release of progeny virus. Large-scale functional genomic screens have identified a wide variety of novel host factors whose silencing inhibited viral replication in cultured cells. A more detailed examination of these HIV-dependency factors (HDFs) showed a significant enrichment for factors involved in the trans-Golgi network (TGN), a pathway which had not been previously implicated in early events in HIV replication. To examine the role that the TGN-HDFs play in the HIV-1 life cycle, cells silenced for the appropriate HDF were subjected to biochemical and cell biological analysis. The Conserved Oligomeric Golgi (COG) complex is a heteromer complex that functions as a tethering factor in concert with the Rab-family GTPase (RAB1) and the t-SNARE syntaxin 5 to facilitate the recruitment, interaction and fusion of membranes from early or late endosomes to the trans-golgi network (TGN). Our results show that the targeted silencing of components of the COG complex each impaired HIV-1 replication, as measured by intracellular p24 staining. These results suggested that the impairment in HIV-1 replication in these silenced cells occurred prior to gag translation. Analysis of the different forms of the HIV-1 genome showed that the defect in HIV-1 replication occurred prior to late RT product formation. This inhibition of HIV-1 is dependent on the HIV-1 envelope glycoprotein since viral particles pseudotyped with the vesicular stomatitis virus glycoprotein (VSVg) were unaffected. The identification of a novel role for HDFs in early events in the HIV-1 life cycle could suggest potential novel therapeutic approaches for the inhibition of viral transmission and replication.

595F

ïDifferential Gene Expression in Granulosa Cells from Polycystic Ovary Syndrome Patients with and without Insulin Resistance: Identification of Susceptibility Gene Sets through Network Analysis. *S. KAUR[1], KJ. ARCHER[2], MG. DEVI[3], A. KRIPLANI[4], JF. STRAUSS[5], R. SINGH[1].* 1) ZOOLOGY, UNIVERSITY OF DELHI, DELHI, India; 2) Department of Biostatistics, Virginia Commonwealth University, Richmond, Virginia; 3) Gouri Hospitals, Delhi 110007, India; 4) Department of Obstetrics and Gynecology, All India Institute of Medical Sciences, New Delhi 110029, India; 5) Department of Obstetrics and Gynecology, Virginia Commonwealth University, Richmond, Virginia.

Context: Polycystic ovary syndrome (PCOS) is a heterogeneous, genetically complex, endocrine disorder of uncertain etiology in women. Objective: Our aim was to compare the gene expression profiles in stimulated granulosa cells of PCOS women with and without insulin resistance vs. matched controls. Research Design and Methods: This study included 12 normal ovulatory women (controls), 12 women with PCOS without evidence for insulin resistance (PCOS non-IR), and 16 women with insulin resistance (PCOS-IR) undergoing in vitro fertilization. Granulosa cell gene expression profiling was accomplished using Affymetrix Human Genome-U133 arrays. Differentially expressed genes were classified according to gene ontology using ingenuity pathway analysis tools. Microarray results for selected genes were confirmed by real-time quantitative PCR. Results: A total of 211 genes were differentially expressed in PCOS non-IR and PCOS-IR granulosa cells (fold change1.5; P0.001) vs. matched controls. Diabetes mellitus and inflammation genes were significantly increased in PCOS-IR patients. Real-time quantitative PCR confirmed higher expression of NCF2 (2.13-fold), TCF7L2 (1.92-fold), and SERPINA1 (5.35-fold). Increased expression of inflammation genes ITGAX (3.68-fold) and TAB2 (1.86-fold) was confirmed in PCOS non-IR. Different cardiometabolic disease genes were differentially expressed in the two groups. Decreased expression of CAV1 (3.58-fold) in PCOS non-IR and SPARC (1.88-fold) in PCOS-IR was confirmed. Differential expression of genes involved in TGF- signaling (IGF2R, increased; and HAS2, decreased), and oxidative stress (TXNIP, increased) was confirmed in both groups. Conclusions: Microarray analysis demonstrated differential expression of genes linked to diabetes mellitus, inflammation, cardiovascular diseases, and infertility in the granulosa cells of PCOS women with and without insulin resistance. Because these dysregulated genes are also involved in oxidative stress, lipid metabolism, and insulin signaling, we hypothesize that these genes may be involved in follicular growth arrest and metabolic disorders associated with the different phenotypes of PCOS.

596W

Insulin-Induced Kinase Signaling Contributes to Individual Differences in Response to Insulin. *I.X. Wang[2], V.G. Cheung[1,2].* 1) Howard Hughes Medical Institute; 2) Genetics/Pediatrics, Univ Pennsylvania, Philadelphia, PA.

Individual differences in sensitivity to insulin underlie diseases such as diabetes and influence cell growth and cancer susceptibility. Insulin mediates cellular functions by triggering signal transduction pathways that lead to subsequent changes in gene expression. In this study, we focused on individual variation in signal transduction. Even though signaling is the first step in cellular response to nutritional demands, individual variability in insulin-induced activation of signaling pathways is poorly understood. To address this, we exposed primary skin cells from normal individuals to insulin and measured phosphorylation of kinases. We found extensive individual variation in insulin-induced activation of key signaling factors, including ERK whose induction differs by more than 20-fold among our subjects. By genetic analysis, we identified DNA variants that influence signaling response and downstream changes in gene expression and cell proliferation. The effect of variation in kinase activation on gene expression is substantial; for example, differences in ERK activation contribute to about 16 to 44% of individual variation in insulin-induced changes in gene expression. To better understand the relationship between the signaling proteins and their target genes, we took advantage of the variability and constructed networks which identify the connections within the signaling pathways and the extensive crosstalk between signaling regulators and their target genes. In this presentation, I will describe our results that demonstrate how signal transduction is an important contributor to insulin sensitivity, therefore offer kinase modulators as promising therapeutics for diseases characterized by insulin resistance.

597T

Relationship between mitochondrial DNA mutations and aging. Estimation of age at death. *S. C. Zapico, D.H. Ubelaker.* Anthropology, Smithsonian NMNH MRC 112, Washington, DC.

According to the aging theory presented by Harman, the production of free radicals rises with age and plays a key role in the degenerative processes of senescence. This oxidative stress increase could be the origin of cellular molecule damage. Particularly, the highest levels of oxidative stress are generated in the mitochondria due to the Electron Transport Chain (ETC). This oxidative stress increase induces accumulation of nonrepaired lesions in mitochondrial DNA (mtDNA). There are some studies that point out the relationship between mtDNA mutations and age in different tissues. These studies are potentially interesting for forensic identification because they can help to improve the estimation of age at death. Since teeth are the hardest tissue of human body, and one of the most abundant types of biological remains available in forensic cases, the present study aims to evaluate the mutations in mtDNA from dentin and pulp and their relation with the age. 30 healthy erupted third molars were extracted for valid clinical reasons from individuals from the Northwestern region of Spain aged 20 to 70. DNA was isolated from the dentin and pulp of each molar. We used specific primers of Hypervariable region 2 (HV2) of the mitochondrial D-loop to assess the mutations in each type of tissue by PCR. We found a decrease in the amplification of this region with age in dentin and variations in pulp. In order to quantify the amplification, we used Q-PCR and found the same results. There was an age-dependent decrease in the gene HV2 copy number in dentin. Using regression analysis, a negative significant linear correlation was found between mtDNA amplification and age in dentin. In contrast, we have not found a correlation between mtDNA amplification and age in pulp. The reason of this variation is the projection of the odontoblastic processes from pulp cells to dentin, which have numerous mitochondrias. As a result, the majority of oxidative stress is generated in the dentin, inducing mutations in mtDNA that are possible to relate to age. The findings from this research provide a new quantitative tool for the estimation of age at death that, in combination with traditional age markers, could improve identification accuracy in forensic cases. Future research may be able to expand on these results, using different types of teeth, analyzing different populations and extending the age range.

598F

Genome-wide analysis identifies heat shock induced gene and chromatin regulatory protein network in Saccharomyces cerevisiae. *R. Li[1], M.D. Ritchie[2].* 1) Bioinformatics and Genomics, The Pennsylvania State University, State College, PA; 2) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, State College, PA.

In Saccharomyces cerevisiae, hundreds of different proteins are involved in transcription regulation, yet much of the regulation mechanisms remain largely unknown. Previous studies have gained insights into transcription regulation from studying the protein factor binding responses induced by heat shock on cells; however, they were limited in scale. Recently, a comprehensive study of almost all DNA binding proteins, chromatin regulators, general transcription proteins, and elongation regulators in Saccharomyces generated ~200 proteins' genome-wide binding levels using ChIP-chip assays in normal and heat shock conditions. Basic statistical approaches have been used to identify important proteins that exhibit the highest fold changes in DNA binding levels between the two conditions. However, this approach can only interrogate individual proteins and it fails to uncover any potential interactions among the proteins involved in gene regulation. Thus we applied ATHENA (the Analysis Tool for Heritable and Environmental Network Associations), which is a software package for modeling the underlying interaction relationships between protein factors using a machine-learning technique, grammatical evolutionary neural networks (GENN). This proof-of-concept in yeast is the first evaluation of ATHENA for epigenetic data. Our model identified important protein factors that could differentiate the two conditions with 80–95% accuracy based on their genome-wide binding levels. Our top models consisted of approximately 10 protein factors selected from ~200 factors based on how well they can distinguish genes under normal condition or genes under heat shock condition when interacting with other protein factors. We replicated several protein factors known to be involved in heat-shock or stress related responses and we also identified a few novel factors, especially at the gene terminator regions. The network models showed that the protein factors exhibited non-additive relationships, indicating the presence of interactions among the protein factors. We also uncovered different regulation networks at the gene promoter regions and gene terminator regions, which shows that different sets of factors were affected by heat shock at distinctive biological locations. This machine learning approach demonstrates the ability to uncover interactions in protein binding data with high accuracy and it has the potential to be extended to any type of genomics data.

599W

Human neuropathy target esterase rescues SWS Drosophila neurodegeneration. *N.D. Hein[1], A.K. Sujkowski[2], A.M. Taylor[1], R.J. Tabano[1], A. Harrell[1], S. Rainier[1], R.J. Wessell[2], J.K. Fink[1,3].* 1) Neurology, University of Michigan, Ann Arbor, MI; 2) Internal Medicine, University of Michigan, Ann Arbor, MI; 3) Geriatric Research Education and Clinical Center, Ann Arbor Veterans Affairs Medical Center.

Neuropathy target esterase (NTE) is a membrane phospholipase A2 with cAMP-dependent protein kinase (PKA) regulatory activity. NTE inhibition by organophosphorous (OP) compounds leads to chronic OP-induced delayed neuropathy. NTE mutations cause autosomal recessive motor neuron disease (NTE-MND), a form of hereditary spastic paraplegia (SPG39).

Drosophila Swiss cheese (SWS) shares 39% protein sequence identity with human NTE (hNTE). Drosophila SWS mutants exhibit progressive locomotor impairment, early lethality, and vacuolar degeneration and glial hyperwrapping in brain. SWS^5 has mutation G648R. SWS Drosophila neurodegeneration can be rescued by wildtype sws or wildtype murine NTE transgenes.

We examined the ability of wildtype hNTE to ameliorate the SWS neurogenerative phenotype. Flies transgenic for RU486-inducible elav-GAL4 (generated by A. Sujkowski) were crossed with transgenic UAS-NTE^{wt} flies; and progeny crossed with SWS^5 flies (generously provided by Dr. Doris Kretzchmar, Oregon Science & Health Univ.) and control flies to create flies with RU486-inducible, neuron-specific $hNTE^{wt}$ expression. Flies were maintained at 29°C and locomotor ability assessed at day 10 in a run-to-exhaustion protocol using iterative negative geotaxis (Tinkerhess et al, 2012). NTE protein (western blot), PKA activity, and NTE activity (phenyl valerate substrate) were assessed using published methods.

Western blot analysis confirmed the presence of hNTE protein in flies expressing wildtype hNTE (RU486-induced SWS^5, $elav^{GAL4\text{-}hNTE}$ and Berlin $K^{elavGAL4\text{-}hNTE}$ control flies). Locomotor activity of SWS^5 flies was approx. half that of wildtype flies. In contrast, SWS^5 flies expressing wildtype NTE (RU486-induced $SWS^{5/elavGAL4\ hNTE}$) had locomotor ability that was very similar to control flies. Histopathology studies of SWS^5, $SWS^{5/elavGAL4\ hNTE}$, and control flies is in progress.

Conclusion. These studies demonstrate the capacity of hNTE to functionally compensate for the SWS^5 Drosophila neurodegenerative phenotype. Together with the previously reported correction of SWS mutant phenotype by mNTE, these studies underscore the functional homology between mammalian NTE and SWS proteins. This *in vivo* assay of hNTE functional activity can be used to evaluate the functional significance of identified hNTE variants, including those in NTE-MND subjects.

600T

Microbiomic profiles and clinicopathologic outcome markers in oropharyngeal cancers. *P. Funchain[1], G. Bebek[1,3], K. Bennett[1], F. Niazi[1], N. Fowler[4], B. Burkey[4], C. Eng[1,2].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Department of Genetics and CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; 3) Case Center for Proteomics and Bioinformatics Case Western Reserve University, Cleveland, OH; 4) Head and Neck Institute, Cleveland Clinic, Cleveland, OH.

Recent studies of the human microbiome suggest that non-human genetic material from the spectrum of commensal microorganisms which inhabit the human body, otherwise known as the human microbiome, provide a large contribution to human health and disease. In particular, investigations of the aerodigestive tract implicate the importance of the microbiome to immune-related disease such as inflammatory bowel disease and in cancers including gastric, colon and pancreas. Here, we hypothesize that microbiomes differ between tumor and adjacent normal tissue in persons with head and neck squamous cell carcinoma (HNSCC), and propose a relation to epigenetic modifications in inflammation- and HNSCC-associated genes. Matched tumor and adjacent normal fresh frozen tissue specimens were collected from 55 prospectively enrolled HNSCC patients. Metagenomic profiles were obtained by Sanger sequencing of 16S rDNA PCR products. Methylation status of four genes previously linked to HNSCC or inflammation (MDR1, IL8, RARB, TGFBR2) was assessed in 49 samples. Principle component analysis was used to identify bacterial subpopulations significantly associated with HNSCC and relevant clinical variables. In this cohort, median age is 62, male:female ratio is 3:2, 16% are HPV+. We find that microbial populations can separate tumors by tobacco ($p<0.005$) but not alcohol ($p>0.6$) status. Specific bacterial subpopulations, including Spirochaetes and Bacillales, significantly associate with HNSCC over normal tissue ($p<0.01$). MDR1 promoter methylation is seen in tumor samples but not in normal oral mucosa (22/49 vs 0/49), and associates with specific microbiomic subpopulations including Enterobacteriaceae and Tenericutes. Importantly, bacterial profiles are independent of HPV status. We conclude that the microbiomes of normal and tumor tissue in persons with HNSCC differ. Preliminary analysis suggests specific bacterial subgroups associate with specific clinicopathologic features, which suggesting a role for metagenomics in HNSCC as a novel area of investigation for diagnosis, prognostication, and therapeutic targeting.

601F

Metagenomic diagnosis of culture-negative infective endocarditis. *A. IMAI[1], Y. ASANO[1], S. NAKAMURA[2], K. GOTOH[2], D. MOTOOKA[2], Y. SAKATA[1], T. IIDA[2], T. HORII[2], S. TAKASHIMA[1].* 1) Department of Cardiovascular Medicine, Osaka University Graduate School of Medicine, Suita, Osaka, Japan; 2) Department of Genome Informatics, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan.

Infective endocarditis (IE) is still a lethal disease and detecting causative microorganisms is essential for choosing an optimal antibiotic and effective dosing periods. However, culture-negative IE is frequently observed, especially among cases that included premedication with antimicrobial agents. Conventional valve culture has disadvantages that it has low sensitivity and can detect only viable and target microorganisms with the culture medium currently available. Recently, the metagenomic approach using next generation sequencing (NGS) technology emerged as a comprehensive method for exploring causative agents of infectious diseases without prior culture .Therefore, we assessed the viability of metagenomic analyses to detect causative microorganisms in resected valves from IE patients who had underwent valve replacement surgery. We studied two representative cases, each of which is a patient with culture-positive IE and culture-negative IE, respectively. After the operation of a valve replacement therapy for IE we assessed causative bacteria in the resected valve both by cultivation survey and by metagenomic sequencing approach. The former case was affected with IE of the native aortic valve. The blood culture at the primary hospital was positive and negative at the time of referral to our hospital with antibiotic pretreatment. After the operation the resected valve culture was positive for the same bacteria species as detected in the previous hospital. By defining this patient as the case with culture-positive IE, we applied metagenomic approach to detect bacterial genome fragments with genome DNA extracted from the resected valve. Using BLAST search, the dominant parts in bacteria-derived sequences were mapped on the genomes of candidate bacterial species. The latter case was also affected with IE of the native aortic valve. There was no medical record of any cultivation test in previous hospital. Both the blood and valve cultures were negative at our hospital, assumed to be due to oral antibiotic treatment prior to admission. Metagenomic data of the resected valve showed S. sanguinis as a causative bacterium confirmed by Gram-staining in pathological specimen, in which Gram-positive cocci were identified in the valve tissue. In conclusion, comprehensive metagenomic approach using NGS could detect causative organisms in culture-negative IE case.

602W

Exome Capture from Saliva Produces High Quality Genomic and Metagenomic Data. *D. Bobo[1], T. Sharpton[4], P. Norman[5], M. Carpenter[3], M. Sikora[3], C. Gignoux[6], N. Gorgani[5], M. Guadalupe[7], K. Pollard[4,8], P. Parham[5], M. Feldman[9], J. Wall[8], C. Bustamante[3], J. Kidd[2], B. Henn[1].* 1) Ecology & Evolution, Stony Brook University, Stony Brook, NY; 2) Departments of Human Genetics, and Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Department of Genetics, Stanford University, Stanford, CA 94305; 4) The J. David Gladstone Institutes, University of California, San Francisco, San Francisco CA, 94158; 5) Department of Structural Biology, Stanford University, Stanford CA 94305; 6) University of California, San Francisco, CA, 94158; 7) Agilent Technologies, Genomics Division, Cedar Creek, TX 78612; 8) Institute for Human Genetics, and the Departments of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco CA, 94143; 9) Department of Biological Sciences, Stanford University, Stanford CA 94305.

Targeted capture of genomic regions reduces sequencing cost while generating higher coverage by allowing biomedical researchers to focus on specific loci of interest, such as exons. Targeted capture also has the potential to facilitate the generation of genomic data from DNA collected via saliva or buccal cells. DNA samples derived from these cell types tend to have a lower human DNA yield, may be degraded from age and/or have contamination from bacterial or other ambient oral flora. However, thousands of samples have been previously collected from these tissue types, and saliva collection has the advantage that it is a non-invasive form of collection, appropriate for a wide variety of research. We demonstrate successful enrichment and sequencing of 15 South African KhoeSan exomes with samples initially derived from saliva. The expanded exome dataset enable us to make several novel discoveries. First, the KhoeSan hunter-gatherers exhibit extremely high levels of genetic diversity in comparison to other individuals of African ancestry. Second, we discover and independently verify twenty previously unknown KIR alleles using methods we developed to accurately map and call the highly polymorphic HLA and KIR loci from exome capture data. Finally, we show that capture of saliva-derived DNA yields sufficient non-human sequences to characterize oral microbial communities, including detection of bacteria linked to oral disease (e.g. Prevotella melaninogenica). Metagenomic profiles obtained from exome-capture results are similar to those obtained from direct sequencing of saliva derived DNA, indicating that metagenomic analysis of saliva derived samples holds promise for future metagenomic studies as a 'free' addition to human exome sequencing.

603T

Host genetic control of the human gut microbiome. *J.K. Goodrich[1,2], R. Blekhman[2,3], O. Koren[1,2], A.C. Poole[1,2], M. Beaumont[4], J.T. Bell[4], T.D. Spector[4], A.G. Clark[2], R.E. Ley[1,2].* 1) Dept of Microbiology, Cornell University, Ithaca, NY, USA; 2) Dept of Molecular Biology & Genetics, Cornell University, Ithaca, NY, USA; 3) Dept of Neurology & Neuroscience, Weill Cornell Medical College, New York, NY USA; 4) Dept of Twin Research & Genetic Epidemiology, King's College, London, U.K.

The gut microbiome differs markedly among individuals and is increasingly viewed as a risk factor in chronic diseases such as obesity, inflammatory bowel disease, and diabetes. The composition of the gut microbiome is a target for emerging therapies, which require an understanding of the factors shaping the microbiome, including host lifestyle, physiology, health, and their interactions. Importantly, the role of host genotype in modulating microbiota composition remains to be elucidated. In animal models, host gene deletions often shift microbiota composition, and genetic mapping has linked specific loci with gut microbial abundances. Similar studies are lacking in humans, however, recent twin-based studies comparing microbiota of monozygotic (MZ) and dizygotic (DZ) twin pairs failed to detect a genotype effect, possibly due to small sample sizes. Here, we characterized the faecal microbiota of 572 genotyped adults, including 114 MZ twin pairs and 159 DZ twin pairs. We determined the degree to which the microbial tree is heritable using the classical ACE twin model, which partitions the total variance of a phenotype into additive genetic effects (A), common environment (C), and unique environment (E). We identified components of the gut microbiota that have moderate heritability such as members of the families Ruminococcaceae, Lachnospiraceae, and Catabacteriaceae. We validated these findings by applying this approach to recently published data from two smaller scale studies of US twins. Furthermore, abundances of some bacteria are associated with specific genetic variants implicated in inflammatory bowel disease. Our results indicate that while the composition of an individual's microbiota is largely influenced by environmental factors, host genetics also plays an appreciable role for specific bacterial taxa. This report highlights the importance of understanding the effects of host genetics on the microbiome, and how host-microbe interactions can influence an individual's susceptibility to disease.

604F

Temporal variation in human gut microbiome composition in the Hutterites. *E.R. Davenport[1], O. Mizrahi-Man[1], K. Michelini[1], L.B. Barreiro[2], C. Ober[1], Y. Gilad[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Sainte Justine Hospital Research Centre Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, QC, Canada.

The bacterial composition of the human fecal microbiome is influenced by several environmental factors. Diet is believed to be one of the most important determinants, however, the extent to which dietary fluctuations alter the gut microbiome over time is unknown, both within an individual and between individuals. In this study, we examined the relationship between seasonal variation in produce consumption and variation in gut microbiome composition. To do so, we sampled stool and collected dietary survey information for 60 individuals of a founder population, the Hutterites, during both summer and winter months. These individuals live and eat communally; therefore many environmental factors - including diet - are similar across individuals. Additionally, menus are remarkably stable throughout the year, with the exception that fresh produce is primarily served during the summer and autumn months, allowing us to examine the association of produce availability to gut microbiome composition. We observed that although there is stability in the composition of the gut microbiome within individuals over time, there are also consistent and significant population-wide shifts in microbiome composition between seasons. Seasonal differences were detected in both (i) the abundance of particular taxa (FDR <0.05), including highly abundant phyla Bacteroidetes and Firmicutes, and (ii) gut microbiome diversity (by Shannon diversity; P = 0.001). Seasonal dietary differences in produce consumption likely explain, at least in part, the seasonal compositional shifts observed in the gut microbiota. For example, high levels of fresh produce consumed during the summer, containing complex starch and fiber, might explain our observations of increased abundance of Bacteroidetes in summer, a phyla containing complex carbohydrate digesters, and decreased levels of Actinobacteria, which have previously been negatively correlated to fiber content in food questionnaires. In conclusion, our observations demonstrate the plastic nature of the human gut microbiome in response to variation in diet.

605W

Using blood-born miRNA profiles and machine-learning techniques to predict inflammatory phenotypes. *M. Huebenthal[1], G. Hemmrich-Stanisak[1], Z.G. Du[1], S. Nikolaus[2], S. Zeissig[2], S. Schreiber[1,2], A. Franke[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; 2) Department of Internal Medicine, University Hospital Schleswig-Holstein, Kiel, Germany.

With accuracies ranging from 80% to 90% invasive methods such as ileocolonoscopy are still the standard for diagnosis of the complex diseases Crohn's disease (CD) and ulcerative colitis (UC). For these two major subphenotypes of inflammatory bowel disease (IBD) differences in miRNA expression can be shown. In this study we detected specific miRNA signatures, which allow stable distinction between the different phenotypes and therefore non-invasive prediction of patients' disease states. Based on 142 whole blood samples (62 CD, 64 UC and 16 healthy controls) expression profiles of 1733 miRNAs were determined using Small RNA sequencing (Illumina HiSeq2000). Classification and biomarker selection was performed using different types of Support Vector Machines and Recursive Feature Elimination, respectively. After model assessment based on manifold hold-out sampling sets of 107 miRNAs for CD and 48 miRNAs for UC were determined to distinguish between cases and healthy controls with estimated balanced accuracies of 87% and 94%. With balanced accuracy of 97% the distinction of IBD cases from healthy controls based on 93 miRNAs performed comparably well. Resulting in a balanced accuracy of 63%, 129 miRNAs were necessary to discriminate between CD and UC. In combination with the models constructed the reported sets of putative biomarkers allow for solving binary as well as multinomial classification tasks arising from the diagnosis of IBD. Our preliminary evaluation already yields to very high accuracies and is therefore of potential relevance for clinical application.

606T

Transcriptional profiling of lncRNAs reveals important biological roles in psoriasis. *L. Tsoi[1], M. Iyer[2], P. Stuart[3], T. Tejasvi[3], B. Li[4], J. Ding[5], J. Gudjonsson[3], H. Kang[1], R. Nair[3], A. Chinnaiyan[2], G. Abecasis[1], J. Elder[3].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI; 3) Dermatology, University of Michigan, Ann Arbor, MI; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 5) Laboratory of Genetics, National Institute on Aging, Baltimore, MD.

Psoriasis is an inflammatory disease of skin and joints affecting 0.1–2% of the world's population. Previous studies have identified crucial immune and epidermal-differentiation genes that are differentially expressed in the lesional skin of psoriatic patients. However, most of these studies assayed only protein-coding genes. The functional importance of long non-coding RNA (lncRNAs), which are known to play a role in human disease, is yet to be determined for psoriasis. In this study, we applied a computational approach and stringent filtering to predict novel lncRNAs in our RNAseq data comprised of 92 psoriatic and 82 normal skin samples, and we characterized the expression patterns of the identified lncRNAs to infer their biological functions. We identified 7,145 known lncRNAs, and further predicted 319 novel ones, yielding on average genomic density of 2.5 expressed lncRNAs per Mb. Although the proportions of differentially expressed genes for protein-coding and known lncRNAs were similar (18%), over 60% of our novel lncRNAs were differentially expressed. This result suggests the novel lncRNAs identified here have a skin-specific expression pattern and function. Notably, the transcript with the most significant correlation between expression levels and disease severity as assessed by the local psoriasis area and severity index (PASI) is a lncRNA. We used co-expression analysis to infer the biological functions of lncRNAs, and the most enriched functions include immune response ($p=2\times10^{-7}$) and cytokine interactions ($p=6\times10^{-9}$). Our results also indicate significant enrichment ($p=5\times10^{-3}$) of novel lncRNAs in the epidermal differentiation complex on chromosome 1q21. The characterization of expression signatures for lncRNAs provided by this study suggests an important role for these transcripts in disease ethiology.

607F

Deciphering and exploiting transcriptome-wide microRNA binding profiles in human brain. *R.L. Boudreau[1], P. Jiang[1], B.L. Gilmore[1], R.M. Spengler[1], R. Tirabassi[2], J.A. Nelson[2], C.A. Ross[3], Y. Xing[1], B.L. Davidson[1].* 1) University of Iowa College of Medicine, Iowa City, IA; 2) Oregon Health & Sciences University, Beaverton, OR; 3) Johns Hopkins University School of Medicine, Baltimore, MD.

The orchestration of brain function requires complex gene regulatory networks, which in part, are modulated by microRNAs (miRNAs). These non-coding RNAs associate with Argonaute (Ago) proteins to direct post-transcriptional gene suppression by binding to 3' untranslated regions (3'-UTRs), and growing evidence suggests that even slight aberrations in miRNA activities may alter synaptic function. To better understand how miRNAs contribute to human-specialized brain processes and neurological phenotypes, identifying their targets is of paramount importance. Here, we address the latter by profiling Ago2:RNA interactions using crosslinking immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) to generate the first transcriptome-wide map of miRNA binding sites in human brain. We uncovered ~7000 stringent Ago2 binding sites which are highly enriched for conserved sequences corresponding to abundant brain miRNAs. This dataset points to functional miRNA:target pairs across more than 3000 genes and represents a valuable resource for accelerating our understanding of miRNA function in the central nervous system. We explored this interactome for clinically-relevant miRNA binding sites and identified numerous disease-associated targets for miR-137, a miRNA implicated in schizophrenia. In addition, we discovered miRNA binding sites overlapping single nucleotide polymorphisms linked to common and rare neuropsychiatric conditions, including Parkinson's and Alzheimer's. These findings provide clues which may facilitate the translation of genetic studies of complex neuropsychiatric diseases into novel or refined pathogenic mechanisms and therapeutics. Overall, this work lays the foundation for translating this methodology to characterize the diverse landscapes of miRNA:target interactions throughout sub-anatomical brain structures and across normal and diseased tissues.

608W

Association of non-coding SNPs with decreased levels of miR-9 and alcoholism. *A. Pietrzykowski[1,2], Y. Wang[1], O. Anees[1], E. Mead[1], N. Kinstlinger[1], L. Tejeda[1], A. Hot[1].* 1) Animal Sciences, Rutgers University, New Brunswick, NJ; 2) Genetics, Rutgers University, New Brunswick, NJ.

Alcoholism has strong genetic underpinnings. The focus on protein-coding genes has only partially revealed the genetic basis of this complex, multigenic disorder. There is increasing recent evidence that microRNAs (miRNAs) play a key role in alcoholism and other types of addiction. miRNAs are produced by non-coding genes, whose products are small RNA molecules with a powerful role as master regulators of mRNA and protein expression. Many miRNAs can simultaneously control the expression of several genes. Our previous results using rodent models indicated that a particular miRNA, miR-9, is regulated by alcohol and has an essential role in the development of molecular tolerance to this drug. Here, we seek to determine an association of single nucleotide polymorphisms (SNPs) in the regions regulating miR-9 expression in alcoholism using human samples. miR-9 is encoded by three distinct genes, with each gene located within a larger host gene. Thus, six different promoters can control expression of mature miR-9. We used the UCSC Genome Browser to determine the length of each promoter (4 to 6 kb) and perform SNP discovery in all 6 regions. We used 282 alcoholic samples from the Collaborative Studies on Genetics of Alcoholism (COGA) collection, and 255 non-alcoholic controls from the NIMH control collection. Nested PCR and direct sequencing of PCR products were performed on each sample. GWAS-derived ancestry informative markers were used to correct for ancestry. Association, allele frequency, odds ratio, heterozygosity and Hardy-Weinberg equilibrium were determined for all samples. MAPPER2 or MatInspector were used to test the effects of SNPs on transcription factor binding. We established the presence of 142 SNPs in 2 categories: known SNPs (present in dbSNP) and new SNPs. 8 known SNPs independently have strong and significant association with alcoholism (they are a topic of a sister abstract submitted to the World Congress of Psychiatric Genetics). Interestingly, some of the rare new SNPs encode the same type of polymorphism, and can potentially change the binding of the same transcription factor. When analyzed as a group they have an association with alcoholism and could have a significant impact on miR-9 expression. Our results reveal the presence of alcoholism-associated SNPs in non-coding regions of the human genome important for miR-9 function. These findings identify novel genetic underpinnings and molecular mechanisms of the development of alcoholism.

609T

De novo discovery of distant regulatory elements by enhancer RNA expression. *H. Wu[1], A. Nord[1], J. Akiyama[1], M. Shoukry[1], V. Afzal[1], E. Rubin[1,2], L. Pennacchio[1,2], A. Visel[1,2].* 1) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA; 2) DOE Joint Genome Institute, Walnut Creek, CA.

Enhancer DNA elements can be actively transcribed to produce short and long transcripts, named enhancer RNAs or eRNAs. Recently, several groups showed that expression of eRNAs appears to correlate with enhancer activity in cell lines. However, whether this is true and can be used to identify tissue-specific in vivo enhancers remain unknown. Here we have investigated the expression dynamics of eRNAs in mouse embryonic tissues (E11.5), heart and limb. We show that active enhancers are transcribed in vivo, and there is global correlation between eRNA expression and tissue-specific enhancer activity. Both p300-dependent and -independent enhancers are marked by eRNA expression. By investigating the expression pattern of eRNA, it shows that the regulatory effects of enhancers are not limited to the nearest genes and the transcription of eRNA tend to be coordinated with their potential target genes. Taken together, this pilot study demonstrated that tissue-specific eRNA expression can be used as a novel method for in vivo enhancer predictions; the transcription of eRNAs and their targets may share common regulatory mechanisms.

610F

MicroRNA regulation in the inner ear and link to deafness. *K. Avraham[1], A. Rudnicki[1], K. Ushakov[1], O. Isakov[2], N. Shomron[2].* 1) Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 2) Department of Cell and Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Mutations in microRNAs (miRNAs) have been discovered to lead to deafness in both humans and mice. miRNAs play a critical role in the development and regulation of sensory systems, including the inner ear, which is responsible for hearing and balance in mammals. Mutations are associated with human hearing impairment (Mencia et al. *Nat Genet* 2009). The study of miRNAs in humans, however, has been hampered by the unavailability of inner ear RNA from human subjects, making the mouse an invaluable model for studying miRNAs in the inner ear. For example, removal of Dicer using the *Cre-loxP* recombination system, driven by a Pou4f3 hair cell-specific promoter, led to depletion of hair cell miRNAs and resulted in complete deafness by one month of age (Friedman et al. *PNAS* 2009). To further dissect the function of miRNAs in the mammalian inner ear, we performed RNA-Seq on RNA isolated from mouse inner ear sensory epithelia. qRT-PCR confirmed the expression of these miRNAs and *in situ* hybridization was used to identify their spatial expression in the mouse inner ear. We used bioinformatics software to predict the targets of these miRNAs and verified them by *in vitro* over-expression and luciferase assays. A total of 7,732,589 and 8,452,794 small RNAs were found in the cochlear and vestibular samples, respectively. These included miRNAs, snoRNAs, transfer RNAs and ribosomal RNAs. Reads were aligned to the mature *Mus musculus* miRNA database (http://www.mirbase.org). miRDeep2 was used for novel microRNA prediction. We chose miRNAs for further study with a seed region that is conserved between mouse and humans and with high expression in the inner ear. One example is a miRNA gene that is included in the intron of the gene *Tectb*, which is expressed in the mouse inner ear and is associated with deafness. The gene produces two mature miRNAs, 5p and 3p, and the human orthologues of this miRNA, hsa-miR-6715a-3p and hsa-miR-6715a-5p, were previously identified in a study on miRNAs and piRNAs in the human epididymis (Li et al. *Gene* 2012). Temporal and spatial expression revealed expression in the cochlear and vestibular sensory epithelia, and spiral and vestibular ganglia. Arhgap12 was found to be a target of miR-6715a-3p, a protein of the RhoGAP family, implicating this miRNA-target pair in cell adhesion, inner ear morphogenesis and actin reorganization. Interactions between inner ear miRNAs and targets will shed light on the role RNA regulation plays in the mammalian ear.

611W

Integrated microRNA and mRNA signature associated with the transition from the locally confined to the metastasized renal cell carcinoma. *J. Billaud[1], Z. Wotschofsky[2,3], K. Jung[2,3], H. Meyer[2,4].* 1) Ingenuity Systems, Redwood City, CA; 2) Department of Urology, Charité - Universitätmedizin Berlin, Berlin, Germany; 3) Berlin Institute for Urologic Research, Berlin, Germany; 4) Institute of Physiology, Charité - Universitätmedizin Berlin, Berlin, Germany.

MicroRNAs (miRNAs) are endogenous small non-coding RNAs that regulate gene expression by interfering translation or stability of target transcripts. One miRNA can interact with several hundred mRNAs, while one mRNA can be regulated by several miRNAs. This interplay between miRNA and their mRNA has been proposed as an important process in cancer development and progression. We have investigated molecular networks impacted by predicted mRNA targets of differentially expressed miRNAs in patients with clear cell renal cell carcinoma (ccRCC) diagnosed with or without metastasis. miRNA and mRNA microarray expression profiles derived from primary clear cell renal cell carcinomas from patients with (in total 16 samples) or without diagnosed metastasis (in total 22 samples) were used to identify anti-correlated miRNA-mRNA interaction in ccRCC. For this purpose, Ingenuity pathway analysis microRNA Target Filter, which enables prioritization of experimentally validated and predicted mRNA targets was used. By applying an expression pairing tool, the analysis was focused on targets exhibiting altered expression in our analysis, finding miRNAs and their target genes with opposite or same expression. The resulting identified interactions were revalidated by RT-qPCR in another cohort of RCC patients. The predicted miRNA-mRNA interactions were also tested by functional analyses using miRNA knock-down and over expression experiments in renal cancer cell lines. Among the significantly differentially expressed miRNAs, we have identified 3 miRNAs (miR-146a, miR-128a and miR-17-5p) that were upregulated in primary tumors from patients without metastasis and down regulated in primary tumors from patients with metastasis. We have further identified the mRNA targets which expression were inversely correlated to these 3 miRNAs, and have been previously experimentally demonstrated in cancer setting in humans. Specifically we showed that BRAC1, MCM10, CDKN3, UHRF1, IL8 were downregulated and targeted by miR-146a-5p. The relation between these identified target genes and miRNA-146a was validated in cell culture experiments. In conclusion, we have identified novel target genes of dysregulated miRNA which are involved in the transition from primary RCC without metastases into tumors generating distant metastasis.

612T

Global Patterns of miRNA Variation and Population-specific Differentiation Based on Whole Genome Sequence Data. *R.A. Rawlings-Goss, S. Tishkoff.* Genetics, University of Pennsylvania, Philadelphia, PA.

MicroRNAs (miRNA) are evolutionarily conserved regulators of protein expression, and are responsible for controlling transcription of up to 60% of expressed genes. For this reason, miRNA profiling strategies are increasingly under clinical trial as biomarkers for complex diseases such as cancer and diabetes. Despite the important role of miRNA in disease, normal miRNA functional variation among different ethnic groups remains an open question, specifically in under-sampled populations such as ethnically and geographically diverse Africans. Here, we examine worldwide variation in miRNA genes by analyzing whole genome sequencing (60x coverage) in a panel of 69 unrelated individuals from 14 different ethnic groups. We include diverse African groups, not yet included in HapMap or 1000 genomes datasets, in an effort to identify novel human variation within miRNAs, as well as population-differentiated miRNA variants. We identified 198 novel variants within miRNAs, not present in dbSNP, with 29 novel variants located in the highly conserved seed sequence of mature miRNA. Additionally, between our African and non-African samples we observe highly population differentiated miRNA (PD-miRNA) variants, suggestive of global differences in miRNA landscape. Seven of the highly differentiated PD-miRNA are currently associated with disease (including several forms of cancer). Taken together, we outline PD-miRNA variants that could be acting individually or in combination to alter expression of tumor sensitive genes and hypothesize that this variation is contributing in part to the observed ethnic disparities seen in some forms of cancer.

613F

Evaluation and identification of master regulatory microRNAs on protein levels. *A.L. Stark[1], R.J. Hause, Jr[2, 3, 4], R.B. Jones[2, 3, 4], M.E. Dolan[1].* 1) Dept of Medicine, Univ Chicago, Chicago, IL; 2) Committee on Genetics, Genomics, and Systems Biology, Univ Chicago, Chicago, IL; 3) Ben May Institute for Cancer Research, Univ Chicago, Chicago, IL; 4) Institute for Genomic and Systems Biology, Univ Chicago, Chicago, IL.

To elucidate a clearer understanding of the mechanisms involved in the cellular dynamics of regulation of gene expression, there has been considerable focus on studying its relationship to genetics, epigenetics, and microRNA. Many studies have focused on the relationship between miRNA and mRNA expression; thus, we sought to expand the regulatory understanding of the role of miRNA by evaluating its relationship to protein expression levels. We quantified 220 miRNAs and 441 protein isoforms across 68 LCLs derived from Yoruba cell lines from Ibadan, Nigeria. Two miRNAs, mir768-3p and mir29c, correlated with STAT3 protein expression with a Bonferroni-corrected $p < 0.05$. These relationships were not predicted bioinformatically or at the mRNA level. We then evaluated all miRNAs for the number of protein levels correlated with each miRNA. Using permutation analysis, each miRNA was found to be associated with more protein levels than would be expected by chance. This enrichment was robust independent of the significance level of the protein-miRNA relationship ($p < 0.05$, $p < 0.01$, $p < 0.005$). We identified two master regulatory miRNAs whose expression levels were associated with greater than ten other proteins at $p < 0.001$: mir125a-5p and mir768-3p. Expression of mir768-3p was significantly associated ($p < 0.01$) with response to five different chemotherapeutic agents (cisplatin, carboplatin, cytarabine, etoposide, and daunorubicin). Mir125a-5p has been widely implicated in the literature for its involvement in many complex processes, including carcinogenesis, inflammation, and metabolism. Our work annotates protein levels that may be regulated by microRNA. Further study of the role of microRNAs in protein expression will lead to a better understanding of dynamic cellular processes and pharmacologic phenotypes.

614W

Study of microRNAs regulated by hypoxia in cells latently infected by Kaposi's sarcoma-associated virus. *C. Viollet-Djelassi[1,2], D. Davis[2], C. Camps[1], M. Reczko[3], F. Pezzella[1], R. Yarchoan[2], I. Ragoussis[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) HIV and AIDS Malignancy Branch, Center for Cancer Research, National Cancer Institute, National Institute of Health, Bethesda, MD; 3) BSRC Al. Fleming, Vari, Athens, Greece.

Kaposi's sarcoma-associated herpesvirus (KSHV) infection, especially when combined with immunodeficiency, can lead to cancers such as Kaposi's sarcoma (KS), primary effusion lymphoma or multicentric Castleman's disease. KS is a highly vascular tumour, and KSHV genes have been shown to be induced by hypoxia and in turn to activate hypoxia-inducible factors, indicating a role of hypoxia in the KSHV life cycle. Additionally, virus-encoded microRNAs (miRNAs) have been increasingly recognised as contributors to viral cancer pathogenesis. Investigating their role in KSHV-related diseases is therefore of interest. We hypothesise that hypoxia and/or KSHV infection will alter cellular and viral miRNA levels that play a role in cancer progression and viral pathogenesis. To address this, we compared miRNA expression profiles of KSHV infected cells and non-infected cells under hypoxia and normoxia using Illumina small RNA and total RNA deep sequencing. In infected endothelial cells, we found that 112 mature human miRNAs and 4 viral miRNAs are differentially expressed in hypoxia compared to normoxia. In particular, hsa-miR-663b was the most down-regulated miRNA in hypoxia (>>32-fold), while hsa-miR-210 was one of the most up-regulated (7.275-fold) and its increases in 1% O_2 were further validated by qRT-PCR in different KSHV-positive cell lines (B-cells BCBL1 and endothelial cells SLKK). Interestingly, KSHV-miR-K12-3-3p, a known promoter of KSHV latency, was also significantly down-regulated in hypoxic infected cells (2.129-fold). Furthermore, the expression patterns of total RNA and miRNAs were analysed using Ingenuity Pathway Analysis (IPA). Canonical pathways including glycolysis, atherosclerosis signalling and coagulation system were found differentially regulated by hypoxia compared to normoxia in endothelial SLKK cells. The effect of the host derived and virus derived miRNAs on the expression patterns of both host and virus RNAs is being elucidated through integrated analysis in order to identify targets. The target mRNAs and respective proteins are being investigated in relation to KSHV-associated diseases and the hypoxic pathway. The outcomes of the present study will aid our understanding of how KSHV uses the host RNA silencing machinery to its advantage and provides clues as to how this intersects with the use of the cell's response to hypoxia. This research was supported by the Wellcome Trust and the Intramural Research Program of the NIH, NCI.

615T

Functional Assessment of snoRNA derived microRNAs in Prader-Willi Syndrome. *V. Williamson, M. Mamdani, G. McMichael, V. Vladimirov.* Psychiatry, Virginia InstPsychiatric & Behavioral Genetics, Richmond, VA.

Prader-Willi Syndrome (PWS) is a neurodevelopmental disorder, characterized by hyperphagia, obesity and self-inflicting behaviors such as skin-picking, temper tantrums and impulsivity. Approximately, 70% of subjects with PWS carry a deletion on chromosome 15 (q11.1-11.3) containing a small nucleolar RNA (snoRNA) cluster, HBII-85. Recently, snoRNA were shown as a source for other small non-coding (ncRNAs) molecules, called microRNA (miRNA). MiRNAs were shown to have important neurodevelopmental properties through their ability to control gene expression. Our hypothesis is that HBII-85 affects PWS development through HBII-85 derived miRNAs and that deletion of this region in PWS eliminates miRNAs integral to proper neuronal functioning. Sequence analysis of the region surrounding the HBII-85 cluster (HBII-85-12, -16, -17, -18, -19, -21, -22) in publically available deep sequencing datasets suggests that multiple miRNAs are possible and expressed at consistently higher levels in specific cell lines such as HL60 cells. Further, target prediction using an intersection of the programs PITA, MiRanda, and TargetScan from a custom database (N = 349) for these HBII-85 "snomiRs" yielded 41 potential gene targets. The database included genes previously implicated in PWS studies as well as those identified through ontology terms neuronal or brain development. When compared to predictions generated using an non PWS related cancer database (N = 487) we see statistically significant enrichment (16% vs. 8.4%, df = 1; $p = 0.003$, $\psi 2$-test, two tailed). Additionally the predicted targets of two non-PWS, non-cancer miRNAs, hsa-mir-132 and -137 demonstrate no statistically significant enrichment of targets in either database (9.7% vs. 10.8%, df = 1, $p = 0.32$, $\psi 2$-test, two tailed). Luciferase (luc) and RNA immunoprecipitation analyses of the proposed relationship between these likely miRNA candidates and their gene targets is on-going. If successful, this study demonstrates a decidedly new view of the disorder and presents the possibility for new therapies in the future.

616F

New clues that shape microRNA expression in human neurospheres. *C. Palacios[1, 2], A. Espinoza[3], A. Contreras[4], N. Najera[1], I. Rubio[1], A. Conejo[2], I. Ita[1], I. Palma[1, 5].* 1) Laboratorio de Morfología Celular y Molecular, Sección de Estudios de Posgrado e Investigación, Escuela Superior de Medicina, Instituto Politécnico Nacional, Plan de San Luis y Díaz Mirón, SN, Col. Casco de Santo Tomás, México, 11340, D.F., México; 2) Unidad de Investigación Médica en Genética Humana, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Av. Cuauhtémoc # 330 Col. Doctores. 06720 México, D.F., México; 3) Unidad de Medicina Genómica, Hospital General de México, Eje 2A Sur (Dr. Balmis) 148 Doctores, Cuauhtémoc, 06726 México, D.F., México; 4) Departamento de Biología del Desarrollo y Teratogénesis. Hospital Infantil de México Federico Gómez. Doctor Márquez 162 Doctores, Cuauhtémoc, 06720 México, D.F., México; 5) 7 Departamento de Morfología, Facultad de Medicina Veterinaria y Zootecnia (FMVZ); Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria No. 3000, 04510, México D.F., México.

The olfactory neuroepithelium (ONE) represents a site where neural stem cells (NSC) can be obtained, as well as being very accessible. These cells can be cultured by neurosphere assay, forming spherical clusters of multipotent and progenitor cells in suspension with the ability to differentiate. These features make it a model for studying neural molecular and cellular processes. The ability of these cells to maintain an undifferentiated state and their capacity to differentiate requires that the cell can modify its expression, therefore depending on epigenetic mechanisms modulating molecules that promote the multipotentiality and that suppress cell differentiation. One such mechanism depends on microRNAs, small molecules of non-coding RNAs with roles in the regulation of expression. Its functions are both, neural lineage determination and in adult neural tissue functions. Obtaining human NSCs is a powerful tool to investigate neural processes. With this in mind, we cultured ONE cells from nasal swabs from 4 healthy volunteers. Neural marker expression was determined by by immunofluorescence. Subsequently a miRNA expression assay was performed with plates TLDA A y B of applied V 2.0. Cq values equal or greater than 36 were discarted. We assessed the expression level according to the DCt method, normalizing with the global mean and forming groups according to the mean plus / minus 1 or 2 standard deviations. Finally, we performed an in silico analysis looking for possible nervous system pathways in which expression of microRNAs are involved. We analyzed a total of 667 microRNAs. 246 microRNAs were expressed in at least 3 of the 4 samples. According to the expression level, 56 miRNAs had very high expression (23%), 23 high (9%), 66 medium (27%), 34 low (14%) and 67 very low (28%). Some of the microRNAs with higher expression are miR-222, miR-200c, miR-191, miR-30a *, miR-30e *, miR-484, miR-146a, miR-378, miR-24 miR-574-3p. Some miRNAs are already described with roles in neurogenesis as miR-184, miR-132 family members' miR-200c, and miR-30, among others. As for the expressed miRNAs possible role, predictions point to a possible involvement in neurodevelopmental pathways as well as in neural pathologies.

617W

LncRNAs regulation in insulin resistance of the adipose tissue. *M. Pradas-Juni[1,2], X. Bofill[4], F. Hanzu[2,3], C. Fillat[4], R. Gomis[1,2,3], E. Fernandez-Rebollo[1,2].* 1) Diabetes and Obesity Research Laboratory, IDIBAPS, Barcelona, Spain; 2) Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Diseases (CIBERDEM), Barcelona, Spain; 3) Endocrinology Unit, Hospital Clínic de Barcelona, Barcelona, Spain; 4) Gene Therapy and Cancer, IDIBAPS, Barcelona, Spain.

The non-coding RNAs (ncRNAs) have been largely underestimated but lately are being recognized as essential regulators of translational regulation and other processes. The ncRNAs are aberrantly expressed in a variety of human diseases; and thus demonstrating potential roles in cellular development and metabolism. NcRNAs are classified into two major classes based on transcript size: small and long ncRNAs (lncRNAs). LncRNAs are mRNA-like transcripts longer than 200 nucleotides, lacking significant open reading frames, generally are transcribed by RNA polymerase II and are emerging as key regulators playing a major biological role in epigenetics, alternative splicing and even as regulators of mRNA decay. On the other hand, type 2 diabetes mellitus (T2DM) is the most common metabolic disorder in the world and obesity, meaning visceral adiposity, is the core problem. T2DM is characterized by hyperglycemia and impaired insulin action and/or secretion. The main role of obesity in T2DM is due to the visceral adipose tissue (VAT) insulin resistance. Based on this information we hypothesize that lncRNAs are key regulators in the insulin resistance of the VAT in T2DM. Aim: Characterize the lncRNAs and their 'co-regulated' mRNAs involved in insulin resistance of VAT. Methods: VAT from 8 obese and 8 obese with T2DM patients was used for RNA isolation to perform human lncRNA microarray V2.0 (Arraystar), which is designed for the global profiling of human lncRNAs (33,945) and protein-coding transcripts (30,215 mRNAs). To identify the highly correlated mRNA and lncRNA pairs, coding-non-coding gene co-expression network analysis (CNC) was performed. Results: First analysis identified 1,859 upregulated and 286 downregulated lncRNAs, and 1,136 upregulated and 238 downregulated mRNAs in T2DM obese patients compared to the obese controls (Fold Change ≥ 2.0, P-value ≤ 0.05). Specificity was increased by reducing the P-value (≤ 0.01) and obtaining 215 CNCs, and taking into account the lncRNA genomic localization in respect to mRNA, we obtained 40 CNCs. Finally these 40 CNCs were confirmed by qRT-PCR in three groups (n=4/group): obese, obese with T2DM and lean. This last filter shows that 9 CNCs are clearly deregulated in adipocytes from obese patients with T2DM. We are currently performing loss and gain-of-function approaches in human visceral adipocyte cell lines to understand the 9 CNCs function in adipogenesis and increased insulin resistance in the T2DM obese patients.

618T

Characterization of piRNA genomic distribution and expression variation in human individuals. *J. Xing[1,2], H. Ha[1,2], J. Song[1,3], S. Wang[1,2], K.C. Chen[1,3].* 1) Dept of Genetics; 2) Human Genetic Institute of New Jersey; 3) BioMaPS Institute for Quantitative Biology; Rutgers, The State University of New Jersey, Piscataway, NJ.

Piwi-interacting RNAs (piRNAs) are a class of recently discovered small non-coding RNAs whose best known function is to repress mobile element activity in animal germlines. To date, virtually all piRNA studies have been conducted in model organisms and little is known about piRNA diversity, target specificity and the mechanism of mobile element regulation in humans. In this study, we performed high-throughput piRNA sequencing in three human testis samples and generated over 50 million putative piRNA reads. Using this data, we characterized the piRNA cluster distribution across the human genome and measured the extent of piRNA expression variation among the three samples. Overall we identified ~10,000 piRNA clusters in the human genome. The piRNA clusters range from 1kb to 276kb in size and occupy ~3% of the genome. PiRNA clusters within genes are enriched in the 3' UTR region, consistent with previous findings in mouse. To examine the role of piRNAs in mobile element regulation, we determined the piRNA mapping density in the consensus sequence of Alu and L1 elements, the most abundant mobile elements in the human genome. We found that piRNAs preferentially mapped to specific motifs in the consensus of these mobile elements, and some piRNA mapping peaks show sequence/position patterns consistent with the ping-pong mechanism. In addition, we showed that human piRNA clusters show smaller variation in their expression level among human individuals than in Drosophila. Our study provide a comprehensive characterization of piRNA diversity and their interaction with genes and mobile elements in humans.

619F

Alternative splicing of MAP/microtubule affinity-regulating kinase 4 (MARK4) in glioma: search for involvement of Polypyrimidine-tract binding protein (PTB). *L. Fontana, D. Rovina, C. Novielli, I. Magnani, L. Larizza.* Medical Genetics, Department of Health Sciences, Università degli Studi di Milano, Milan, Italy.

MARK4 is a Ser-Thr kinase that phosphorylates MAPs taking part in the regulation of microtubule dynamics involved in cell cycle control. The MARK4 gene encodes two alternatively spliced isoforms: the canonical MARK4S, featuring 18 exons, and the alternative MARK4L, derived from skipping of exon 16. In glioma we pointed out an imbalance between the MARK4 isoforms with decreased MARK4S expression associated with overexpression of MARK4L. A high L/S ratio also characterizes human glioblastoma-derived stem cells and mouse neural stem cells and appears proportional to cellular de-differentiation and tumor grade. Since the deregulation of MARK4 expression in glioma is not due to mutations or copy number loss/gain, we hinted that alterations in alternative splicing (AS) may be at the origin of the observed MARK4 isoforms imbalance. It's well known that specific splice variants are commonly enriched in cancers as a consequence of splicing factors up-regulation. In glioma, in particular, overexpression of PTB, a key component in regulating neural stem cells proliferation and differentiation, drives an oncogenic splicing switch favoring isoforms, like MARK4L, derived from exon skipping. Bioinformatic analysis of the MARK4 sequence by SFMap software revealed three putative PTB binding sites in both introns flanking exon 16. A functional role of these sites is suggested by the high conservation in mouse, as confirmed by CLUSTAL W alignment, and by the surrounding polypyrimidine rich context, required for PTB activity. Western blot analysis showed a significant overexpression of PTB in our astrocytoma and glioblastoma samples, that correlates with MARK4L expression. Deletion of the last 80 nt of MARK4 IVS15 foster exon 16 inclusion in a splicing minigene system, revealing the presence of a functional intronic splicing silencer (ISS) in this region. However, mutagenesis of the predicted PTB binding site contained in the deletion does not affect minigene splicing, suggesting that PTB may bind to a non canonical ISS. Sequential deletions and mRNA-protein immunoprecipitation are in progress to validate the involvement of PTB in MARK4 AS. Our data suggest that PTB overexpression in glioma may favor MARK4L expression causing the observed imbalance between the two isoforms. The identification of the splicing factors involved in MARK4 AS appears essential to understand the basis of the imbalance observed in gliomas and to identify underlying oncogenic pathways.

620W

A Synonymous Change, p.Gly16Gly in MECP2 Exon 1, Causes a Cryptic Splice Event in a Rett Syndrome Patient. *T.I. Sheikh[1,2], J. Vincent[1,2].* 1) Institute of Medical Sciences, University of Toronto, Toronto ON Canada; 2) Center of Addiction and Mental Health, Toronto ON Canada.

Background: Mutations in MECP2 are the main cause of Rett Syndrome. To date, no pathogenic synonymous MECP2 mutation has yet been identified. Here, we investigated a de novo synonymous variant c.48C>T (p.Gly16Gly) identified in a girl displaying a possible typical RTT phenotype. Methods: In silico analyses to predict effects of sequence on mRNA splicing were employed, followed by sequencing and quantification of lymphocyte mRNAs from the subject for splice variants MECP2_E1 and MECP2_E2. Results: Analysis of mRNA confirmed predictions that this synonymous mutation activates a splice-donor site at an early position in exon 1, leading to a deletion (r.[=, 48_63del]), codon frameshift and premature stop codon (p.Glu17Lysfs*16) for MECP2_E1. For MECP2_E2, the same premature splice site is used, but as this is located in the 5'untranslated region, no effect on the amino acid sequence is predicted. Quantitative analysis specific to this cryptic splice variant also revealed a significant decrease in the quantity of the correct MECP2_E1 transcript, which therefore suggests that this is the etiologically significant mutation in this patient. Conclusion: These findings suggest that synonymous variants of MECP2 as well as other known disease genes~and de novo variants in particular~ should be re-evaluated for potential effects on splicing.

621T

Role of RNA Editing in ER Stress Response. *A. Richards[1], I. Wang[2], V. Cheung[2,3,4].* 1) Cell and Molecular Biology Graduate Program, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA; 4) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA 19104, USA.

Adenosine Deaminase Acting on RNA (ADAR) deaminates adenosine to inosine in double-stranded RNA. There are thousands of adenosine to inosine editing events in human cells. While some editing sites, such as those in ion channels, including the AMPA receptor, have been well studied, the role of RNA editing in a more general cellular context is less well understood. In this project, we aim to understand how RNA editing of genes involved in the endoplasmic reticulum (ER) and the Golgi affects their functions, particularly in ER stress response.

We have identified ~1,000 human genes edited by ADAR, as confirmed by a decrease in editing levels following ADAR knock-down. Among them, 108 are genes localized to the endoplasmic reticulum and the Golgi body. To study these editing events, we sequenced the mRNA and corresponding DNA from B-cells of 10 individuals. We found about 710 editing sites within these 108 ADAR target genes; on average, there are 6 editing sites and a range of one to 54 sites per gene. The editing levels vary across individuals. For example, the editing level of a site in the 3' UTR of *EIF2AK2* varies by greater than 4-fold across individuals (range 13%-57%). In addition, the editing levels differ at different sites. To study if editing levels change following cellular stress, we treated B-cells of 10 individuals with tunicamycin to induce ER stress. We then sequenced the DNA and mRNA of these cells before and at 2 and 8 hours following ER stress. About 275 of the 710 editing sites in ER and Golgi genes show an increase in editing level of >1.5-fold following ER stress. Editing level changes following ER stress also show individual variability. For example, the editing level of a site in the 3' UTR of *VHL* increases 1.8- to 7-fold across individuals following ER stress. Over 60% of ADAR editing sites in the ER and Golgi genes are located in the 3' UTR. Therefore, we aimed to determine if RNA editing influences gene expression to affect ER stress response. We used luciferase reporter constructs with the edited sequence of 3' UTRs and found that regions of various 3' UTRs can repress or enhance reporter gene expression.

In this presentation, we will illustrate the features of RNA editing in ER and Golgi genes. We will, further, describe how RNA editing levels respond to ER stress and how editing, in turn, may impact ER stress response.

622F

Understanding translational regulation using RNA-Seq of Ribosome Protected mRNA fragments. *J. Pease[1], S. Kuersten[1], P. De Araujo[2], D. Vo[2], S. Burs[2], M. Qiao[2], A. Radek[1], E. Bahrami[3], P. Uren[3], A. Smith[3], L. Penalva[2].* 1) Epicentre, Madison, WI; 2) University of Texas Health Science Centre at San Antonio, San Antonio, TX; 3) University of Southern California, Los Angeles, CA.

Musashi1 (Msi1) is an evolutionarily conserved RNA-binding protein that has been implicated in a variety of cellular processes such as stem cell maintenance, nervous system development, and tumorigenesis. Msi1 is highly expressed in many cancers including glioblastoma, and is emerging as a potential therapeutic target in both regenerative medicine and cancer. Our goal is to better understand the regulatory role of Msi1 by identifying the RNAs targeted by Msi1 at the level of translation. Our approach is to use high-throughput RNA sequencing (RNA-Seq) to compare total mRNA to the Ribosome Protected mRNA Fragments (RPF) in order to provide quantitative analysis of translationally active mRNAs versus repressed pools of mRNAs present at a particular time or condition. Consequently, we developed a simplified method for preparing RNA-Seq libraries from Ribosome Protected fragments. A small hairpin (sh) RNA was used to knock down Msi1 activity in the glioblastoma cell line U251. RNA-Seq libraries were prepared from the ribosome protected fragments of control and shMsi1-treated cells and sequenced on the Illumina platform. Our results show RNA-Seq of ribosome protected fragments offers a viable approach to understanding the translational regulation role of Msi1.

623W

Integrator Complex Subunit 8 mutation associated with cortical and cerebellar malformations results in disruption of the Integrator complex and spliceosomal defects. *G.M. Mancini[1], R. Oegema[1], D. Baillat[4], R. Schot[1], D. Heijsman[2], L. van Unen[1], S. Kheradmand Kia[1], J. Hoogeboom[1], A. Kremer[2], F.W. Verheijen[1], P. van der Spek[2], R.M. Hofstra[1], E. Wagner[4], M. Fornerod[3].* 1) Clin Genetics Erasmus Med Ctr, 3000CA Rotterdam, Netherlands; 2) Bioinformatics Erasmus Med Ctr, 3000CA Rotterdam, Netherlands; 3) Biochemistry Erasmus Med Ctr, 3000 CA Rotterdam, Netherlands; 4) Biochemistry and Molecular Biology, University of Texas Medical School, TX 77030 Houston, USA.

The Integrator Complex is responsible for the 3'-end processing of the spliceosomal U snRNAs. Maturation of these snRNAs is essential for the proper function of the major/minor spliceosome. Although there is accumulating evidence that the complex is necessary for normal vertebrate development, no association with human disease has been reported yet. Three siblings presented with severe intellectual disability, cerebellar hypoplasia and periventricular nodular heterotopia (PNH) of unknown cause. Using whole genome sequencing (Complete Genomics), and filtering for recessive inheritance, we discovered co-segregating compound heterozygous mutations in Integrator Complex Subunit 8 (INTS8) in the three sibs. Quantitative RT-PCR showed 2-fold decreased INTS8 RNA levels and lower levels of the catalytic subunits INTS4, INTS9 and INTS11 were observed in patient fibroblasts by western blot, indicating that the integrity of the complex is disrupted. Quantitative RT-PCR analysis uncovered significantly misprocessed levels of U1, U2 and U4. Pathway analysis (DAVID and IPA) of expression data (Affymetrix U133 expression arrays and Gene Chip 1.0 ST exon arrays) of RNA extracted from patient fibroblasts showed abnormal expression of genes required for mRNA splicing and posttranscriptional modification. Expression data of (developing) human brain structures (Allen Brain Atlas) show that expression of INTS8 peaks prenatally, especially in the ganglionic eminences, (sub)ventricular zone and hindbrain, similar to the developmental expression in mouse embryo (Emage database). Interestingly, these areas, from which neuronal precursors migrate to the cortex, are compatible with the regions affected by PNH and cerebellar hypoplasia in the patients. We propose that dysfunction of the Integrator Complex, possibly through snRNA misprocessing and spliceosomal defects, leads to severely disrupted brain development in humans.

624T

Predicting the impact of non-coding genetic variants on complex traits. *F. Luca, G. Moyerbrailean, R. Pique-Regi.* Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI.

Genome-wide association studies (GWAS) have been instrumental to identify a large number of genetic variants associated with complex traits. However, risk variants tend to fall in non-coding regions and likely affect gene regulatory mechanisms that are not yet well defined. Projects such as ENCODE have generated comprehensive catalogs of tissue-specific regulatory regions, but it is not clear how these are affected by a sequence change. Identification and functional characterization of variants in regulatory elements is a crucial step in understanding phenotypic expression in complex traits. To identify genetic variants that can potentially disrupt binding in a specific tissue, we developed a novel extension of the CENTIPEDE model that builds a new sequence motif model. First, we applied the CENTIPEDE approach to identify DNase-seq footprints where access to the DNA is blocked by a bound factor. Then, we calculated tissue-specific binding probabilities and how they change in the presence of a genetic variant. We used this approach across 280 cell-types from the ENCODE project, and then computed allele-specific binding (ASB) probabilities for all known variants (from the 1000Genomes project) overlapping motifs. We annotated >8.7 million sequence variants in footprints. For >5.6 million of these variants we predicted >2-fold difference in binding affinity. Among these predicted functional variants, 1,972 were associated with a complex trait from the NHGRI GWAS catalog. For example, we identified 6 independent variants predicted to disrupt binding of DEAF1, a factor involved in maintaining peripheral immune tolerance (in lymphnodes) and in cell proliferation. Of these variants, one is associated with Kawasaki Disease and is predicted to disrupt binding in lymphoblast and 3 are associated with different types of cancer. We then validated a subset of our predictions using ASB on heterozygous SNPs for six cell-types. 75% of the heterozygous SNPs with significant ASB are concordant with the predictions. These results represent one of the most extensive catalogs of tissue-specific regulatory sequences and genetic variants. We anticipate that this resource will be useful to annotate GWAS loci for many diseases and complex traits by providing new advantages in three key aspects: i) single base pair resolution, ii) tissue-specificity for more than 200 tissues, and importantly iii) a suggested molecular mechanism beyond physical annotation.

625F

Mutation profiling of exonic-enhancers using massively parallel reporter assays. *R.Y. Birnbaum[1], R.P. Patwardhan[2], M.J. Kim[1], G. Finlay[2], J. Zhao[1], R. Bell[3], R.P. Smith[1], A.A. Ku[1], J. Shendure[2], N. Ahituv[1].* 1) Department of Bioengineering and Therapeutic Sciences, Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA; 2) Department of Genome Sciences, University of Washington, Seattle, Washington, USA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94143, USA.

Protein coding exons that also function as enhancers (eExons) have been shown to regulate transcription of the genes they reside in or nearby genes. Mutations in these eExons could lead to multiple phenotypes due to alterations in protein function and/or transcriptional regulation. However, the functional consequences of these mutations are not well known. Here, using ChIP-seq with enhancer marks (p300, H3K27ac, H3K4me1) on human hepatocytes and mouse liver, we show that ~6% of all ChIP-seq peaks overlap coding exons (excluding 1st exon and promoter regions) and demonstrate that 8 of 15 tested sequences function as enhancers in mouse liver and HepG2 cells. Using massively parallel reporter assays, we further dissect in vivo the enhancer activity of three liver eExons (*SORL1*, *PPARG*, *TRAF3IP2*) in a single nucleotide resolution. We find that deleterious enhancer mutations are correlated with the location of transcription factor binding sites and synonymous and non-synonymous mutations have a similar effect on enhancer activity. We also show that the regulatory repertoire that controls the enhancer activity of these eExons is different across cell types leading to differences in deleterious mutation profiles. Combined, these results demonstrate that eExons mutations can disrupt both the protein structure and enhancer activity with differential effect across cell types and can cause multiple phenotypes.

626W

HSA21 Single-minded 2 (Sim2) binding sites co-localize with super-enhancers and pioneer transcription factors in pluripotent mouse ES cells. *A. Letourneau[1], G. Cobellis[2], F. Santoni[1], E. Falconnet[1], A. Vannier[1], M. Guipponi[1], C. Borel[1], S.E. Antonarakis[1,3].* 1) Genetic Medicine & Development, University of Geneva Medical School, Geneva, Switzerland; 2) Department of Biophysics, Biochemistry and General Pathology, Seconda Università di Napoli, Italy; 3) iGE3 Institute of Genetics and Genomics of Geneva, Switzerland.

Down syndrome (DS) results from trisomy of chromosome 21 (HSA21). Some DS phenotypes may be directly or indirectly related to the increased expression of specific HSA21 genes, in particular transcription factors. The HSA21 Single-minded 2 (SIM2) transcription factor has key neurological functions and appears therefore as a good candidate for some DS features, in particular mental retardation. In order to identify DNA binding sites and downstream targets of Sim2, we deep sequenced Sim2-immunoprecipitated DNA from a mouse embryonic stem cell (mESC) line overexpressing a Flag-tagged mouse Sim2 under the control of a Tet-off system. Non-expressing mESCs were used as controls. Reads uniquely mapped with BWA were submitted to HOMER for the identification of Sim2 targets and the discovery of binding site motifs. We identified 1229 potential Sim2 DNA-binding sites, mainly located in intergenic regions (57%). A gene ontology analysis on the nearby genes confirmed the importance of Sim2 in developmental and neuronal processes and revealed a potential role of master transcription regulator. Sim2 gene targets were then validated by RNA-Seq transcriptome comparison of the Sim2 expressing and non-expressing mESCs. The identification of the Sim2 binding motifs revealed high sequence similarities with motifs previously described for pioneer transcription factors such as Oct4, Nanog or Klf4. Correspondingly, a significant fraction of Sim2 binding sites overlaps with loci experimentally known to be occupied in mES cells by one or several of these master transcription factors including Oct4, Nanog, Klf4 or Sox2. In addition, we found that Sim2 binding sites overlap significantly with marks of open chromatin and active enhancer elements such as H3K27ac, P300 or DNaseI hypersensitive sites. More importantly, 67% of the identified Sim2 peaks overlap with regions bound by the Mediator coactivator Med1, suggesting that some of the Sim2 binding sites could even predict a particular sub-category of enhancers known as super-enhancers. All together, we provide evidence that Sim2 binds specific enhancer elements thus explaining how Sim2 can regulate its gene network in DS neuronal features. We hypothesize that those enhancer loci are initially occupied by pioneer transcription factors like Oct4, Sox2 and Nanog known to actively open chromatin enabling the recruitment of Sim2.

627T

Splicing QTL analysis from primary immune cells identifies regulatory effects putatively associated with autism and Alzheimer's disease. *J.M. Replogle[1,2,3,4], T. Raj[1,2,3,4], K.L. Rothamel[5], C. Benoist[2,3,5], B.E. Stranger[4,6,7], P.L. De Jager[1,2,3,4], Immunological Variation Consortium.* 1) Department of Neurology, Brigham & Women's Hospital, Boston, MA; 2) Program in Medical & Population Genetics, The Broad Institute, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA; 5) Department of Microbiology and Immunobiology, Harvard Medical School; 6) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 7) Institute of Genomics and Systems Biology, University of Chicago.

Alternative splicing acts as an abundant source of transcriptional and phenotypic diversity in humans. Recently, advances in sequencing and array technologies have facilitated high-throughput quantification of mRNA expression at the exon level, and previous studies highlight significant tissue- and population-specificity of alternative splicing. Several splicing quantitative trait loci (sQTL) studies have identified putative functional links between variants associated with complex traits and alternative splicing. However, many of these studies have identified sQTLs in immortalized cell lines, and we have little knowledge about how these discoveries will translate to primary cell-types, which may be most relevant for unraveling disease phenotypes. Here we performed an sQTL analysis in two primary immune cell-types, CD4+ T lymphocytes and CD14+CD16- monocytes, obtained from 377 individuals of European ancestry as part of the ImmVar Consortium. To maximize statistical power, we limited the number of tests performed by focusing our analysis on genetic variants with a high probability of influencing splicing: SNPs located in gene splice sites. Using 851 potential splice variants, we found 783 genes containing at least one potential splice variant. To deconvolute variants' effects on overall gene expression with effects on exon abundance, we calculated a normalized intensity (NI), or splicing index, of expression for each exon, dividing each exon's expression by the overall gene expression level. Testing the association between exon NI and splicing SNP genotype, we identified 53 significant sQTLs (27 genes) in CD4+ lymphocytes. In CD14+ monocytes, we failed to detect any significant sQTLs. Although our array-based expression quantification is more susceptible to reporting spurious associations than RNA-seq data, we replicated previously-reported, experimentally-verified sQTLs in *OAS1* and *CAST*. Examining overlaps between our results and Genome-Wide Association Studies, we found that rs1143674, previously associated with autism susceptibility, correlates with alternative splicing of *ITGA4*, and rs3826656 and rs3865444, previously associated with Alzheimer's disease risk, correlate with alternative splicing of *CD33*. In the future, we plan to expand our sQTL analysis to include CD4+ and CD14+ cells from individuals of African American and East Asian ancestry in order to increase power and characterize the sQTL variation between populations.

628F

Transcription in the human brain: DNA methylation in a single nucleosome may regulate an opioid gene as a region-specific `epigenetic switch'. *G. Bakalkin, T. Yakovleva, H. Watanabe, O. Kononenko, R. Henriksson, I. Bazov.* Uppsala University, Uppsala, Sweden.

Regulation of gene transcription in human brain has not been yet systematically addressed. We focus on the opioid genes including prodynorphin (PDYN) that control emotions, processing of nociceptive information, and reward. Dynorphin opioid peptides may induce a number of pathological processes associated with substance addiction and depression, and also when mutated cause profound neurodegeneration in SCA23 subjects. To address transcriptional / epigenetic mechanisms of PDYN regulation, we analysed 1.7 kB PDYN promoter DNA methylation in post-mortem specimens of the human prefrontal (PFC) and motor (MC) cortices (discovery sample: n = 14 subjects / group; replication sample: n = 8 subjects / group) using bisulfite treatment / pyrosequencing. Methylation of each CpG in the promoter short CpG island (CGI), but not that of CpGs located proximally or distally demonstrated strong and negative correlation with PDYN mRNA levels. Strong positive correlations between CpGs in the CGI, but not between these CpGs and those in the proximal and distal domains were also observed. The CPI encompassing approximately 150 bp apparently formed a single nucleosome. Search for transcription factor (TF) binding sites found a canonical E-box in the CGI, and Upstream Regulatory Factor-2 (USF2) was identified as a dominant E-box binding factor in the human brain by EMSA. USF2 was i) colocalized with PDYN protein in neurons using immunostaining; and ii) bound to the PDYN CGI using ChIP-qPCR. PDYN mRNA / peptides, CpG methylation in the CGI, and USF2 correlated with each other. These correlations were impaired and the CGI methylation was decreased in the model human pathology analysed, i.e. human alcoholism (DSM-4; discovery sample: n = 14 subjects / group; replication sample: n = 8 subjects / group), which was characterized by PDYN upregulation. No changes were evident in the MC, which showed no PDYN activation in alcoholics. We hypothesize that the PDYN CGI functions as the brain-area specific 'epigenetic switch' affected by alcohol. The CGI demethylation may promote USF2-mediated recruitment of histone modifiers to the promoter resulting in PDYN activation in alcoholics. These findings also propose a novel transcription function for short CGIs, which may form a single nucleosome and which DNA methylation may be selectively regulated. Support: Swedish FAS, VR, FORMAS. Tissues were received from the NSWTRC, University of Sydney supported by NIH/NIAAA R24AA012725.

629W

Sex chromosomes and sexual dimorphism in human transcriptome. *D. Nguyen, C. Disteche.* Pathology, University of Washington, Seattle, WA.

X-linked recessive mutations that increase male-fitness but are potentially disadvantageous to females are more easily fixed than autosomal mutations because of X hemizygosity in males and random X inactivation in females. Mutations that enhance female-specific fitness are also preferentially selected because the X chromosome spends 2/3 of its time in females. In addition, up to 25% of human X-linked genes appear to escape X-inactivation and thus would have higher expression in females, although only a limited number of tissues, cell lines, and individuals have been assayed. We have measured the sexual dimorphic expression of X-linked, Y-linked, and autosomal genes in a wide range of human tissues based on more than 500 RNA-seq datasets and 8000 expression arrays in 65 tissues from disease-free individuals deposited in public databases. We used global normalization to compare arrays across tissues, and RPKM for RNA-seq datasets. A standard student-t test with step-down Benjamini-Hockberg correction at 5% FDR was applied throughout to assess genes with at least 1.2 fold expression difference between the sexes. We selected tissues with at least 15 individuals per tissue and similar number of female and male individuals. Genes located within the largest human pseudoautosomal region (PAR1) had significantly higher expression in male tissues, possibly due to spreading of silencing in the inactive X in females. In contrast, escape genes outside the PAR had significantly higher, but rarely doubled, expression in females compared to males, in a tissue-specific manner. This female bias was most pronounced for genes known to escape X inactivation but was also seen for other X-linked genes perhaps due to partial escape in specific tissues. Such female-specific bias was particularly pronounced in the gyrus and the cortical regions of the brain, in the skin, in CD4 positive blood cells, and in reproductive organs. These studies led us to identify the first comprehensive sex biased human transcriptome in brain, non-brain, immune cells, and sex organs, which consisted of a female biased gene set containing 279 X-linked and 7229 autosomal genes (21% of the genome) and a male biased gene set containing 95 X-linked and 3015 autosomal genes (9% of the genome).

630T

FREM1 Regulates Genes Important for HIV-1 Replication and Cell Migration. *M. Luo[1,2], J. Sainsbury[1], P. Lacap[1], F. Plummer[1,2].* 1) HIV and Human Genetics, Public Health Agency of Canada, Winnipeg, Manitoba, Canada; 2) Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada.

Introduction: FREM1 has been identified as a novel candidate gene in resistance to HIV-1 infection in the Pumwani sex worker cohort established in Nairobi, Kenya. Several molecular features of FREM1 suggest its potential role in HIV-1 vaginal transmission. It is an essential component for epidermal integrity and in the right path of HIV-1 infection. A splice variant of FREM1 is a co-receptor for IL-1R1 and toll-like receptor involved in enhancing NFKB activation. However, the precise role of FREM1 in HIV-1 infection and how its variants influence resistance and susceptibility to HIV-1 infection need to be investigated. In this study we studies the role of FREM1 in regulating genes important in HIV-1 replication and cell migration, the two important factor in vaginal HIV-1 transmission. **Method:** We knocked down FREM1 expression in 293F cells and over expressed FREM1 in the HeLa cells derived from cervical tissue and examined the effect on the expression of genes in signal transduction pathways important for HIV-1 transmission by real time PCR array (SABioscience). **Results:** The results showed that knocking down or over expression of FREM1 influences the expression of many important genes involved in NF-κB and inflammatory response, apoptosis, epithelial adhesion and cell migration. As immune response is of paramount importance to HIV-1 transmission and replication, regulation of immune system genes by FREM1 may be a major factor in mediating resistance to mucosal acquisition of HIV-1. **Conclusion:** These suggest that FREM1 may play an immunomodulatory role in cellular activation, which is important for HIV replication. FREM1 may also influence target cell recruitment and HIV infection at vaginal mucosa as suggested by its potential role in epithelial integrity and cell migration through its different functional domains.

631F

***In vivo* UAS$_{Gal}$ gene regulation analysis using the a novel approach.** *H. Guillen Ahlers[1], A. Ludwig-Kubinski[1], K. Lazarova[1], A.M. Greene[1], J. Kennedy-Darling[2], M. Levenstein[2], C. Anderson[1], J. Barfknecht[1], R. Knoener[2], M. Scalf[2], Y. Yuan[2], R. Cole1[1], M. Shortreed[2], L. Cirillo[1], L. Smith[2], M. Olivier1[1].* 1) WI CEGS , Medical College of Wisconsin, Milwaukee, WI; 2) WI CEGS, University of Madison, Madison, WI.

The Wisconsin Center for Excellence in Genomics Science is developing a novel technology to identify the proteins that are associated with DNA at any desired region of the genome. Sequence-specific hybridization to formaldehyde-crosslinked chromatin is used to capture the target region along with its interacting proteins. The novel approach builds on the previous GENECAPP technology developed by the present group and does not require restriction enzyme digestion. The new approach utilizes an optimized capture oligonucleotide mixture that greatly increases the efficiency of the process while reducing costs. This technology was used to study the Gal upstream activator sequence (UAS$_{Gal}$) in *S. cerevisiae*. Cells were grown using either glucose or galactose as the carbon source, so that the difference in gene modulation at the UAS$_{Gal}$ region could be observed. Bound proteins were identified by tandem mass spectrometry using an Orbitrap Velos instrument. The development of this technology has shown steady improvements and highlights the great potential offered for an unbiased analysis of DNA-protein interactions. Funded by the Wisconsin CEGS through NIH/NHGRI grant 1P50HG004952.

632W

Functional impact of polymorphic inversions on gene expression in humans. *M. Oliva[1], S. Villatoro[1], L. Pantano[1], C. Aguado[1], D. Vicente-Salvador[1], D. Izquierdo[1], M. Puig[1], T. Marques-Bonet[2,4], JI. Lucas Lledó[1], R. Castelo[3], M. Cáceres[1,4].* 1) Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; 2) Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Barcelona 08003, Spain; 3) Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; 4) Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

Despite the significant advances made over the last few years, our understanding of the prevalence and functional impact of inversions in the human genome is scarce. Moreover, little is known about inversions that associate to gene expression changes in humans. The only exceptions are a few studied examples, such as the association of the 17q21.31 inversion with decreased MAPT expression or the 8p23.1 inversion with decreased PPP1R3B expression. Here, we analyzed the overlap of a set of polymorphic inversions in the human genome with genes and the correlation between inversion genotypes and gene expression profiles in a genome-wide fashion, looking for both *cis* and *trans* effects. First, we recorded the gene content of 48 human inversions (identified by our group) and generated 1000 permuted sets to build an empirical null distribution of gene counts for each inversion. We confirmed that inversions break fewer genes than expected by chance ($p<0.05$). Next, we looked for associations of 37 validated inversions with gene expression changes by analyzing microarray expression datasets of lymphoblastoid cell lines of 90 Yoruba, 90 European, 45 Chinese and 45 Japanese HapMap individuals. For each gene expression profile, we searched for inversion eQTLs using linear models in which we adjusted for confounding factors using surrogate variable analysis (SVA) and tried to quantify the effect of each inversion. Possible influencing variables (gender, population) were also taken into account in the model. We further analyzed the two well-characterized inversions (17q21.31, 8p23.1) to evaluate the reliability of our methodology in detecting known effects of inversions on the expression of particular genes. Our results show that 30% (11/37) of studied inversions associate to the expression of at least one gene in *cis* (FDR<10%) and a subset of them also seem to affect expression of genes in *trans*. Additional analyses of expression datasets derived from alternative technologies (RNA-Seq) were performed to assess the robustness of our findings. Insight gained in this study could contribute to a better understanding of the role of polymorphic inversions in the regulation of gene expression across the entire human genome. Support: European Research Council (ERC) Starting Grant (INVFEST) under the European Union Seventh Research Framework Programme (FP7).

633T

Targeted sequencing of promoter-associated tandem repeats identifies common functional effects on gene expression levels in the human genome. *A.J. Sharp[1], A. Guilmatre[1], P. Garg[1], G. Highnam[2], D. Mittelman[2,3].* 1) Genetics & Genomics Sci, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA, USA; 3) Department of Biological Sciences, Virginia Tech, Blacksburg, VA, USA.

Tandem repeats (TRs) are stretches of DNA comprised of two or more contiguous copies of a motif arranged in a head-to-tail pattern, and comprise ~2% of the human genome. They are characterized by high mutation rates and account for >25% of indel variants, therefore representing an important source of genetic variation. However, due to technical difficulties in studying them, TRs remain poorly studied and have often been considered as mere 'junk DNA'. We hypothesized that many TR variations might operate as expression quantitative trait loci (eQTLs). To investigate this hypothesis we have conducted a cis-association analysis of variation in promoter-associated TRs with neighboring gene expression levels. Utilizing a custom solution-based capture approach (PMID: 23696428) we studied 120 CEU and YRI HapMap individuals, obtaining a median coverage of 47x informative reads for >4,000 TRs located within 1kb of RefSeq gene transcription start sites. TR genotypes were called using the RepeatSeq algorithm (PMID: 23090981). We then performed an eQTL association analyses using published RNAseq data, identifying 432 TRs in the CEU and/or YRI population that show significant association ($p<0.05$) with the expression level of adjacent genes. This included many TR loci that have not previously been reported as polymorphic. TRs scored as significant eQTLs were enriched for overlaps with transcription factor binding sites and putative enhancers, providing a strong biological rationale for their effects. After phasing of SNP and TR genotypes using BEAGLE, we analyzed patterns of linkage disequilibrium between TRs and nearby SNPs, and observed that most TR variants are poorly tagged by SNP markers. Only 8% of TRs had $R^2 \geq 0.8$ with any SNP within 250kb, and the majority had $R^2<0.3$ with the best tagging SNP. Thus although many TR variants show clear evidence of functional effects, this indicates that the majority are not effectively assayed by SNP-based GWAS approaches, potentially explaining some of the 'missing heritability' of the genome. Our eQTL study represents the first systematic attempt to assign biological significance to TR variations in the human genome, and suggests that potentially there are many thousands of TR variations in the genome that exert functional effects via alterations of local gene expression or epigenetics. We conclude that specific studies that focus on genotyping TR variants are required to fully ascertain functional variation in the genome.

634F

Targeted CD4+ Effector Memory T-Cell Gene Expression Profiling Identifies State-Specific cis-eQTLs Among Rheumatoid Arthritis and Celiac Disease Risk Variants. *X. Hu[1,2,3,4,5], H. Kim[1], C. Baecher-Allan[4,6], T. Raj[3,6], P. Brennan[1,4], P. De Jager[3,6], M. Brenner[1,4], S. Raychaudhuri[1,2,3,4,5,7,8].* 1) Division of Rheumatology, Immunology and Allergy, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; 3) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) Harvard Medical School, Boston, MA USA; 5) Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA USA; 6) Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA; 7) Partners Center for Personalized Genetic Medicine, Boston, MA, USA; 8) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

Motivation: We recently demonstrated that common single-nucleotide polymorphism (SNP) variants associated with rheumatoid arthritis (RA) and celiac disease (CeD) both implicate genes specifically expressed in CD4+ effector memory T (TEM) cells. We hypothesize that some variants may be expression quantitative trait loci (eQTL) that contribute to the development of disease by regulating gene transcription in a cell-specific manner. Methods: We performed genome-wide SNP analysis in 174 healthy, non-Hispanic, Caucasian volunteers. In conjunction, we isolated highly purified CD4+ TEM cells (CD45RA-, CD45ROhigh, CD62Llow/-) from the peripheral blood of the same individuals, and assayed the expression of 215 genes within RA and CeD loci along with 15 control genes for calibration using Nanostring nCounterTM, before and after T cell receptor (TCR) stimulation by anti-CD3/anti-CD28 antibodies. We assessed the presence of cis-eQTL within 1Mb of the transcription start site changes of each gene. Results: The genes with the largest fold changes in expression following stimulation were GZMB (average fold change: 182.2; range: 6.5-573.4) and IL2RA (average fold change: 105.9; range: 8.7-711.0). The genes with the greatest fold change were also those most specifically expressed in CD4+ TEM cells (Spearman rho = 0.26, $P = 5.4\times10^{-4}$). Six of the 35 RA variants in high-density genotyped regions, and four of the 50 CeD variants were in tight linkage disequilibrium with cis-eQTLs (permutation-based $P < 0.01$, $R2 > 0.7$). One disease-associated eQTL (rs1980422/CD28, shared between RA and CeD) influenced CD28 expression specifically in cells before stimulation; rs12936049 and rs4840565 were associated with differential expression of GSDMB and BLK, respectively, both before and after stimulation. Remaining disease-associated eQTLs were associated with differential expression only after stimulation. None of the eQTL SNPs were significantly associated with the proportion of CD4+ TEM cells or with cell proliferation. Conclusion: Most RA- and CeD-associated eQTLs are specific to either resting or stimulated CD4+ TEM cell state. Genetic variants may contribute to the development of autoimmune diseases by regulating gene transcription in specific cell populations such as CD4+ TEM cells.

635W

Mapping the genetic architecture of gene regulation in whole blood in the KORA study. *K. Schramm[1,2], C. Marzi[3], C. Schurmann[4], M. Carstensen[5], E. Reinmaa[6,7], C. Gieger[8], E. Mihhailov[7,9], R. Mägi[7], A. Peters[3,10,11], K. Strauch[8,12], M. Roden[5,13], T. Illig[3,14], T. Meitinger[1,2,10], A. Metspalu[6,7], C. Herder[5], H. Grallert[3], H. Prokisch[1,2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 2) Institut für Humangenetik, Technische Universität München, München, Germany; 3) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 4) 4Ernst-Moritz-Arndt-University Greifswald, Interfaculty Institute for Genetics and Functional Genomics, Greifswald, Germany; 5) Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany; 6) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 7) Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 9) Estonian Biocenter, Tartu, Estonia; 10) Munich Heart Alliance, München, Germany; 11) Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 12) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, München, Germany; 13) Department of Metabolic Diseases, University Hospital Düsseldorf, Heinrich-Heine University, Düsseldorf, Germany; 14) Medical School Hannover, Hannover Unified Biobank, Hannover, Germany.

Introduction: Analysis of whole genome expression quantitative trait loci (eQTL) provides a means for transcriptional regulatory relationships at a genome-wide scale and thus for identifying regulatory pathways affecting disease susceptibility and other relevant traits **Methods:** We performed eQTL analyses in 890 randomly selected fasting participants from the population-based KORA F4 study and replicated the results in independent non-fasting samples (EGUT, N=842). Linear regression models using additive effects with adjustments for principal components were applied. Genome-wide statistical significance was defined as 0.05/number of tests performed. **Results:** 4,210 eQTLs, (4,116 cis-eQTLs, defined as 500 kb window around the transcription start and end site of a gene, and 94 trans-eQTLs) reached genome-wide significance. Of those, 87% and 80% of the cis- and trans-eQTLs, respectively, were confirmed in the replication samples. (The replication in a second cohort is ongoing and looks similar.) Network analysis for the significant eQTLs using Ingenuity Pathway analysis software identified an enrichment of pathways involved in the development and the activity of the immune system and a central role of the HLA-system. Furthermore, for the set of significant cis-eQTL we observed (1) an overlap of 19% of genes detected in genome-wide analysis studies (GWAs) and recorded in the GWAs catalog so far (http://www.genome.gov/gwastudies, July, 18th, 2012), (2) major cross-tissue similarity (46–70%) with previously published cis-eQTLs found in monocytes, LCLs, lymphocytes, lung tissue, and liver tissue, and (3) five chromosomal regions with simultaneous impact on multiple gene expression levels. Amongst the set of significant trans-eQTLs, a triangular relationship between an eQTL-SNP residing in a gene desert on chromosome *6q24.1*, the gene expression probe of a known type 2 diabetes susceptibility gene (*IGFBP2*), and adiponectin was identified. **Conclusion:** The present study identified numerous eQTLs in whole blood in a large Western European sample and provided evidence that these results offer a valuable resource for investigators studying the genetic architecture of regulatory pathways in whole blood. The high replication rate also in non-fasting subjects demonstrates the robustness of the regulatory effects in whole blood. Furthermore whole blood seems to be an informative tissue for an abundance of transcriptional regulatory relationships also in other tissues.

636T

Paired eQTL analysis of monocytes and differentiated macrophages. *S. Makino[1], V. Naranbhai[1], J. Knight[1], B. Fairfax[1,2].* 1) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 2) Oxford Cancer Centre, Churchill Hospital, Oxford, OX3 7LJ.

Monocytes and macrophages form crucial cellular subsets of the innate immune system. Monocytes are short-lived circulating cells that upon migration into inflamed or damaged tissue differentiate into resident macrophages, promoting either resolution or chronicity of inflammation. Multiple eQTL analyses have determined that a proportion of eQTL exhibit high cellular specificity. The degree to which the activity of genetic determinants of gene expression is retained during cellular differentiation is unclear however. Here we aimed to investigate whether eQTL observed in CD14+ monocytes are maintained after their differentiation into macrophages. Primary CD14+ monocytes were isolated from peripheral blood mononuclear cells of 64 healthy Europeans. They were ex vivo differentiated into macrophages over18 days with Tumour Necrosis Factor (TNF) and Granulocyte-Macrophage Colony Stimulating Factor (GMCSF). Differentiation into a macrophage phenotype was confirmed with quantitative PCR (qPCR) of the monocyte and macrophage markers, CD14, CD68 and CD163. Gene expression was subsequently analyzed with total RNA from monocytes and macrophages using Illumina HumanHT-12v4 BeadChips. Individuals were genotyped at 730,000 markers using Illumina OmniExpress v1 beadchips. eQTL analysis was performed using a linear model incorporating expression principal components as covariates. Here we report the results of this analysis. Major observations include the diminishment of the previously reported master regulatory region at 12q15 at the LYZ-YEATS4 locus upon macrophage differentiation, a result validated with parallel qPCR of LYZ. This analysis supports the plastic nature of eQTL activity which is dependent in part upon the cellular differentiation state. It also provides the first reported analysis of eQTL from differentiated macrophages.

637F

Human Cytomegalovirus pUL23 Protein Potentially Associated with Viral Latency. *H. Li, B. Zeng, J. Li, H. Liu, Y. Ran, Y. Li, T. Zhou, Z. Deng.* College of Life Science and Technology, Jinan University, Guangzhou, Guangdong, China.

Most human individuals are latently infected with human CMV, a prototypic β-herpesvirus, frequently acquired during early childhood. The latent infection can be reactivated in the absence of adequate immune control, and the otherwise asymptomatic infection can be dangerous and play a pathological role in some life threatening diseases. This study is aiming on identifying potentially interesting virus-host interactions responsible for CMV latency. o enable efficient replication and to maintain lifelong latency in immunocompetent hosts, CMVs have evolved numerous molecules mediating immune evasive properties, targeting both innate and adaptive immune responses. One of the striking immune evasive strategies is to interfere with JAK/STAT signal transduction, block IFN-stimulated gene (ISG) expression following viral gene expression, also during an initial ISG activation phase. CMVs encode more than 100 genes that are nonessential for growth in vitro and hence are likely to modulate the virus-host interaction in vivo, including immune modulatory genes. UL23 gene, encoding pUL23 protein, is one member of the human cytomegalovirus (HCMV) US22 gene family. The intracellular target of the pUL23 protein was investigated by using a yeast two-hybrid screening system with pUL23 as bait. Following the two-hybrid screen, a list of interesting interactors was generated and the interactions validated using Western blots using commercial antibodies recognizing the interactors identified. From these results, we are focusing on specific partners associated with STAT1 and IFN receptor. Glutathione S-transferase pull-down experiment revealed some interactors associated with HCMV pUL23 protein with a higher affinity than STAT-1 did. Therefore, it is suggested that pUL23 protein has the ability to interact strongly with host proteins and consequently to bring about the disruption of the complex formed from STAT-1 and the IFN receptor, probably resulting in suppression of the IFN signal transduction pathway.

638W

Resolving regulatory genetic variants in severe sepsis due to community acquired pneumonia by mapping context specific expression quantitative trait loci. *E.E. Davenport[1], J. Radhakrishnan[1], P. Humburg[1], T. Mills[1], P. Hutton[2], C. Garrard[2], C. Hinds[3], J.C. Knight[1], The GAinS Investigators.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Adult Intensive Care Unit, John Radcliffe Hospital, Oxford, United Kingdom; 3) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, London, United Kingdom.

Severe sepsis remains a major area of unmet clinical need with a mortality rate of over 30% despite optimal current antibiotic therapy and intensive care support. New insights into disease pathophysiology and opportunities for early effective intervention and patient stratification are urgently required. We present data for a large cohort of 300 patients of European ancestry with severe sepsis due to community acquired pneumonia. Detailed transcriptomic profiling has been carried out for a total leukocyte cell population rapidly purified at the bedside using the Ambion LeukoLOCK Total RNA Isolation System. We quantified gene expression from serial samples taken following admission to the intensive care unit (ICU) using Illumina HumanHT-12 v4 Expression BeadChip arrays for 47,000 probes. We hypothesised that specific genetic modulators of gene expression may be important in this disease context and we proceeded to compliment detailed transcriptomic profiling by mapping gene expression as a quantitative trait (eQTL). Genotyping was performed for 730,525 SNPs using the Illumina HumanOmniExpress BeadChip. Following quality control, eQTL analysis was carried out for 240 patients using 16,874 probes and 644,390 SNPs. We incorporated principal components analysis (PCA) to define 4010 unique probes associating with local, likely cis-acting, expression associated SNPs (eSNPs) that are within 1Mb of the probe (FDR<0.05). Extensive clinical phenotyping including survival up to six months after ICU admission allows analysis of gene expression and eQTL data in the context of outcome and resolution of endophenotypes. Additionally, we will show how with complex heterogeneous clinical datasets, such as presented here, defining variance using PCA and inclusion of known covariates significantly increases the yield of eQTL identified. We will demonstrate how analysis of context specific eQTL in a disease setting reveals novel eSNPs not currently identified by analysis of lymphoblastoid cell lines or primary cells from healthy volunteers. This novel dataset and approach is informative for genome-wide association studies in sepsis, infectious disease and other immune-related traits.

639T

A comparative transcriptome analysis identifies FGF23-regulated genes in HEK293 cells stably expressing KLOTHO. *S. Diener[1], T. Schwarzmayr[1], A. Schmittfull[1], T. Wieland[1], B. Lorenz-Depiereux[1], T.M. Strom[1,2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Bavaria, Germany; 2) Klinikum Rechts der Isar der Technischen Universität München, Institute of Human Genetics, Munich, Bavaria, Germany.

Phosphate is the most abundant anion in the human body and is crucial for various biological functions like cellular activity and bone mineralization. Phosphate homeostasis is regulated in a complex process that involves the interplay of different organs, tissues and systems. A key regulator of phosphate homeostasis is the fibroblast growth factor 23 (FGF23). It is mainly secreted from osteocytes and osteoblasts, circulates in the blood and binds to a specific receptor complex composed of FGF receptor 1 (FGFR1) and KLOTHO in the kidney. As elevated FGF23 levels are the main cause of hypophosphatemia in monogenic disorders of phosphate homeostasis (XLH (MIM 307800), ADHR (MIM 193100), ARHR1 (MIM 241520) and ARHR2 (MIM 613312)), further studies on the regulation of phosphate metabolism are necessary to identify possible therapeutic targets. FGF23 activates FGFR1/KLOTHO to inhibit renal phosphate reabsorption and to suppress 1,25-dihydroxyvitamin D_3 synthesis. However, little is known about FGF23/FGFR1/KLOTHO signaling and downstream targets of FGF23 contributing to its phosphaturic action. For this purpose, we established an *in vitro* cell system of FGF23-inducible HEK293 cells that stably express KLOTHO (HEK293-KL). To find differentially expressed FGF23-induced transcripts, we performed whole transcriptome analysis. We used the technology of RNA-Seq, which is a massively parallel sequencing approach to allow genome-wide analysis of gene expression profiles at a far higher resolution than is available with microarray-based methods. Genome-wide transcriptional changes in HEK293-KL cells specifically caused by FGF23 were defined by comparing the transcriptome of FGF23-induced HEK293-KL cells with the transcriptome of not induced HEK293-KL cells. We tried to identify new FGF23-responsive genes that might belong to a network of factors involved in the regulation of phosphate homeostasis.

640F

Circulating miRNAs associated with High Altitude Sickness at the Qinghai-Tibetan Plateau. *NE. Buroker[1], X-H. Ning[2], Z-N. Zhou[3], K. Li[4], W-J. Cen[4], X-F. Wu[3], W-Z. Zhu[5], CR. Scott1[1], SH. Chen[1].* 1) Pediatrics, 356320, University of Washington, Seattle, WA; 2) Division of Cardiology, Seattle Children's Hospital. Institute. Foundation, Seattle, WA; 3) Laboratory of Hypoxia Physiology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 4) People's Hospital of the Tibet Autonomous Region, Lhasa, China; 5) Center for Cardiovascular Biology and Regenerative Medicine, University of Washington, Seattle, WA.

Circulating miRNAs isolated from dried blood spots (DBS) were found to be associated with high altitude sickness (HAS) patients in Tibet. HAS arises from two different diseases which are acute (AMS) and chronic (CMS) mountain sickness. Circulating miRNAs differences were found between AMS Han Chinese patients and normal Han controls and between CMS Tibetan Chinese patients and normal Tibetan controls. HAS arises from hypoxia which afflicts some high altitude inhabitants or visitors and not others. The difference results from each individual's genetic makeup where hypoxia related genes have been shown to be a major contributor to these sicknesses. Several fold changes increases (up regulation) were found in the hypoxia associated miRNAs let-7f-5p, miR-9-5p, miR-19a-3p, miR-23a-3p, miR-98-5p, miR-125a-5p, miR-181b-5p, mir-202-3p, miR-372, miR-381-3p, miR-519d, miR-520d-3p, and miR-656 for both HAS groups compared to their controls. Other miRNAs (miR-19a-3p, 302c-3p and 875-3p) were found to be up regulated in one HAS group and down regulated in the other HAS group indicating the genetic differences between the two sickness groups.

641W

Genetic and epigenetic regulation of human LincRNAs gene expression variation. *K. Popadin, M. Gutierrez-Arcelus, E.T. Dermitzakis, S.E. Antonarakis.* Department of Genetic Medicine and Development,, University of Geneva Medical School, Geneva, Switzerland.

Large intergenic non-coding RNAs (lincRNA) are still poorly functionally characterized in spite of the fact that they represent at least one fourth of the number of protein-coding genes in the human genome. To provide transcriptome-wide description of human lincRNAs, we have analyzed the natural variation of lincRNA expression levels by RNA-Seq (10–50 M reads) as well as the genetic (Illumina 2.5M Omni chip) and epigenetic (450K Illumina Infinium HD Methylation Assay) regulation of lincRNAs in the GenCord collection of three cell types (fibroblasts, lymphoblastoid cell lines and T-cells) from 195 unrelated European individuals. We have observed eight hallmarks of lincRNA functionality: (i) a negative correlation between the lincRNA gene size and their level of expression; (ii) high conservation score of ubiquitously expressed lincRNAs; (iii) genomic co-localization of expressed lincRNA genes with protein-coding genes involved in zinc-ion binding; (iv) a higher abundance of cis expression Quantitative Trait Loci (cis-eQTLs) in lincRNAs compared to protein-coding genes and the prevalent localization of these cis-eQTLs very close to Transcription Start Sites (TSS) of lincRNAs; (v) regulation of lincRNA expression by genetic variation independently of regulation of the neighboring protein-coding genes; (vi) an independent transcription when lincRNA and protein-coding genes found near each other; (vii) epigenetic regulatory patterns similar to protein-coding genes: presence of both positive and negative correlations between DNA methylation and gene expression, with negative correlations being closer to the TSS, and similar landscape of passive and active roles of DNA methylation in gene regulation; (viii) an enrichment of frequently expressed lincRNAs in enhancer chromatin marks. Based on these results we rank all human lincRNAs according to the described evidences of functionality in the investigated human cell types. Computational analyses provide strong signatures of independent functions of the lincRNAs in these cell types.

642T

Heritability of Gene Expression Levels in Genome-Wide Analyses. *T. Huan[1], C. Liu[1], R. Joehanes[1,2], X. Zhang[1], M. Larson[3], B. Chen[1], C. Yao[1], A. Johnson[1], P. Munson[2], P. Courchesne[1], C. O'Donnell[1], D. Levy[1].* 1) Division of Intramural Research, National Heart, Lung and Blood Institute; the NHLBI's Framingham Heart Study; 2) Mathematical and Statistical Computing Laboratory, Division of Computational Bioscience, Center for Information Technology, NIH; 3) Department of Mathematics and Statistics, Boston University, Boston, Massachusetts;.

Genome-wide expression quantitative trait locus (eQTL) mapping studies reveal common genetic variants regulating gene expression. In addition to mapping eQTLs, we systematically evaluated the heritability (h^2) of the whole blood transcriptome in Framingham Heart Study (FHS) families, and explored the proportion of the heritability of gene expression explained by *cis*- and *trans*- components, as a means of identifying the role of eQTLs in promoting phenotype differences and disease susceptibility. Peripheral whole blood samples were collected and large-scale transcriptomic microarray measurements were performed on 5,626 FHS participants. The pedigree structure consisted of 704 extended pedigrees (of size ≥2) from two-generations of FHS participants. Heritability estimates of ~18,000 transcripts were obtained by variance-component methodology. Of all transcripts, about 40% displayed heritability $h^2>0$ with $p<0.05$, and 10% displayed $h^2>0.2$ ($p <$ 1.8e−6). We then investigated the proportion of *cis*-/*trans*- eQTLs in each of the heritability categories at h^2 = 0.2-0.29, 0.3-0.39, 0.4-0.49, 0.5-0.59 and ≥0.6. We discovered that transcripts with higher heritability estimates tended to have larger proportion of *cis*-eQTLs. In contrast, there was no apparent trend for the proportions of *trans*-eQTLs in the different heritability categories. In addition, we discovered that single *cis*-eQTLs explain 27-68% of variance in gene expression levels with $h^2>0.2$. However, *trans*-eQTLs only account 1-6% of variance in transcripts with $h^2>0.2$. Interestingly, we observed that the top *cis*-eQTL tended to explain more variance in the respective transcript as the heritability of the transcript increased, but the top *trans*-eQTL tended to explain more variance in the respective transcript when the heritability of the transcript decreased. By cross-linking the eQTLs with GWAS results of multiple metabolic traits, we discovered that a few trait-associated GWAS single nucleotide polymorphisms (SNPs) explained a large proportion of genetic variance in multiple transcripts, although these SNPs only explained a small proportion of phenotypic variance for the metabolic traits. Using body mass index and blood pressure as case studies, we found that several differentially expressed genes were associated with the same GWAS SNPs. These gene signatures explained a larger proportion of variance in the respective traits than did the GWAS SNPs.

643F

Cis and Trans Effects of Human Variations on Gene Expression. *J. Bryois[1,2,3], A. Buil[1,2,3], D.M. Evans[4,5], J.P. Kemp[4,5], S.B. Montgomery[1,7], D.F. Conrad[6], K.M. Ho[5], S. Ring[5], M. Hurles[6], P. Deloukas[6], G.D Smith[4,5], E.T Dermitzakis[1,2,3].* 1) Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva (iGE3), Geneva, Switzerland; 3) Swiss Institude of Bioinformatics (SIB), Geneva, Switzerland; 4) MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, United Kingdom; 5) School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 6) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 7) Departments of Pathology and Genetics, Stanford University, Stanford, California, United States of America.

Gene expression is a heritable cellular phenotype that defines the function of a cell and can lead to diseases in case of misregulation. In order to detect human variations affecting gene expression, we performed genome-wide association analysis of single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) with gene expression measured in 869 lymphoblastoid cell lines in cis and in trans. Using a threshold defined by permutations, we discover that 3925 genes (false discovery rate (FDR)=4%) are affected by an expression quantitative trait locus (eQTL) in cis and 83 genes (FDR<20%) are affected in trans. We found that CNVs are more likely to be eQTLs than SNPs (odd ratio for cis-eQTLs=7.62 (pval<2.2e−16), odd ratio for trans-eQTL=44.1 (pval=1.7e−7)). In addition we found that GWAS SNPs are enriched for cis and trans eQTLs (pval<0.01) and that trans-eQTLs are enriched for cis-eQTLs (pval<0.01). As a variant affecting both a gene in cis and in trans suggests that the cis gene is functionally linked to the trans gene expression, we looked specifically for trans effects of cis-eQTLs. We estimated the proportion of probes affected by each cis-eQTL from the pvalue distribution of their trans effects and discover that many of them have pleiotropic effects. Using a threshold defined by permutations, we found that 51 cis-eQTLs are associated to 151 genes in trans (FDR<11%) with the cis-eQTLs of the transcriptions factors BATF3 and HMX2 affecting the most genes with 54 and 23 genes affected respectively. We then explored if the variation of the level of expression of the cis genes were causally affecting the level of expression of the trans genes using Bayesian networks and a causal inference test. We discover 19 causal relationships between variation in the level of expression of the cis gene and variation of the level of expression of the trans gene. However, most of the trans associations (73.8%) are independent of the cis gene expression, implying that the trans associations are due to other functional variants in linkage with the cis-eQTLs or through other functional elements regulated in cis. This analysis shows that a large sample size allows the discovery of secondary effects of human variations on gene expression that can be used to construct short directed gene regulatory networks.

644W

Common genetic variation within transcription factor binding sites is associated with bipolar disorder. *D.T.W. Chen[1], N. Akula[1], L. Hou[1], L. Jing[1], G. Hawariat[1], S. Detera-Wadleigh[1], X. Jiang[1], BiGS. Consortium[2], F.J. McMahon[1].* 1) Human Genetics Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD; 2) Department of Psychiatry, University of San Diego, La Jolla, CA 92093-0603.

Genome-wide association studies (GWAS) have uncovered a number of loci associated with bipolar disorder (BD), but functional alleles have not been identified. Some GWAS signals may reflect genetic variation in sequences regulating gene expression via binding of transcription factors (TFs). This study tested the hypothesis that signals obtained via GWAS reflect genetic variation in transcription factor binding sites (TFBS). Genetic association data were extracted from a meta-analysis of worldwide GWAS in BD comprising ~14,000 cases/controls. This set of 700,000+ single nucleotide polymorphisms (SNPs) was mapped onto published TFBS identified in human lymphoblastoid cells. GWAS signal enrichment near TFBS was tested with the Kolmogorov-Smirnov (K-S) rank-sum statistic. Biological function of the TFBS was explored by use of the bioinformatics tools INRICH, GREAT, and DAVID. Significant enrichment of GWAS signals (empirical p= 9.1×10-3) was noted among SNPs near TFBS. This enrichment was not attributable to linkage disequilibrium between SNPs, or differing allele frequency spectra among SNPs detected in the GWAS. Genes nearest these SNPs significantly clustered into relatively few functional pathways. This clustering was not due to gene size or variable assignment of SNPs to nearby genes. Eight genes emerged most often (CAMK2G, ERN1, DDR1, MAP3K5, MAP4K5, MARK2, RAF1, BMPR1A); four of which encode proteins in the serine/threonine-protein kinase pathway. These findings suggest that regulation of gene expression by transcription factors plays an important role in the genetic etiology of bipolar disorder.

645T

Effect of Transcription Factor Binding Variation Depends on Genomic Context. *D.A. Cusanovich[1], B. Pavlovic[1,2], J.K. Pritchard[1,2], Y. Gilad[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Howard Hughes Medical Institute, University of Chicago, Chicago, IL.

A major goal of human genetics is to understand the regulatory logic of transcription factors (TFs) binding to DNA, namely TF-DNA interactions that result in functional output. Research in this field to date has largely relied on DNA sequencing technologies (e.g. ChIP-seq, DNase-seq, etc.) to determine transcription factor binding locations and the chromatin modifications associated with such binding. However, factors are often bound throughout the genome and it is unclear to what extent transcription factor binding influences gene expression levels at a given locus. In order to better tease apart the genomic context of functional transcription factor binding, we knocked down 59 different TFs using RNA interference and measured the resulting global gene expression levels in a Yoruba HapMap lymphoblastoid cell line. The number of genes differentially expressed (DE) in each of the knockdown experiments ranged from 39 to nearly 4,000 (FDR = 0.05). We intersected the gene expression data with transcription factor binding data from previous studies to identify functionally relevant binding events. As expected, we found that only a subset of genes whose regulatory regions were bound by a TF were measurably perturbed by the knockdown of the TF. On average, 14.3% of bound genes were differentially expressed in the knockdown experiments. Using annotations of chromatin states across the genome, we determined that genes identified as differentially expressed in a particular factor knockdown were likely to have that factor bound in a nearby enhancer (e.g. adjusted P-value = 1.2×10^{-4} for the *IRF4* knockdown) and were unlikely to have that factor bound in an active promoter (e.g. adjusted P-value = 4.9×10^{-3} for the *RELA* knockdown). These results were consistent across multiple factors and suggest a model whereby factor binding to a genomic location otherwise marked as an active enhancer is likely to contribute to gene expression regulation and binding at a promoter may be robust to changes in factor concentration in the cell. In conclusion, these experiments give us insight into the role of DNA binding factors in determining gene expression levels and knowledge of both their direct and indirect targets.

646F

Localizing ancient causal regulatory variants from global genetic analyses of gene expression. *M.K. DeGorter[1,2], S.B. Montgomery[1,2].* 1) Department of Pathology, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA.

The availability of dense genetic and gene expression data from diverse human populations, when combined with rich epigenomic data, provides new opportunity to localize ancient causal regulatory variants. Specifically, by taking advantage of the divergent haplotype structure of diverse populations, we are able to more precisely map causal variants whose functional mechanism can then be determined. Using dense genetic data for 540 individuals representing seven populations, we identify shared *cis*-expression quantitative trait nucleotides (*cis*-eQTN) between populations. In particular, we combine 2.2 million phased SNPs (Illumina Omni2.5 BeadChip) genotyped by the 1000 Genomes Project and gene expression data obtained from lymphoblastoid cell lines using the Human-6 Expression BeadChip (v2) to discover *cis*-eQTN. These shared *cis*-eQTN are then intersected with data from the Encode project to elucidate their overall enrichment in functionality and their specific mechanistic roles. Furthermore, shared variants are intersected with GWA data to illuminate particular disease-predisposing variants that have escaped purifying selection. This activity provides enhanced resolution of causal regulatory variants and genes that have remained polymorphic since early out-of-Africa migration.

647W

Regulatory Function of *CACNA1C* Schizophrenia-Associated Variants. *N. Eckart[1], R. Wang[2], M. Szymanski-Pierce[1], M. Zeledon[1,2], S. Goswami[2], D. Valle[1], D. Avramopoulos[1,2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD.

Schizophrenia (SZ). is a complex psychiatric disorder affecting approximately 1% of the population with an estimated heritability between 70-90%. Genome wide association studies have identified many potential susceptibility variants for SZ, some of which have also been shown to be associated with Bipolar Disorder (BP). One such variant, rs1006737 in an intron of the CACNA1C gene, has been reported to be associated with both disorders. Two others in the same gene, rs7972947 and rs4765913, have been reported to be associated with SZ or BP. However, little is known about the identity of the causative variants and the mechanisms by which they contribute to pathogenesis. Because these variants are in non-coding regions, we hypothesized that they contribute to disease by disrupting normal regulation of gene expression. To test this hypothesis, we genotyped the three variants with Taqman assays and used RT-qPCR to measure expression of 3 alternative transcripts of CACNA1C in postmortem superior temporal gyrus (STG) brain samples from pathology-free controls. Then, we used generalized linear models to identify correlations between genotype and transcript levels. We found that the minor allele of rs1006737 is associated with decreased abundance of all CACNA1C transcripts in the STG. To identify the functional variant(s) underlying this effect, we made a series of dual luciferase constructs for all SNPs tagged by rs1006737 with an r2 > 0.80 and included approximately 1kb of genomic context in each case. Dual luciferase reporter assays in two cell types showed that 4 of the 12 genomic loci harboring tagged variants had statistically significant allele-specific effects on luciferase expression. Furthermore, electrophoretic mobility shift assays (EMSAs) showed that 4 of the 6 SNPs included in these 4 constructs also have allelic differences in binding protein complexes from nuclear extracts of the same cell lines. From this data, we conclude that rs1006737, the non-coding variant in CACNA1C that is associated with both SZ and BP, marks a regulatory haplotype that differentially affects the expression of multiple CACNA1C transcripts in the STG. The presence of multiple functional variants in the haplotype raises many interesting questions. Further planned studies include identification of the proteins that bind the regulatory elements in an allele specific manner and characterization of any complex chromatin interactions that mediate the observed effects.

648T

Extensive Variation in Chromatin States Across Human Individuals and Populations. *F. Grubert[1], M. Kasowski[1], S. Kyriazopoulou-Panagiotopoulou[2], J. Zaugg[1], A. Kundaje[3], Y. Liu[4], A. Boyle[1], Q. Zhang[1], F. Zakharia[1], D. Spacek[1], J. Li[1], D. Xie[1], L. Steinmetz[1,5], M. Kellis[3], S. Batzoglou[2], M. Snyder[1].* 1) Dept Genetics, Stanford University, Stanford, CA; 2) Department of Computer Science, Stanford University, CA; 3) Department of Computer Science, Massachusetts Institute of Technology, Cambridge MA; 4) Department of Chemistry, Stanford University, Stanford, CA; 5) Genome Biology EMBL Heidelberg, Germany.

The vast majority of disease-associated variants identified by GWAS lie outside protein-coding regions, suggesting that variation in regulatory regions may play a major role in disease predisposition. Here, we study differences in chromatin states using six histone modifications, cohesin, Pol2 and CTCF in lymphoblastoid lines from 19 individuals of diverse ancestry. We find extensive signal variation in regulatory regions as well as switches of chromatin state across individuals, most frequently between active and repressed states. Enhancer activity is particularly diverse among individuals, and is strikingly divergent across populations. Consistently, chromatin marks show strong inheritance in family trios. Overall, our results provide fundamental insights into epigenetic differences of humans and how regulatory elements might evolve within a species.

649F

Identification of enhancer-promoter interactions in the mammalian genome. *Y.-C. Hwang[1], Q. Zheng[2,4], C.-F. Lin[3,5], O. Valladares[3,5], B.D. Gregory[1,2,4], L.-S. Wang[1,3,4,5].* 1) Genomics and Computational Biology Graduate Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Biology, University of Pennsylvania, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Penn Genome Frontiers Institute, University of Pennsylvania, Philadelphia, PA; 5) Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Genome-wide association studies have shown the majority of disease-associated DNA variations are within non-coding DNA regions and gene deserts. One of class of DNA regulatory element that is non-coding in nature is enhancer elements. Because an enhancer element can be distal and orientation-independent from the gene it regulates, probing all possible pairs of DNA-DNA contacts could be extremely laborious. In order to systematically uncover the enhancer-target gene interactions in mammalian cells, we reanalyzed published Hi-C data from human (e.g. LCL, K562, hESC, and IMR90) and mouse cells (mESC and cortex), with read depths spreading from 60M to 1612M. We applied a geometric distribution-based analysis pipeline to identify DNA interacting hotspots within the Hi-C datasets. By further classifying these Hi-C interaction hotspots, we then identified candidate enhancer elements as hotspots overlapping known enhancer-associated histone modifications from ENCODE and interacting with a promoter element. In addition, we implemented a streamlined analysis pipeline to ensure that this information can be quickly and efficiently extracted from all future Hi-C datasets, which uses a set of Hi-C paired-end FASTQ files as input. The pipeline automates the analyses from mapping the reads, identifying hotspots, detecting hotspot interactions, to identifying enhancer-target gene pairs by incorporating also commonly mapping softwares such as BWA, samtools, bedtools. All analysis commands are generated automatically by the pipeline, including job submission, controlled job dependencies and error checking. The pipeline efficiently functions on Oracle Grid Engine allowing jobs to be parallelized. Using our pipeline, we have identified 159,522 to 449,124 Hi-C hotspots and thousands of enhancer-target gene interactions throughout the human and mouse genomes. Additional genomics evidence supports these findings: these identified enhancers are 6- to 17-fold enriched in p300 binding sites and their target promoters are 1.2 to 1.5-fold more likely to be near RNA polymerase II binding sites. We also discovered that the enhancer-regulated genes tend to express in a tissue-specific manner. Using these comprehensive enhancer-target promoter datasets will allow us to identify disease-linked polymorphisms that lie within these regulatory elements, as well as study the evolutionary conservation of enhancer-target gene pairs between human and mouse.

650W

Small introns of firefly luciferases: structural characterization and their insufficient substrates for splicing in CHO cells. *M. Ishii[1,3], R. Kojima[1], S. Fukuda[1], Y. Tanji[1], M. Sakaguchi[1], Y. Sugahara[1], M. Kamaya[2,3], F. Oyama[1,3].* 1) Biotechnology Laboratory, Department of Applied Chemistry, Faculty of Engineering, Kogakuin Univ, Hachioji, Tokyo 192-0015, Tokyo, Japan; 2) Enviroment Analytical Chemistry Laboratory, Department of Environmental and Energy Chemistry, Faculty of Engineering, Kogakuin Univ, Hachioji, Tokyo 192-0015, Tokyo, Japan; 3) The Firefly-Breeding Project at Hachioji Campus, Kogakuin University, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan.

Most eukaryotic genes contain segments of coding sequences (exons) interrupted by noncoding sequences (introns). Introns are removed through splicing, which is nearly universal in eukaryotes. However, the general function and evolutionally importance of introns remains unclear. We cloned and sequenced the genomic and cDNA clones encoding luciferase in the Japanese firefly, Luciola cruciata. The luciferase genes contained six introns, all of which were very short (less than 100 bases long). We also found that mouse acidic mammalian chitinase (AMCase) possessed the small intron (intron 9, 86 bases long). We carried out comparative analyses and subsequent the structural characterization of the small introns. Although the GT and AG dinucleotides at the 5' and 3' splice sites and an A nucleotide required for branch point formation were conserved in the introns of the firefly luciferase gene, they were characterized by a high A+T content (70–80% average) and lacking the pyrimidine tract located between the branch point and 3' splice site, which were observed in the small intron of the AMCase. The luciferase gene (with 6 introns) and cDNAs (without introns) were inserted into mammalian expression vectors and transiently expressed in CHO cells respectively. Western blot analysis and photoluminescence measurement indicated that the luciferase gene didn't express functional luciferase in the CHO cells, whereas cDNA did. RT-PCR analysis also showed that all of firefly introns didn't be spliced from pre-mRNA transcribed the luciferase gene. These results suggest that there are different mechanisms in splicing and splicing factors between insect and mammal.

651T

A comprehensive genomic landscape of NRSF binding in various cell types. *P. Jain, FP. Behn, GM. Cooper, RM. Myers.* HudsonAlpha Institute for Biotechnology, Huntsville, AL.

Neuron Restrictive Silencing Factor (NRSF/REST) is a master regulator of neuronal genes and is dysregulated in various neurodegenerative diseases and cancers. To more comprehensively understand the roles of NRSF, we have used ChIP-seq, RNA-seq, DNA Methylation (Reduced Representation for Bisulphite Sequencing), DNAse I hypersensitivity, and histone modifications assays for genome wide mapping of regulatory profiles of NRSF and its cofactors within multiple cell lines. We also leveraged ENCODE data to systematically compare NRSF binding activity with binding of dozens of other transcription factors. We have identified key components of differential regulation of target genes by NRSF. We find that NRSF repression function is greatly influenced by other co-binding TFs and not restricted to its known cofactors like sin3a and CoREST. The data suggest that co-occupancy of other TFs with NRSF reduces NRSF repression activity, with transcriptional repression usually observed only in the absence of co-factor binding. Additionally, higher levels of DNA methylation at NRSF target genes coincide with lower expression. We also find that NRSF motif conservation is very high within binding sites near repressed genes. Our data suggest that while NRSF binds widely in human cells, only a small subset of target genes, that tend to also be methylated, have highly conserved motifs, and be depleted for binding of other TFs, are actually repressed by NRSF. This conclusion poses interesting questions as to the nature and consequences, or lack thereof, of most NRSF binding activity in human genomes.

652F

Distal co-regulated regions are crucial for human phenotypes. *K.J. Karczewski[1,2], A. Battle[3], D. Knowles[3], M. Snyder[1], S.B. Montgomery[1,4].* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA; 3) Department of Computer Science, Stanford University, Stanford, CA; 4) Department of Pathology, Stanford University School of Medicine, Stanford, CA.

The human genome is a complex and efficiently-packed system, where regions on different chromosomes may be close to each other in conformational space and thus, may be co-regulated. However, accurate detection and functional characterization of these interactions has been difficult: large sample sizes are required to discover trans eQTLs, while long-range interaction data, such as Hi-C, do not provide a functional framework. Characterizing the layout and interplay between these regions will be crucial for understanding protein function, and thus, human phenotypes and disease. Previously, the most common model for long-range associations has involved SNPs affecting transcription factor function, which then modulate transcription in trans, as seen in model organism trans-eQTL studies. However, direct contact may be an additional unexplored component to this type of regulation. Using trans-eQTL and Hi-C data, we observe that reciprocally regulated inter-chromosomal regions are enriched for co-localization in 3-D conformational space. We investigate the properties of these reciprocally regulated regions, compared to known cis and trans eQTLs. Additionally, these regions are enriched for shared functions as well as shared disease associations. Finally, we investigate the role of these regions on allele-specific expression and highlight a sequencing-based validation strategy. Our data suggest that long-range regulation is crucial for understanding the complexity of human phenotypes.

653W

Regulatory motif centric validation, dissection, and construction of transcriptional enhancers. *P. Kheradpour[1,2], J. Ernst[1,2,3], A. Melnikov[2], P. Rogov[2], L. Wang[2], X. Zhang[2], J. Alston[2], T.S. Mikkelsen[2], M. Kellis[1,2].* 1) MIT Computer Science and Artificial Intelligence Lab, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) UCLA Department of Biological Chemistry, Los Angeles, CA.

Large scale studies have associated thousands of variants with phenotypes. However, due to genomic linkage structure, only a fraction of these variants likely play a causative role. While variants that may affect protein coding sequences can be computationally predicted, identifying variants with regulatory effect is much more difficult. To assess the feasibility of introducing variants and measuring the regulatory , we predicted causal regulators and employed a large-scale experimental approach to validate their instances. Using motif enrichments and depletions in ENCODE candidate enhancer regions, we predicted five activators (HNF1, HNF4, FOXA, GATA, NFE2L2) and two repressors (GFI1, ZFP161) to be active in K562 and HepG2 cell lines. We synthesized and performed a massively parallel reporter assay (MPRA) on 2,104 wild-type sequences centered on motif instances for these factors and an additional 3,314 engineered enhancer variants containing targeted motif disruptions.

We find robust evidence that activator motif disruption, even with just 1-bp changes, abolishes enhancer function, while silent or motif-improving changes do not. We also find that evolutionary conservation, nucleosome exclusion, motifs for other factors, and strength of the motif match predict sequences whose manipulation would result in less enhancer activity. Moreover, we found that scrambling repressor motifs leads to aberrant reporter expression.

We then swapped the motif with variable surrounding context for the ten most expressed sequences for each of four of the tested activators. We found that the context plays a significant cell type specific role. We also examined the role of regulatory motifs in promoters by testing evolutionarily conserved instances for 70 factors. We found that these instances appears to be considerably less cell type specific, suggesting they may be easier to experimentally dissect because identifying a relevant cell type may be less vital.

654T

Discovery and analysis of over 50,000 common functional regulatory variants in the human genome. *M.T. Maurano, E. Haugen, E. Rynes, R. Humbert, J.A. Stamatoyannopoulos.* Department of Genome Sciences, University of Washington, Seattle, WA.

Common disease- and trait-associated variants preferentially localize in regulatory DNA marked by DNaseI hypersensitive sites. Here we perform genomic footprinting across 81 cell types, and show that high density, allelically resolved in vivo DNaseI cleavage mapping can systematically delineate functional regulatory variants on a large scale. We identified 348,521 heterozygous variants in regulatory regions, of which 50,936 mark significantly altered chromatin accessibility in vivo. Of these, 15.5% lie in promoters, and the remaining majority are in distal enhancer regions. Functional variants systematically perturb the recognition sequences of hundreds of individual transcription factors. We show that these variants can be used to functionally map the in vivo protein-DNA interface, providing a new and powerful code for accurately interpreting the functional significance of noncoding regulatory variation.

655F

The Impact of Post-Mortem Sample Collection on Gene Expression in the Genotype-Tissue Expression Project. *T.R. Young, D.S. DeLuca, G. Getz, K. Ardlie, The GTEx Consortium.* The Broad Institute of MIT and Harvard, GTEx Project, Cambridge, MA.

The Genotype-Tissue Expression project (GTEx) is creating a vast public resource for the study of gene expression in up to 30 human tissues using tissue samples collected from deceased donors. Although GTEx samples are collected within 24 hours of death, RNA degradation throughout the post-mortem interval (PMI) is a key concern of the project. While some studies report that post-mortem tissue can yield high quality RNA and retain biologically relevant gene expression profiles, others report alterations in the expression profiles of some genes measured in post-mortem vs. premortem tissue. Using publicly available gene datasets from the GTEx Pilot Study and the Gene Expression Omnibus (GEO), we have attempted to characterize the impact of PMI on the gene expression profiles of GTEx samples. In order to assess how well the GTEx samples represent the gene expression patterns of samples collected from living donors we performed a metagene projection as described by Tamayo and colleagues. Here, we first train a support vector machine (SVM) classifier on the GTEx expression data and then use the SVM to predict the tissue type of 532 GEO samples from 23 GEO datasets. We examined the changes in gene expression patterns that accrue during the post-mortem interval via differential gene expression analyses on two sets of GTEx samples: pre-mortem vs. postmortem blood and low-PMI vs. high-PMI muscle. Finally, we investigated the suitability of GTEx data for expression Quantitative Trait Loci (eQTL) analysis by attempting to replicate known tissue-specific eQTLs within the GTEx Pilot Study. The metagene projection was highly accurate in classifying GEO samples indicating that the biologically relevant expression patterns are retained. In contrast, differential expression analyses identified several thousand differentially expressed (DE) genes and subsequent GO enrichment revealed DE gene sets unique to each tissue type. After correcting for the first principal component of the gene expression data, which is highly correlated with PMI, we were able to replicate the majority of known eQTLs. These results indicate that although post-mortem collection does alter the gene expression profiles of GTEx samples, the samples remain highly representative of living tissue.

656W

Human Transcriptome Landscape Characterized by Deep RNA Sequencing in 957 Individuals. *x. zhu[1], S. Mostafavi[2], A. Battle[2], K. Beckman[3], C. Haudenschild[4], C. McCormick[4], D. Koller[5], AE. Urban[1,5], DF. Levinson[1], SB. Montgomery[5,6].* 1) Psychiatry, Stanford University, Palo Alto, CA; 2) Computer Science Department, Stanford University, Palo Alto, CA; 3) Biomedical Genomics Center, University of Minnesota, Minneapolis, MN; 4) DNA sequencing services, DNA Sequencing Business, Illumina, Inc. Hayward, CA; 5) Deparment of Genetics, Stanford University, Palo Alto, CA; 6) Department of Pathology, Stanford University, Palo Alto, CA.

Various RNA species represent the direct interpretations of genetic information in a cell: they convey the message from DNA to proteins, as well as perform regulatory roles such as catalyzing biological reactions, controlling gene expression, or participating signal transduction. The recent ENCODE project reported a complete catalogue of transcribed regions based on 15 human cell lines, including polyadenylated (poly-A) and non-poly-A RNA, long and short fragments within various cellular compartments. While being highly dynamic in terms of expression levels, cellular localization and isoform compositions, around ~75% of the human genome can be transcribed, which is much higher than the previous estimate. Despite the recent progress in deciphering novel transcription activities, however, the functional levels of these new transcripts remains unclear, due to limited number of test subjects and extremely low expression levels. Our group expanded the study into a much larger cohort of 957 individuals with European-ancestry from two primary tissues: 922 fresh whole blood and 35 brain specimens. All subjects were genotyped and their polyadenylated RNA molecules were studied with high depth sequencing (~ 70 million reads). New algorithms were developed to account for the stochastic noises in the transcription machinery and the sequencing experimental procedures. We then carried out a systematic characterization for a number of novel transcription events: 1) candidate novel exons as intronic elements expressed and spliced to nearby exons 2) alternative splicing 3) alternative 5' untranslated region (5' UTR) and upstream open reading frames (uORF) 4) alternative 3' untranslated region (3' UTR) and polyadenylation sites 5) Large intergenic non-coding RNAs (lincRNA) and 6) viral RNA expression. In addition to compile these findings to a new catalogue, we also studied the expression quantitative trait loci (eQTLs) that are associated with the newly defined transcription events. Together with a number of other features such as evolutionary conservation scores, we were able to estimate the functional levels of the new transcripts. The recent population genetic studies have discovered a great number of genetic variations that may contribute to disease phenotypes. Our study focused on an extensive characterization on the intermediate message layer, and this will help us to further understand the connections between genotypes and phenotypes.

657T

Chromatin enhancer maps in early differentiating T helper cells reveal regulatory SNPs associated with autoimmune diseases. *D. Hawkins[1], A. Larjo[2], S. Tripathi[3], U. Wagner[4], Y. Luu[4], S. Raghav[3], L. Lee[4], S. Raghav[3], R. Lund[3], B. Ren[4], H. Lahdesmaki[2], R. Lahesmaa[3].* 1) Medicine and Genome Sciences, University of Washington, Seattle, WA; 2) Department of Information and Computer Science, Aalto University, Aalto, FI; 3) Turku Centre for Biotechnology, Turku, FI; 4) Ludwig Institute for Cancer Research, Department of Molecular and Cellular Biology, UCSD, San Diego, CA.

Naïve CD4+ T cells can differentiate into specific helper and regulatory T cell lineages in order to combat infection and disease. The correct response to cytokines and a controlled balance of these populations is critical for the immune system and the avoidance of autoimmune disorders. To investigate how early cell fate commitment is regulated, we generated the first human genome-wide maps of histone modifications that reveal enhancer elements after 72 hrs of in vitro polarization toward T helper-1 (Th1) and T helper-2 (Th2) cell lineages. Our analysis indicated that at this very early time point, cell-specific gene regulation and enhancers are at work directing lineage commitment, and likely opposed a role in cell maintenance as in fully differentiated cells. We determined a unique set of enhancer elements relative to recent ENCODE data in fully differentiated cells. Nucleosome-free regions were determined within enhancer chromatin structures. Examination of enhancers at nucleosome resolution identified transcription factor binding site (TFBS) motifs for expressed TFs with known and unknown T cell roles as putative drivers of lineage-specific gene expression.

An integrative analysis of immunopathogenic associated single nucleotide polymorphisms (SNPs) suggests a role for distal regulatory elements in disease etiology. We found SNPs overlapping enhancers associated with asthma, Crohn's disease, multiple sclerosis (MS), psoriasis, rheumatoid arthritis (RA), type 1 diabetes (T1D) and ulcerative colitis. Guided by the motif analysis, we used DNA Affinity Precipitation Assays (DAPA) to determine if autoimmune disease-associated SNPs overlapping TFBS motifs could disrupt TF binding. We successfully found that individual SNPs associated with rheumatoid arthritis, type 1 diabetes and ulcerative colitis are enough to disrupt TF binding at enhancer sequences.

Lastly, we are using computational approaches to predict target genes of enhancers that overlap autoimmune disease-associated SNPs. We are also mapping histone modifications to find novel or alternative promoters that may also overlap associated SNPs. Collectively, our approach sheds new light on the functionality of non-genic disease-associated SNPs and begins to provide novel insight on the etiology autoimmune diseases.

658F

Analysis of allele specific expression in mouse liver by RNA-Seq: marked differences compared to cis-eQTL identified using genetic linkage. *P.-F. Roux[1, 2, 3], S. Lagarrigue[1, 2, 3], L. Martin[4], F. Hormozdiari[5, 6], A. van Nas[6], O. Demeure[1, 2, 3], A. Ghazalpour[6], E. Eskin[5, 6], A.J. Lusis[4, 6, 7].* 1) INRA, UMR1348 Pegase, Rennes, France; 2) Agrocampus Ouest, UMR1348 Pegase, Rennes, France; 3) Université Européenne de Bretagne, France; 4) Department of Medicine/Division of Cardiology, University of California, Los Angeles, United States of America; 5) Department of Computer Sciences, University of California, Los Angeles, United States of America; 6) Department of Human Genetics, University of California, Los Angeles, United States of America; 7) Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, United States of America.

We report an analysis of allele specific expression [ASE] and parent-of-origin expression in adult mouse liver using next generation sequencing (RNA-Seq) of reciprocal crosses of heterozygous F1 mice from the parental strains C57BL/6J and DBA/2J. The genes exhibiting ASE differed markedly from the putative cis-acting expression quantitative trait loci (cis-eQTL) identified in an intercross between the same strains. While about 60% of the ASE, mapped by RNA-Seq, were found in the eQTL gene set, only a small fraction of the eQTL, mapped by linkage analysis, were found in the ASE gene set. We discuss the various biological and technical factors that contribute to these differences, in particular strengths of the two approaches in making a distinction between local and cis eQTL.

We also identify genes exhibiting parental imprinting and complex expression patterns. Our study demonstrates the importance of biological replicates, which is not currently the norm, to limit the number of false positives with such RNA-Seq data.

659W

A large-scale transcriptome study in the Sardinian population. *M. Pala[1,2,6], M. Marongiu[2], Z. Zappala[1], A. Mulas[2], R. Cusano[2], F. Crobu[2], F. Reinier[3], R. Berutti[3,6], M.G. Piras[2], C. Jones[3], D. Schlessinger[5], G. Abecasis[4], A. Angius[2], S. Sanna[2], F. Cucca[2,6,7], S. Montgomery[1,7].* 1) Pathology and Genetics Dept, Stanford University, Stanford, CA; 2) Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato, 09042, Italy; 3) CRS4, Advanced Genomic Computing Technology, Pula, Italy; 4) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 5) Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA; 6) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 7) co-senior authors.

The interpretation of genome-wide association studies (GWAS) is very challenging since most candidate loci fall in non-coding regions and are difficult to interpret. Given the important role of regulatory variation in phenotypic complexity, expression quantitative trait loci (eQTLs) have been proposed as an informative intermediate phenotype between genetic variation and human disease. To study the role of transcriptional mechanisms on regulating eQTL targets, we sequenced the polyA RNA fraction of peripheral blood mononuclear cells (PBMCs) isolated from 624 related individuals (259 families). These individuals had been subject to whole genome sequencing and characterized for more than 800 quantitative traits (including > 250 immune cell traits) by the SardiNIA project. To assess the impact of eQTLs on lincRNAs, we also sequenced the RNA of a subset of 80 individuals after ribosomal RNA depletion. Our preliminary results show that the expression levels of 10,389 genes are associated with a genetic variant (top SNP with p-value $\leq 10^{-8}$ uncorrected), 40 of which are present in the GWAS catalog. We also assessed splicing events and identified 27,904 donor splice sites that are associated with a genetic variant (top SNP with p-value $\leq 10^{-8}$ uncorrected), of which only 4 are present in the GWAS catalog. Using a subset of 68 individuals, we identified 9,427 heterozygous sites which show allele-specific expression (p-value $\leq$ 0.05) 39 of which show high allelic imbalance (a mean of 0.80 for the most expressed allele) and low variability between individuals. These sites are enriched for missense variants and located near genes implicated in disease. In addition, we developed a statistical method to identify genes that are specifically and significantly over or under expressed within single families compared to the majority of the population. We identified 39 genes that have family specific expression (p-value $\leq$ 0.01; FDR 0.462). Interestingly, these genes are also enriched for rare variants within 70kb upstream of their transcriptional start site, and we hypothesize that they may have a role in rare diseases. We are now extending these analyses to the entire cohort and will present these analyses. Our goal is to identify associations between genotype and gene expression in order to enhance our overall understanding of gene regulation and correlate these molecular events with clinical data in order to characterize their role in phenotypic complexity and disease.

660T

ChIP-Enrich analysis of ENCODE ChIP-seq data reveals biological and technical insights into the regulation of pathways. *M.A. Sartor[1,2], R.G. Cavalcante[1], C. Lee[1], N.A. Comment[1], S. Ramdas[1,4], N.M. Dehaan[3], R.P. Welch[2], S. Patil[1], L.J. Scott[2].* 1) Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI., USA; 2) Biostatistics, University of Michigan, Ann Arbor, MI., USA; 3) Environmental Health Sciences, University of Michigan, Ann Arbor, MI., USA; 4) Human Genetics, University of Michigan, Ann Arbor, MI., USA.

A given transcription-related factor (TRF) can bind and influence similar or disparate sets of genes in different cell types. Furthermore, a TRF may regulate the same pathway in multiple cell types by regulating different genes with similar functions. Regulation of similar pathways via different genes may not be detected through assessment of overlapping peaks or target genes. An alternative method to assess differences in TRF binding patterns across cell types is to use gene set enrichment testing to detect biologically relevant pathways or processes. Using ChIP-Enrich, which empirically adjusts for potential confounding by the locus length (the length of the gene body and its surrounding non-coding sequence), we examined ChIP-seq data from four Tier 1 or 2 cell lines from ENCODE (K562 (leukemia), GM12878 (lymphoblastoid), H1-hESC (embryonic stem cells), and HeLaS3 (epithelial carcinoma)), and eight TRFs (GABP, JunD, MAX, MXI1, cMyc, NRSF, POL2, and TBP). We compared the p-values for gene set enrichment across cell types for each TRF by assessing the strength of the correlation between -log10 p-values from each TF between each pair of cell types, and identifying common and cell type-specific pathway regulatory activity. We found similar patterns of gene set enrichment for most TRFs, with POL2 having the highest concordance (Pearson's r = 0.78 - 0.95) among the four cell types. In contrast, H1-hESC had the most tissue-specific enrichment, most strikingly for NRSF (Pearson's r = 0.058 - 0.20) with the 3 other cell types. And overall, NRSF had the lowest correlation across cells types (r= 0.058 - 0.77). The two carcinoma cell lines (K562 and HeLaS3) showed greatest correlation (r= 0.82 - 0.95) for 5 of the 8 TRFs (GABP, MAX, MXI1, POL2, and TBP) even though these two cell lines are derived from different types of cancer. We are investigating the sensitivity of these findings to differences in number of peaks and peak calling algorithms as we have seen in other ENCODE data that use of MACS tends to result in more powerful enrichment testing results than use of spp, and that including more peaks (using a less stringent cutoff) generally results in more powerful enrichment results. In summary, we find that cell type-specificity of regulation can vary widely by TF but that similar cells have more similar patterns of enrichment. The use of gene set enrichment testing can shed light on the tissue-specificity of regulatory networks of TRFs.

661F

Establishing an eQTL map of the Japanese population. *M. Narahara, F. Matsuda, K. Higasa, Y. Tabara, T. Kawaguchi, R. Yamada.* Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Variation in gene expression is one of the major factors that lead to phenotypic variation and disease susceptibility. Many loci identified by genome-wide association studies (GWAS) are located in non-coding regions, which suggests the importance of transcript differences regulated by genetic variations on phenotypic variations. Expression quantitative trait loci (eQTL) mapping is an approach to locate genetic loci that regulate transcription. Genome-wide eQTL map can significantly improve our understanding of local and distal genetic transcriptional regulation, and improve interpretability of results of GWAS. Establishing ethnic-specific eQTL map is of great necessity because ethnic specificity has been demonstrated in several studies. In this study, we report genome-wide eQTL mapping in the Japanese population.

After a standard quality control, we obtained genotypes of 1,425,832 SNPs and 34,872 gene expression phenotypes from peripheral blood cells of 298 unrelated subjects recruited from the Japanese population. We carried out a genome-wide eQTL mapping for all the phenotypes. We identified 207,462 local eQTLs (located within 500kb of a gene) that are significantly associated by FDR < 5%. The genotypes of the local eQTLs explained more than 10% of variance for 2,099 probes. We observed that magnitudes of association were related to distance from eQTL to gene, and eQTLs located at upstream of gene tend to have stronger effect than those at downstream. Distribution of statistics obtained from distal eQTLs suggested only a small fraction of distal eQTLs have regulatory effects. We identified 9,946 distal eQTLs that are significantly associated by Bonferroni-corrected P < 5%, all of which explains more than 15% of phenotypic variance. 55% of the distal eQTLs were located on the same chromosome as their target genes, and thus, intra-chromosomal effects were enriched. Both local and distal regulations were identified for 50 probes.

This study provides a genome-wide eQTL map of the Japanese population. Our analysis indicates that, although local eQTLs are detected more abundantly than distal eQTLs, many distal eQTLs have strong effects. It is suggested that eQTL studies should focus on both local and distal regulation.

662W

Genetic, evolutionary, and structural properties of protein fragility. *D. Vuzman, C. Cassa, D. Jordan, S. Sunyaev.* Brigham and Women's Hospital and Harvard Medical School, Bosto, MA.

Whole genome and exome sequencing data reveal that disease-associated genes harbor numerous completely benign missense mutations alongside pathogenic variants. Moreover, variants in some genes, such as LDLR, are almost exclusively pathogenic, while other genes, such as Dicer1, contain mostly benign variants. Factors determining robustness of individual genes to sequence variations are unclear. We suggest that the robustness of genes is based on intrinsic structural properties of their encoded proteins, which we name 'fragility'. We have constructed 'fragility score', which estimates the chance for a random amino-acid change to critically impact protein function, using DNA sequencing data on thousands of individuals from the general population from the Exome Sequencing Project (ESP) dataset and data on pathogenic alleles reported in the human gene mutation database (HGMD). High fragility score was assigned to proteins with significantly fewer missense mutations in the general population than in disease-associated populations, since most substitutions are deleterious. Using statistical and computational approaches, we estimated important evolutionary and functional predictors of pathogenicity, such as selective constraints and conservation of sequence across species. We analyzed hundreds of protein tertiary structures from the protein data bank to determine structural and thermodynamic determinants of protein fragility. Our results demonstrate that protein fragility derived from sequencing data is correlated to evolutionary conservation scores, to folding free energy change due to a missense mutation, and to the extent of intrinsically disordered regions within proteins. Characterization of fragility elucidates the biophysical constraints of protein structure and can inform various aspects of research on human genetic variation, ranging from basic population and evolutionary genetics, to genetic of complex traits, and clinical genetic diagnostics.

663T

Whole Transcriptome Array to Identify Alternative Splicing Signals Controlled by SNPs in Human Brain. *H. Zhou, L. Chen, A. Kutlu, C. Molony, D. Stone.* Merck, Boston, MA.

The goal of this study was to use a novel, whole-transcriptome array to identify SNPs that affect alternative splicing in human brain. RNA was extracted from the prefrontal cortex of 371 samples (190 Alzheimer's disease, 74 Huntington's disease, and 107 controls) collected at the Harvard Brain Tissue Resource Center. A custom Affymetrix array with ~1 million probe sets specific to exons and junctions of 20,000 well-established human genes and un-annotated stretches of the genome was designed. This allows expression measurement at exon level in addition to gene level. Quality control on the new whole-transcriptome data was conducted using principal component analysis. Gene level expression was summarized by averaging expression of constitutive probe-sets (exon probe-sets that cover more than 50% of alternative transcripts). SNP genotyping was conducted on an Illumina 650K array. After quality control and imputation, a total of 11.5 million SNPs were used in the analysis. Linear regression was performed to test associations between individual SNPs and probe-set level expression or gene level expression. An FDR value of 10% based on permutation was used to determine the significance level for cis probe-set level associations and cis gene level associations separately. As a result, 2503 cis gene level expression signals were identified. To further identify splicing signals controlled by SNPs, probe-set level associations where probe-set expression was highly correlated with gene level expression ($r2>0.5$) were excluded. As a result, 7941 genes with cis splicing signals were identified. Based on the patterns of alternative splicing, splicing signals were classified into categories such as exon skipping, alternative donors or acceptors, complex changes of multiple event types. To confirm the identified splicing signals, Illumina targeted RNA-seq is being performed on several candidate signals selected from different categories. To our knowledge, this is the first whole transcriptome array study to identify alternative splicing signals controlled by SNPs using human brain samples. As brain demonstrates a higher rate of alternative splicing than any other tissue and a large number of neurological disorders have been linked in aberrant splicing, the findings of the current study will help us understand genetic mechanisms underlying complex neurological disorders.

664F

Gene Expression Profiling of Young and Adult Mouse Cochlea by RNA-Seq in Strains with Normal and Age-Related Hearing Loss. *A.B.S. Giersch[1,3], J. Shen[1,3], N.G. Robertson[1], K. Wong[2], C.C. Morton[1,2,3].* 1) Pathology, Brigham and Women's Hospital, Boston, MA., USA; 2) Obstetrics and Gynecology, Brigham and Women's Hospital, Boston MA, USA; 3) Harvard Medical School, Boston, MA, USA.

Purpose: Age-related hearing loss is the most common sensory deficit in humans, reducing the quality of life in the aged population. Its social impact will become more pronounced as life expectancy continues to increase. Despite the discovery of many deafness genes, understanding of the pathophysiology of progressive hearing impairment due to the aging process remains elusive. We hypothesize that gene expression profiling of the cochlea from mouse models with various degrees of age-related hearing loss will reveal the molecular mechanism and inform potential target selection for prevention and treatment. Methods: We performed gene expression profiling of mouse cochlea by next-generation sequencing (RNA-Seq). A multi-factorial design was used. Cochleas from mouse strains with good hearing past one year of age (CBA/CaJ and B6.CAST-Cdh23Ahl+) or with documented age-related hearing loss (CochG88E/G88E and Coch-/- in a CBA background, and C57BL/6J) were dissected at discrete ages ranging from one week through late adulthood. Poly(A) selected mRNAs were extracted and the derived cDNA samples were fragmented, indexed, pooled, and sequenced by Illumina HiSeq. Biological replicates were used for all conditions. Expression levels of all transcripts were analyzed and differential gene expression analyses were performed. Results: We have obtained gene expression profiles of mammalian cochlea at various ages by RNA-Seq. With total reads of at least 30 million, more than 16,000 genes were detectable in each sample, and the expression levels were highly reproducible. We detect significant systematic differences in gene expression profiles between C57BL/6J and CBA/CaJ strain backgrounds, regardless of age. Comparing mouse models with or without age-related hearing loss of the same genetic background, we have found that few genes show statistically significant differential expression at young ages before the onset of hearing impairment, but the number of genes dramatically increases to hundreds at later stages. In addition, hundreds of genes show significant temporal changes on both genetic backgrounds. Summary: We have surveyed gene expression in mammalian cochlea by RNA-Seq and identified genes that show age-related differential expression. Systematic differences exist between different genetic backgrounds. Temporal gene expression profiles in the cochlea may suggest candidate targets for prevention and treatment of age-related hearing loss.

665W

Transcriptional response to 1,25(OH)2 vitamin D in the human colon of African- and European- Americans. *S. Kupfer[1], K. Ceryes[1], M. Chase[1], J. Maranville[2], A. Di Rienzo[2].* 1) Medicine, University of Chicago, Chicago, IL; 2) Human Genetics, University of Chicago, Chicago, IL.

$1,25(OH)_2$ vitamin D (vitD) is a steroid hormone with anti-inflammatory and anti-proliferative effects that functions through direct transcriptional mechanisms. In colonic diseases, such as colorectal cancer, vitD is thought to play a protective role; however, it is not known to what extent inter-individual and inter-ethnic differences in transcriptional response could underlie differences in disease susceptibility especially in African-Americans (AA) who have the highest incidence of colorectal cancer in the US. To study differences in transcriptional response in the colon, we have optimized an *ex vivo* culture system using colonic biopsies obtained during colonoscopy, thereby eliminating biological, technical and environmental confounders. Four biopsies obtained from the rectosigmoid junction are cultured; 2 biopsies each are treated with 0.1M $1,25(OH)_2$ vitD or vehicle control (EtOH). Samples are incubated for 6 hours at 37°C and placed in RNAlater. After extraction, RNA from two treatment replicates are pooled and hybridized to Illumina HumanHT-12 expression chips. We have conducted an initial study in 24 individuals equally split between AA and European-Americans (EA). We found 508 of 11,317 differentially expressed genes in response to vitD treatment (FDR< 0.01). Among 307 up-regulated genes, we found several known vitD responsive genes including *CYP24A1* ($p<2.2\times 10^{-16}$), *TRPV6* ($p=6.93 \times 10^{-8}$), and *CD14* ($p<2.2 \times 10^{-16}$). In addition to *CD14*, we found a number of up-regulated immune-related genes including *IRF8* and *TLR4*. In a preliminary study of 24 additional individuals, we have replicated these findings. We also noted that for 33 genes there were significant differences in genome-wide log-fold change between AA and EA. Among these, suppressor of AP-1, regulated by IFN (*SARI* or *BATF2*) was significantly upregulated in response to vitD in EAs but not in AAs. In summary, we found significantly up- and down-regulated genes in response to $1,25(OH)_2$ vitD in human colon biopsies. Several known vitD-responsive genes were up-regulated. We noted several up-regulated immune-related genes including *CD14* and *TLR4*. We also found genes with significantly different log-fold change in response to vitD between AA and EA suggesting that inter-ethnic differences in the vitD pathway include transcriptional response to a given amount of vitD. Further studies of genes identified in this study may reveal important vitD responsive pathways that underlie disease disparities.

666T

Participation of individual genes, signaling pathways and biological processes related to inflammation in gestational diabetes mellitus. *N.B. Cezar[1], A.F. Evangelista[1], D.J. Xavier[1], A.F. Assis[1], T.C. Arns[1], M.C. Foss-Freitas[2], M.C. Foss[2], E.T. Sakamoto-Hojo[3], E.A. Donadi[2], G.A. Passos[1].* 1) Department of Genetics - Faculty of Medicine of Ribeirão Preto - University of São Paulo; 2) Department of Clinical Medicine - Faculty of Medicine of Ribeirão Preto - University of São Paulo; 3) Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, Department of Biology, University of São Paulo.

Gestational diabetes mellitus (GDM) is the most common metabolic disorder found during gestation and has been defined as an abnormal glucose metabolism first diagnosed during pregnancy. Women who have a history of GDM usually present postpartum diabetes, insulin resistance, metabolic syndrome, hypertension and dyslipidemia. Several studies have reported the induction of genes involved with inflammatory response in GDM, providing the basis to propose that inflammation may be associated with the disease. We have previously shown that meta-analysis of the transcription profile of patients with type 1 diabetes are closer to GDM than to type 2 diabetes, and that inflammation genes were primarily modulated. In this study, we compared the transcription profiling of peripheral blood mononuclear cells of 18 GDM patients and 10 healthy pregnant women. Total RNA samples were hybridized to Agilent ® 4 x 44 K oligo microarrays encompassing the whole human functional genome. Differentially expressed mRNAs were obtained by Rank Products analysis, which classifies transcripts according to their fold change values. The hierarchical clustering of mRNAs and samples were performed using the Cluster program, dendrograms and spatial representations of mRNAs were constructed using the TreeView software. A total of 731 differentially and significantly expressed mRNAs were observed when patients were compared to controls. Molecular functions and biological pathways were analyzed using the DAVID database identifying 130 biological processes and nine signaling pathways ($P<0.05$), most of these related to inflammation and immune system. Major induced and biologically relevant transcripts exhibiting greatest fold change values were associated with inflammatory response, including *CXCL2*, *NFKBIA*, *IL1*β and *TNF*, which were individually validated by real-time PCR. Besides these genes, GDM patients exhibited modulated genes previously associated with metabolic abnormalities (*IGFBP2* and *TCF3* up regulated and *OLR1* and *TCF7L2* down regulated) and induced genes related to the Major Histocompatibility Complex (*HLA-DRB6*, *DQB1*, *DQB2*, *DOA* and *DOA2*). The modulation of genes associated with inflammation signaling pathways and immune system response in GDM patients reinforce the hypothesis that inflammation may be associated with hyperglycemia and insulin sensitivity dysregulation, suggesting that mRNAs that regulate inflammatory responses represent major functional genes in GDM patients.

667F

Gene Expression Studies by RNA-Seq in Airway Epithelial Cells (AECs) from Asthmatic and Non-Asthmatic Individuals. *R.A. Myers[1], J. Nicodemus-Johnson[1], D.K. Hogarth[2], J. Sudi[2], J.F. McConville[2], E.T. Naureckas[2], A.I. Sperling[2], J. Solway[2], J.A. Krishnan[2], S.R. White[2], D.L. Nicolae[1,2,3], C. Ober[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

Asthma is a chronic inflammatory disease characterized by reversible airflow obstruction. Airway epithelial cells (AECs) form the interface between the environment and the host and are functionally important in asthma pathogenesis. We hypothesized that transcript abundance in AECs differs between asthmatic and non-asthmatic individuals. In this study, we characterized the transcripts present and identify differentially expressed genes by RNA-Seq in 48 airway epithelial cell samples (27 asthmatics, 21 non-asthmatics) using the Illumina HiSeq 2000 platform. Sequences were mapped to the genome using BWA and the number of sequences mapped to protein coding genes was determined using BEDTools. We used edgeR to test for differential expression controlling for technical and biological factors that influence gene expression. We identified as expressed 14,484 genes with at least one count per million sequences (cpm) in at least nine samples. At a false discovery rate of 10%, 43 genes were differentially expressed, 25 with increased expression and 18 with decreased expression in asthmatics compared to non-asthmatics. The gene with the most significant differential expression is regulator of G-protein signaling 2 (*RGS2*, p-value = 5.39 ×10-9), which showed increased expression in asthmatics. *RSG2* inhibits G-protein signal transduction by increasing GTPase activity of G-protein alpha subunits. As expression of *RGS2* is induced by long acting beta2-adrenoceptor agonists and glucocorticoids, the gene expression pattern may reflect responses to medication in the individuals with asthma. Other differentially expressed genes included *ALOX15B*, involved in arachidonic acid metabolism; *TC1* (*C8orf4*), a regulator of inflammation response; chemokine receptors *CCR2* and *CX3CR1*; and *KCNJ5*, a subunit of a G-protein activated potassium channel. Ten pathways showed enrichment (p < 0.01) of differentially expressed genes (p-values < 0.05 and fold changes > 2), including the antigen presentation pathway (p-value = 0.0005), atherosclerosis signaling, (p-value = 0.0019), and cAMP-mediated signaling (p-value = 0.0027). Through RNA sequencing in AECs, we identified differentially expressed genes and pathways that may be potential targets for asthma treatment. This work is supported by NIH grant U19 AI095320.

668W

Age-related variations in the transcriptome and methylome of human monocytes. *L.M. Reynolds[1], J. Ding[1], K. Lohman[1], J.R. Taylor[1], A. de la Fuente[2], T.F. Liu[1], C. Johnson[3], R.G. Barr[4], T.C. Register[1], K.M. Donohue[4], M.V. Talor[5], D. Cihakova[5], C.C. Gu[6], J. Divers[1], J.Z. Kuo[7], D. Siscovick[3], G. Burke[1], W. Post[5], S. Shea[4], D.R. Jacobs Jr[8], I. Hoeschele[9], C.E. McCall[1], S.B. Kritchevsky[1], D.M. Herrington[1], R.P. Tracy[10], Y. Liu[1].* 1) Depts of Epidemiology and Prevention, Internal medicine, and Translational Research Inst, Wake Forest School of Medicine, Winston-Salem, NC; 2) FBN, Leibniz Institute for Farm Animal Biology, Genetics and Biometry, Mecklenburg-Vorpommern, Germany; 3) Cardiovascular Health Research Unit, Depts of Medicine and Epidemiology, University of Washington, Seattle, WA; 4) Depts of Medicine and Epidemiology, Columbia University Medical Center, New York, NY; 5) Johns Hopkins University, Maryland, MD; 6) Washington University School of Medicine, St. Louis, MO; 7) Medical Genetics Inst, Cedars-Sinai Medical Center, Los Angeles, CA; 8) Div of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 9) Virginia Bioinformatics Inst, Virginia Tech, Blacksburg, VA; 10) Dept of Pathology, University of Vermont College of Medicine, Colchester, VT.

Identifying age-associated variations in the human transcriptome and epigenome may provide novel insights into the molecular basis of aging. Published transcriptomic and epigenomic studies of aging have been limited by small samples sizes, mixed cell types, and the lack of both expression and methylation profiles from the same samples. We analyzed genome-wide mRNA expression and DNA methylation profiles in purified peripheral CD14+ monocytes from 1,264 subjects ranging in age from 55 to 94 years. A total of 2,498 genes were significantly associated with chronological age (false discovery rate (FDR) ≤0.001) after adjustment for technical and biological covariates. We further identified six networks of co-expressed genes that included prominent genes from three correlated pathways: autophagy, oxidative phosphorylation, and translation, with expression patterns suggesting these pathways decline with age. The most significant age-associated gene network (first eigenvector: R^2 = 9.7%, p=1.79×10^{-30}) contained three up-regulated genes including *MCL1* (Myeloid Cell Leukemia sequence 1), an inhibitor of autophagy. Methylomic analyses revealed 2,093 CpG sites whose methylation levels were significantly associated with both age and with mRNA expression of nearby genes (± 1Mb). Mediation analyses (structural equation modeling) support methylation of 279 of these CpG sites as mediators between age and expression of 202 genes, which were enriched for antigen processing and presentation genes. One of the most significant age-associated methylation sites (cg13007871) was located in a predicted enhancer region in the gene body of an MHC Class I gene, *HLA-E*, and was negatively associated age (FDR=7.54×10–24) and with mRNA expression of *HLA-E* (FDR=3.08×10–11), while *HLA-E* expression was positively associated with age (FDR=1.19×10^{-8}). In summary, the data from this large cross-sectional population-based study in humans provide evidence that age-related differences in methylation may mediate the relationship between age and expression of antigen processing and presentation genes, and support previous in vitro and in vivo findings that autophagy and mitochondrial function decline with age. The identification of important molecular features of aging in humans may lead to new insights about the aging process and possible targets for interventions of aging-associated disease and disability.

669T

Transcriptome profiling of human ulcerative colitis mucosa shows altered expression of pathways overlapping genome-wide association susceptibility loci. *C.J. Cardinale[1], Z. Wei[2], J. Li[1], J. Zhu[2], R.N. Baldassano[3], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 3) Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA.

Tissue of the human colonic mucosa which has been altered by inflammation due to ulcerative colitis (UC) displays a drastically altered pattern of gene expression based on microarray transcript profiling. We have utilized three independently-generated, publically-available sets of gene expression data from endoscopic biopsies of the colon in healthy controls, ulcerative colitis inflamed tissue, tissue uninvolved in inflammation in UC patients, and Crohn's disease lesions in both adult and pediatric patient populations. We show that gene expression patterns in active UC lesions are easily distinguishable from healthy control tissue, while the grossly uninflamed tissue or Crohn's lesions may show patterns consistent with healthy mucosa, inflammatory gene expression, or an intermediate pattern. Gene set enrichment analysis (GSEA) illustrates that all three data sets share in common 87 gene sets upregulated in UC lesions and 8 that are downregulated (false discovery rate < 0.05). The upregulated pathways are dominated by gene sets involved in immune function and signaling as well as the control of mitosis. We compared these pathways with GSEA pathway-based meta-analysis obtained from GWAS loci from six European cohorts with 5584 UC cases and 11587 controls, giving 56 gene sets (FDR < 0.05). The upregulated pathways show substantial ovelap, with 33 of the 87 expression-derived gene sets being represented among the 56 GWAS gene sets. Future sequencing work is being conducted to determine allelic expression influences of the major gene expression networks involved.

670F

EQTLs and allele specific expression of HLA haplotypes and amino acids associated to autoimmune diseases. *A. Zhernakova[1], M. vd Sijde[1], J. Gutierrez-Achury[1], P.J. McLaren[2], D.V. Zhernakova[1], P.I.W. de Bakker[3,4], H.J. Westra[1], L. Franke[1], C. Wijmenga[1], J. Fu[1].* 1) Genetics, UMCG, Groningen, Netherlands; 2) 2.School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland; 3) 3.Department of Medical Genetics, and Department of Epidemiology, University Medical Center Utrecht, Utrecht, The Netherlands; 4) 4.Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

HLA is the strongest associated locus in autoimmune diseases. Recent studies in celiac disease (CeD) and rheumatoid arthritis (RA) indicated the major role of DR3-DQ2 and DR4-DQ8 in both diseases, and identified additional independently associated variants both in the DR-DQ locus and in the extended HLA locus. We hypothesized that the mechanism of downstream effect of associated variants and haplotypes is due to an impact on the expression level of HLA genes We investigated this hypothesis in a dataset of 60 unrelated CEU individuals, for whom RNAseq data, dense HLA genotyping and imputation of HLA alleles was available. We selected CeD and RA associated amino acids, HLA-haplotypes and SNPs, and assessed allele specific expression and eQTLs in the HLA locus. eQTL analysis allowed us to assess the dosage effect of HLA variants on gene expression, whereas the allele-specific analysis, performed in heterozygous individuals for each variant, indicated if one or another allele of a SNP or amino acid was preferentially expressed. In both analyses we observed that an amino acid at position 52 of the DQB1 gene (DQB1_AA52) was significantly associated with expression of DQB1 gene both in eQTL and in allele-specific analysis. DQB1_AA52 is associated to CeD independently from the most associated DQB1 variant (AA55). We therefore confirmed that imbalanced allelic expression is the downstream effect of some HLA variants associated with autoimmune diseases. This analysis is currently ongoing in the population cohort of 760 individuals, for whom RNAseq data is available.

671W

Host and pathogen transcriptome profiling during *Leishmania major* infection. *L.A.L. Dillon[1,2], R. Suresh[1], K. Okrah[3], M. Mangione[1], J. Choi[1], H. Corrada Bravo[2,4], D.M. Mosser[1], N.M. El-Sayed[1,2].* 1) Department of Cell Biology & Molecular Genetics, University of Maryland, College Park, MD; 2) Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD; 3) Department of Mathematics, University of Maryland, College Park, MD; 4) Department of Computer Science, University of Maryland, College Park, MD.

Leishmaniasis, caused by protozoan parasites of the *Leishmania* genus, affects roughly 1.5 million people worldwide and is endemic primarily to South America and the Middle East. The parasite's lifecycle is divided between its insect vector, the phlebotomine sand fly, and its mammalian host, where it resides inside of macrophages. In order to establish an infection, the parasite must evade host immune system responses to infect macrophages and to survive and replicate within cellular phagolysosomes. Little is known about changes that take place at the transcriptional level in either the host or the pathogen during an infection and how *Leishmania* is able to direct these changes through manipulation of host cellular machinery.

Using the mouse as a model system, we have performed RNA-seq on the Illumina HiSeq platform to simultaneously measure global changes in gene expression in both *L. major* and host macrophages at multiple time points during the *L. major* lifecycle, including as the parasite replicates inside of host cells. Transcripts were aligned to the mouse and *L. major* genomes using TopHat and differential expression analysis was performed using limma after ComBat batch correction and voom data transformation. Gene ontology category enrichment analysis was done using GOseq. Genes that are differentially regulated in *L. major* include histones, antioxidant genes, cell surface markers, membrane transporters, and signaling molecules while those that are differentially expressed in mouse include those involved in immune system responses (e.g., inflammation, wound healing, leukocyte activation), cell signaling, metabolism, and apoptosis.

The identification and quantification of up- and down-regulated genes has provided evidence regarding the mechanisms used by *L. major* to elude host defenses and survive in the intracellular environment and has enabled connections to be made between changes in gene expression and the parasite's biology and lifecycle. These findings are revealing new insights into the dynamics of an infection and will help identify markers and targets for the prevention, diagnosis, and treatment of Leishmaniasis and related kinetoplastid diseases.

672T

Comparative analysis of gene expression profiles in the cerebellum and granule neurons of the *Cstb*$^{-/-}$ mouse, a model for the Unverricht-Lundborg disease (EPM1). *T. Joensuu[1,2,3,4], S. Tegelberg[1,2,3,4], E. Reinmaa[1,2,4], M. Segerstråle[4,5], P. Hakala[1,2,3,4], H. Pehkonen[6], E.R. Korpi[6], J. Tyynelä[7], T. Taira[4,5], I. Hovatta[5,8], O. Kopra[1,2,3,4], A-E. Lehesjoki[1,2,3,4].* 1) Folkhälsan Institute of Genetics, FI-00290 Helsinki, Finland; 2) Department of Medical Genetics, Haartman Institute, University of Helsinki, FI-00014 Helsinki, Finland; 3) Research Programs Unit, Molecular Neurology, University of Helsinki, FI-00014 Helsinki, Finland; 4) Neuroscience Center, University of Helsinki, FI-00014 Helsinki, Finland; 5) Department of Biosciences, Faculty of Biological and Environmental Sciences and Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, FI-00014 Helsinki, Finland; 6) Institute of Biomedicine, Pharmacology, University of Helsinki, FI-00014 Helsinki, Finland; 7) Institute of Biomedicine, Biochemistry and Developmental Biology, University of Helsinki, FI-00014 Helsinki, Finland; 8) Mental Health and Substance Abuse Services, National Institute for Health and Welfare, FI-00300 Helsinki, Finland.

Progressive myoclonus epilepsy of Unverricht-Lundborg type (EPM1) is an autosomal recessively inherited disorder characterized by stimulus-sensitive myoclonus, epilepsy and progressive neuronal degeneration with ataxia and dysarthria appearing later. EPM1 occurs worldwide, but it is enriched in Finland and in the Western Mediterranean region. It is caused by mutations in the gene encoding a cysteine protease inhibitor cystatin B (*CSTB*). Altogether fourteen *CSTB* mutations have been reported to underlie EPM1, of which an unstable repeat expansion of at least 30 copies in the promoter region is found in 90% of the disease alleles worldwide. Despite identification and characterization of the underlying *CSTB* gene and mutations, the disease mechanisms in EPM1 are poorly understood. A *Cstb*$^{-/-}$ mouse, the model for the disease, shows EPM1 like disease phenotype with myoclonic seizures that develop by one month of age and progressive ataxia by six months of age accompanied with early microglial activation, neuron loss and sensitivity to oxidative stress. In an approach towards understanding the molecular basis of disease mechanisms in EPM1, we have characterized the genomewide gene expression profiles from postnatal day 30 (P30) and P7 *Cstb*$^{-/-}$ mice, as well as from the cultured cerebellar granule cells using a pathway-based approach. The microarray profiling revealed multiple biological pathways related to activation of inflammatory processes, ion homeostasis, synaptic function and plasticity, cell cycle, cellular architecture of cytoskeleton and intracellular transport reflecting the ongoing neuropathology in the *Cstb*$^{-/-}$ mice. As the altered genes from P7 *Cstb*$^{-/-}$ cerebellum implied changes in neuronal functions, which could reflect the hyperexcitablity and motor symptoms seen in EPM1, the GABAergic signaling pathway was selected for further characterization using electrophysiological recordings for spontaneous post-synaptic currents in Purkinje cells, as well as ligand binding and immunohistological analyses. Electrophysiological recordings from *Cstb*$^{-/-}$ Purkinje cells revealed a shift of balance towards decreased inhibition, while the amount of inhibitory interneurons did not change. The ligand binding capacity of GABA receptors was decreased along with reduced number of pre-synaptic GABAergic terminals. These results suggest that changes in GABAergic signaling could lead to reduced inhibition and neuronal hyperexcitability in *Cstb*$^{-/-}$ cerebellum.

673F

Peripheral blood microarrays identify dysregulated genes and pathways unique to psoriatic arthritis compared to psoriasis without arthritis. *R. Pollock[1], V. Chandran[1,2], C. Virtanen[3], F. Pellett[1], C. Rosen[4], K. Liang[5], D. Gladman[1,2].* 1) Psoriatic Arthritis Program, Toronto Western Research Institute, Toronto, Canada; 2) Division of Rheumatology, Department of Medicine, University of Toronto, Toronto, Canada; 3) Ontario Cancer Institute Genomics Centre, Toronto, Canada; 4) Division of Dermatology, Department of Medicine, University of Toronto, Toronto, Canada; 5) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada.

Psoriatic arthritis (PsA) is an inflammatory arthritis that develops in about 30% of patients with psoriasis (PsC). We aimed to identify a gene expression signature of PsA compared to PsC and controls to identify biomarkers and improve our understanding of the pathology of arthritis in psoriasis patients. Gene expression was measured in peripheral whole blood from 20 PsA patients, 20 PsC patients without arthritis, and 20 controls using Agilent Whole Human Genome Oligo microarrays. Statistical analyses were performed using commercial (GeneSpring GX) and open-source (Bioconductor) packages to identify significant differentially expressed genes (DEGs) between PsA vs. controls, PsC vs. controls, and PsA vs. PsC. DAVID and Ingenuity Pathway Analysis were performed to identify enriched annotations and pathways relative to the Agilent probe set reference. Eleven candidate biomarker genes were validated by qPCR using Taqman assays. Expression levels across all three groups were compared using one-way ANOVA and 1547 DEGs were found. Student's t-test found 790 DEGs between PsA vs. controls, 98 between PsA vs. PsC, and 0 between PsC vs. controls. DEGs in PsA vs. controls included genes involved in peptide loading onto the MHC Class I (TMTC3 and PDIA3), epidermal development and maintenance (CSTA), histone acetyltransferase complexes (EP300, CREBBP, MYST4), toll-like receptor 2 signaling (AKAP13), and NK/T cell migration (S1PR5). Two-way hierarchal clustering of the DEGs with a fold change > 2 grouped 16 out of 20 PsA patients together, with the remaining 4 patients clustering with the controls. DEGs in PsA vs. PsC included genes involved in toll-like receptor 4 signaling (LY96), T cell adhesion and activation (CD58), bone remodeling (TNFSF10), NK/T cell activation (CLEC4D, CLEC2B), and inflammation (BCL2A1, S100A12). Clustering of top DEGs grouped 19 out of 20 PsA samples together and 17 out of 20 PsC samples together. Enriched annotations and pathways in PsA vs. controls and PsA vs. PsC were related to the regulation of gene expression and activation of the PPARα/RXRα pathway. DNA double-strand break repair by non-homologous end joining was enriched in PsA vs. PsC ($p=1.7\times10{-}3$). Microarray and qPCR results were highly correlated (overall $r=0.92$, $p<0.001$, 95% CI 0.71-0.98). These data demonstrate that there is a unique peripheral blood gene expression signature of PsA compared to PsC, and microarrays can identify potential biomarkers of PsA in PsC patients.

674W

Deep Sequencing of microRNAs and novel tRF RNAs in human monocytes. *C. Li[1], H. Ge[1], C. Lindvay[1], M. Chhoa[2], E. Lamas[1], S. Escobar[3], H. Arnett[3], T. Juan[2], R. Sandrock[1].* 1) Genome Analysis Unit, Amgen Inc., San Francisco, California, 94080; 2) Protein Sciences, Amgen Inc., Thousand Oaks, California, 91320; 3) Inflammation Research, Amgen Inc., Seattle, Washington 98119, USA.

microRNAs play essential roles in modulating the level of gene expression post-transcriptionally in cells. Certain microRNAs have been shown to be up- or down-regulated during the process of immune responses. Using next-generation sequencing (NGS), we performed transcriptome analysis of small RNAs ranging from 18- to 30-nt in unstimulated human monocytes or those treated with an activating stimulus, lipopolysscharide (LPS). Upon aligning and counting fragments that mapped to the human genome, we found that 20 microRNAs accounted for approximately 75% of the entire microRNA population in monocytes. Within around 5-order range of differential expression, *miR-155* was the most abundant up-regulated gene in LPS-induced monocytes followed by *miR-146a*, *miR-193a*, *miR-9*, *miR-147b*, and *miR-31*. No significant down-regulation of microRNAs was observed in human LPS-treated monocytes. Activation of mature and precursor forms of the signature set of microRNAs was further confirmed in LPS-treated monocytes by a semi-quantitative analysis of tagged RT-PCR. In addition to known microRNAs, we identified 18 *tRNA* fragments (*tRFs*) sequences, an emerging type of novel small RNAs, of which little biological significance is known, that were originally identified in cancer cells that mapped to 50 different *tRNA* gene loci of which 13 were novel. *tRFs* are an emerging type of novel small RNAs, of which little biological significance is known, that were originally identified in cancer cells. Only *tRNA-TyrGTA* showed differential expression in human LPS-treated monocytes. Our study provides a comprehensive quantitative catalog of microRNA expression and demonstrates an analysis pipeline for identification of novel *tRF* RNAs that lays the groundwork for further investigation regard to the biological significance of *tRFs* in human monocytes.

675T

Platelet RNA And eXpression-1 Study Demonstrates RNA Expression Differences that Correlate with Ancestry. *L. Simon[1], E. Chen[1], L. Edelstein[2], P. Bray[2], C. Shaw[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Thomas Jefferson University, The Cardeza Foundation for Hematologic Research and the Department of Medicine, Jefferson Medical College, Philadelphia, PA, USA.

Cardiovascular disease, cancer and many drug responses show striking differences between groups with different ethnicity or ancestral background. The Platelet RNA And eXpression-1 (PRAX1) study was designed to investigate the function of platelets by profiling mRNAs and microRNAs (miRNAs) and identifying inter-individual variation in samples from individuals of European American (EAs) and African American (AAs) ancestry. Platelets present a unique opportunity for study of racial differences for many reasons including the breadth of functional variation and the ease with which samples from normal individuals can be obtained for functional and molecular characterization. We profiled the RNA levels in platelets of 154 healthy adults using genome-wide RNA expression analysis, and discovered large-scale differences in platelet gene expression between EAs and AAs. These observations are consistent with prior findings of other studies, including the 1000 Genomes Project as well as expression profiling of lymphoblastoid cell lines from the HapMap project. These studies have revealed genetic differences and cell line gene expression differences between groups of distinct ancestry. The differences we identified in PRAX1 are observed in purified human platelets both at the mRNA and miRNA level. At the mRNA level, race (ancestry) is significantly associated with the first principal component of variation in the genome wide expression data. In addition, mRNAs with predicted binding sites for miRNAs up-regulated in AAs are enriched among down-regulated mRNAs in AAs. A cluster of miRNAs encoded at the DLK1-DIO3 locus on human chromosome 14q32 is strongly up-regulated in EAs compared to AAs. The results of PRAX1 provide a comprehensive view of the population differences in human platelets measured at the level of RNA expression and suggest that ancestry should be taken into account.

676F

Genetic variation underlying protein expression variation in brain samples and lymphoblastoid cell lines. *R.J. Hause[1,2,3], C.T. Archer[3,4], L. Cheng[5], J.A. Badner[6], C. Liu[5], R.B. Jones[1,2,3].* 1) Ben May Department for Cancer Research, University of Chicago, Chicago, IL; 2) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 3) Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL; 4) Department of Human Genetics, University of Chicago, Chicago, IL; 5) Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; 6) Department of Psychiatry and Behavior Neuroscience, University of Chicago, Chicago, IL.

Genome-wide association analyses have discovered many DNA variants that influence complex psychiatric disorders. Because gene expression is a molecular phenotype that acts as an ‘intermediary’ between genetic and physiological variation, expression quantitative trait loci (eQTL) mapping has been used to try to better understand complex disease by identifying regulatory variation that affects mRNA expression levels. An implicit assumption made in these analyses is that eQTLs explain subsequent differences in protein levels, even though previous experiments have shown imperfect correlations between mRNA and protein expression levels. To improve our understanding of the genetic basis of protein expression variation and how this variation is related to genetic risks for complex disease, we have recently extended micro-western and reverse phase lysate array technologies to quantify 441 protein levels across 68 unrelated Yoruba lymphoblastoid cell lines (LCLs) and examined the correlations between mRNA levels, protein levels, and known physiological covariates. In this study, we sought to translate our protein analysis work to examine protein levels in the brain. We quantified the levels of 50 proteins related to cell signaling, transcription, and neurobiology across brain samples derived from a population of 129 unrelated individuals of Caucasian descent who were classified as normal controls ($n = 43$) or diagnosed as cases with depression ($n = 12$), schizophrenia ($n = 39$), or bipolar disorder ($n = 35$). We first identified several proteins that were differentially expressed between cases and controls, such as brain-derived neurotrophic factor (BDNF), and between particular psychiatric sub-classes and controls, such as thymus-specific serine protease (PRSS16) that was significantly overexpressed in depressed patients. We then performed association analyses on these protein levels to identify protein quantitative trait loci (pQTLs). We observed that up to a quarter of pQTLs previously identified in LCLs replicated in this brain cohort. Our approach represents a general method for the identification of novel protein variation associated with neuropsychiatric disorders and genetic variation that contributes to protein variation. In the future, we will investigate the overlap of pQTLs with SNPs previously implicated in neuropsychiatric disorders and scale up our investigation to analyze a substantially larger protein set.

677W

Identification of novel non-coding RNAs associated with smoking. *M.J. Peters[1,2], F.A.S. Smouter[1], J.G.J. van Rooij[1,2], D. Schmitz[1], M. Jhamai[1], P. Arp[1], F. Rivadeneira[1,2,3], A. Hofman[3], A.G. Uitterlinden[1,2,3], J.B.J. van Meurs[1,2].* 1) Department of Internal Medicine, Erasmus Medical Center Rotterdam, the Netherlands; 2) The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Rotterdam, the Netherlands; 3) Department of Epidemiology, Erasmus Medical Center Rotterdam, the Netherlands.

Background Smoking is the leading cause of premature death from diseases such as lung cancer and chronic respiratory disease; it harms nearly every organ and reduces the health of smokers in general. Since the toxin ingredients in cigarette smoke (like nicotine) are absorbed into the bloodstream, we hypothesized that smoking will alter the gene expression patterns of many coding and non-coding RNAs in the circulation. Our aim was to identify these RNAs changing with smoking. Oxidative stress is thought to be the general cause; smoking causes an imbalance between the systemic manifestation of reactive oxygen species and the system’s ability to repair the resulting damage. Methods We performed RNA sequencing in 92 women of the Rotterdam Study: 30 smokers and 62 non-smokers. Whole-blood was collected using PAXgene-tubes and the RNA was sequenced using the Illumina HiSeq 2000 (4 samples/lane). We aligned the reads to the Human Reference Genome (v19) using TopHat and we estimated the relative abundance using Cufflinks (Tuxedo Suite); transcripts with a coverage <3 were removed and they needed to be present in >2 samples. We used logistic regression to model the outcome variables and adjusted for age, RNA quality, technical batch, and cell counts. We mapped the transcripts to different RNA reference databases to distinguish known and novel coding and non-coding RNAs. Results We identified a total of 112 RNAs differentially regulated between smokers and non-smokers ($p<0.01$), of which 71 were present in any RNA reference database and 41 were novel. Examples of transcripts known to be associated with smoking behavior are ANKRD13B, GFRA2, UBQLNL, PIGO and OLFML1. We newly identified the TXNRD3 transcript: four smokers had transcribed this gene, while none of the non-smokers had detectable TXNRD3 RNA levels. The TXNRD3 protein catalyzes the reduction of thioredoxin, and is implicated in the defense against oxidative stress. 41 novel RNAs were differentially regulated between smokers and non-smokers: most interesting are the ones present in >=10 non-smokers and not present in smokers at all. These new intergenic transcripts are located upstream the LEF1 gene and upstream the CSTF1 gene. Conclusions These results show that transcription profiles alter during smoking. While replication is pursued, identification of both known and novel RNAs associated with smoking will add to our understanding of the molecular response in blood to cigarette smoke exposure.

678T

Using Metabolomics and ex vivo Activity Approaches to Understand the Functional role of Epoxide Hydrolase 2 gene in Anorexia Nervosa.
P. Shih[1], C. Morisseau[2], J. Yang[2], T. Clarke[3], A. Van Zeeland[4], A. Bergen[5], P. Magistretti[6], N. Schork[4], W. Berrettini[3], B.D. Hammock[2], W. Kaye[1], Price Foundation Collaborative Group. 1) Dep of Psychiatry, Univ California, San Diego, La Jolla, CA; 2) Dep of Entomology, Univ California, Davis, Davis, CA; 3) Dep of Psychiatry, Univ of Pennsylvania, Phila., PA; 4) The Scripps Translational Science Institute, La Jolla, CA; 5) SRI International, Menlo Park, CA; 6) The University of Lausanne, Lausanne, Switzerland.

Individuals with Anorexia nervosa (AN) restrict eating and become emaciated. They tend to have an aversion to foods rich in fat. Understanding genetic basis of fat metabolism may unravel AN molecular psychopathology. We recently identified a novel AN susceptibility gene, Epoxide Hydrolase 2 (*EPHX2*), through a series of complementary genetic study designs (GWAS, exon-sequencing, single-locus association and replication studies) in 1205 AN and 1948 controls (p=0.0004 to 0.00000016). The molecular mechanism by which *EPHX2* influences AN risk remains elusive; here we utilize metabolomics and *ex vivo* enzyme activity assays to evaluate the biological function of *EPHX2*. *EPHX2* codes for soluble epoxide hydrolase (sEH) which hydrolyzes and converts derivatives of arachidonic acid (AA), Epoxyeicosatrienoic acids (EETs) to Dihydroxyeicosatrienoic acids (DHETs). We measured AA metabolite markers (8,9-, 11,12- and 14,15-EETs) and their corresponding DHETs using the LC/MS/MS method in 14 AN and 6 age-,sex- matched controls with Arg287Gln (rs751141) and 3'-UTR SNP (rs1042064) genotypes. EET-to-DHETs ratios were calculated as proxy markers of *in vivo* sEH activity. Direct sEH activity was also measured *ex vivo* using a radioactive surrogate substrate on buffy coat in an independent sample of 36 controls. No significant differences in AA metabolite markers were observed between AN and controls except for increased 8,9- and 14,15-EETs in AN (p=0.05). Variant allele carriers of Arg287Gln and 3'-UTR showed decreased levels of all 3 EETs (p=0.04 to 0.00002) and lower 11,12- and 14,15- EET-to-DHETs ratios (p=0.04 to 0.0016), suggesting elevated basal sEH activity in variant allele carriers. However, the *ex vivo* measurement of sEH activity did not show association with genotypes. Our earlier work using the bacculovirus expression system found the Arg287Gln variant allele to associate with decreased *in vitro* sEH activity; here, through joint assessment of sEH's biochemical target, AA epoxide derivatives, we found variant alleles associated with elevated sEH activity. This suggests that sEH expression/activity *in vivo* is complex, context-dependent, and likely to be influenced by interactions of multiple loci instead of single variant(s) with a large effect. This study suggests *EPHX2* influences AN risk through biological interaction with the AA pathway, and demonstrated the importance of a multi-disciplinary approach to examine molecular functions of risk genes.

679F

The African Genome Variation Project Phase II: Down-sampling of African whole genome sequence data.
T. Carstensen[1], D. Gurdasani[1,2]. 1) Genetic Epidemiology, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) Cambridge University, Cambridge, Cambridgeshire, United Kingdom.

/Background Although several studies have focused on the utility of ultra-low coverage (ULC) designs and dense chip arrays in European populations, the utility of such platforms for large scale genomic studies in Africa, which has much greater genetic diversity, is unknown. Here, we present preliminary findings on the utility of ULC whole genome sequencing (WGS) designs and dense chip arrays for genomic research in Africa. /Methods We sequenced whole genomes and genotyped samples from 320 individuals from 3 geographically distinct African populations; i.e. Baganda in Uganda, Zulu in South Africa and Amhara, Oromo and Somali from Ethiopia. Sequencing and genotyping were carried out on the Illumina HiSeq 2000 and Illumina Omni 2.5M platform, respectively. We randomly down-sampled 4x WGS data to 2x, 1x and 0.5x. Variant calling was carried out using GATK followed by genotype refinement with BEAGLE using the multi-ethnic 1000 Genomes Project reference panel. We compared the accuracy and sensitivity of ULC-WGS at different depths and imputed chip array data to 4x WGS. The effective sample size was calculated for all designs as n*R2 where R2 is the correlation between imputed sequence/chip data and actual chip array genotypes, and n are the number of samples that can be sequenced/genotyped for a given cost. /Results The correlation between sequence data and genotype data was 0.8-0.9 for 0.5x and greater than 0.9 for all other ULC-WGS designs for common (MAF>5%) variants. For rare variants (MAF<5%) it was 0.6-0.7, 0.7-0.8, 0.8-0.9 and 0.9 in the 3 populations at 0.5x, 1x, 2x and 4x, respectively. Imputation markedly improved accuracy of WGS data with respect to chip data at all coverage depths, and appeared to be the primary determinant of accuracy. Sensitivity of imputed chip array data was comparable to 2x data with respect to capture of variants obtained with 4x. The effective sample size for both common and rare variants was greater for all ULC-WGS and chip array designs than for 4x. /Conclusions We show that high accuracy in genotype calling can be obtained with ULC-WGS and chip array designs for common variants in African populations when large multi-ethnic panels are used for imputation. Even for rare variants, greater effective sample sizes are achieved with ULC-WGS and chip designs compared to 4x WGS for a given cost. Developing additional Africa-specific reference panels for imputation is likely to augment accuracy of ULC-WGS designs even further.

680W

SRD5A2 gene polymorphisms affect the risk of benign prostatic hyperplasia.
V.K. Choubey[1], S.N. Sankhwar[1], J. Carlus[2], A.N. Singh[1], D. Dalela[1], R. Singh[3]. 1) Department of Urology, King George's Medical University, Lucknow, Lucknow, India; 2) Centre for Genetics and Inherited Diseases (CGID), Taibah University, Al- Madinah, Kingdom of Saudi Arabia; 3) Division of Endocrinology, Central Drug Research Institute, Lucknow, India.

Abstract Background Benign prostatic hyperplasia (BPH), characterized by an enlarged prostate, affects the quality of life in the elderly people, and prostate cancer (PC) is the second leading cause of death in men. Since prostate is an androgen dependent tissue, androgen metabolism is likely to affect the risk of developing BPH and PC. Methods We have analyzed common polymorphisms in the SRD5A2 gene in 210 BPH patients, 192 PC cases and 171 controls. Published data on V89L and (TA)n repeat polymorphisms was pooled for performing meta-analysis to quantitate the BPH risk associated with these polymorphisms. In total, six studies on V89L polymorphism comprising a total of 878 patients and 586 controls and three studies on (TA)n polymorphism comprising a total of 768 patients and 460 controls, were included in meta-analysis. Results We did not observe any polymorphism at the A49T polymorphic site. All the subjects showed 'AA' at this site without any substitution. The presence of 'VV' increased BPH risk (p = 0.047); however, no significant difference in the genotype distributions comparison between PC vs Controls was observed. Comparison of genotypes at (TA)n locus found highly significant difference between cases and controls such that shorter repeats increased BPH risk (p = 0.003). However, no significant difference in comparison of PC vs Controls was observed. Meta-analysis also showed high frequency of 'V' at V89L and (TA)0 repeats in the BPH cases, though the differences were not statistically significant. Conclusion A49T locus of SRD5A2 gene is monomorphic in the study population. The presence of VV at V89L Locus and shorter (TA)n repeats increased BPH risk, though the meta-analysis suggest no overall correlation between SRD5A2 polymorphisms and the risk of BPH. None of these polymorphisms appears to affect the risk of PC. Keywords: SRD5A2, benign prostatic hyperplasia, prostate cancer, 5α-reductase type 2.

681T

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

Meiotic recombination contributes to genetic diversity and ensures the correct segregation of chromosomes. It is initiated by the introduction of double strand breaks (DSBs) by the SPO11 protein and occurs in narrow regions of the genome called hotspots. Computational analysis of patterns of linkage disequilibrium (LD) allowed for the identification of ~30,000 population hotspots in humans. Here, we use an approach previously developed by us to generate high-resolution genome-wide personal maps of meiotic DSBs. The Zn-finger protein PRMD9 has been shown to define the location of the majority of meiotic DSB hotspots. We mapped DSBs in six individuals; two carrying the most common Prdm9 allele (A), two heterozygous for the A allele and a closely related variant, the B allele, one heterozygous for the A allele and the C allele (a variant commonly found in African populations) and one heterozygous for the C allele and a C-type variant, L4. The A and B alleles of Prdm9 defined similar recombination initiation hotspots while the C allele defines a distinct set of hotspots. Approximately 60% of population LD hotspots are explained by A-defined hotspots, while C-defined hotspots explain an additional 12%. This shows that relatively minor alleles significantly contribute to the LD map. We also found that the DSB distribution exhibits a strong telomeric bias which resembles that of male, but not female crossovers indicating that frequency of DSBs is a major determinant of crossover distribution and frequency in human males. Examination of the two AA individuals revealed inter-individual variation at about 5% of hotspots. Whole genome sequencing of these individuals determined that no more than 40% of AA polymorphic hotspots could be explained by a single nucleotide variant at a putative PRDM9 binding site. Interestingly, we also found that polymorphic hotspots were frequently found in clusters. We used H3K4me3 ChIP-Seq signal strength as a proxy for PRDM9 binding affinity. Unlike in mouse, we found that the hotspot strength is not well correlated with the strength of the H3K4me3 signal. In aggregate, these data indicate that PRDM9 binding is not the only factor modulating hotspot strength in humans. Finally, we explored the role of DSB hotspots in genomic rearrangements. We found that DSB hotspots were enriched at structural variants that arise via homology-mediated mechanisms and that meiotic DSBs occur at well known disease-associated chromosomal breakpoints.

682F

Comparative analysis of biological activity of 11 mammalian paralogs of the novel human interferon IFNλ-4 (IFNL4) associated with viral clearance. *W. Tang[1], M. Dennis[2], L. Prokunina-Olsson[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, NCI; 2) Department of Genome Sciences, University of Washington.

Interferons (IFNs) are proteins produced by host cells in response to infections. Recently, we discovered a new type III-interferon, interferon-λ4 (IFNL4). IFNL4 is encoded by a gene located upstream of the IFNL3 (former IL28B) on chromosome 19. We showed that the ancestral dG allele of a dinucleotide variant, ss469415590 (dG/TT), which creates a frameshift and leads to a generation of a full-length functional IFNL4 protein, is strongly associated with the inability to clear HCV infection spontaneously or after treatment (Prokunina-Olsson et.al, Nat Gen, 2013). Utilizing public databases and targeted re-sequencing, we now explored the existence of IFNL4 in other species. The IFNL4 genomic region was absent in mouse and rat, likely due to a deletion in a common murine ancestor, and not found in any non-mammals. All mammals with available genomic sequence were found to be monomorphic for the presence of the IFNL4-generating dG allele. We created expression constructs for 11 synthetic IFNL4 paralogs based on predicted open-reading frames deduced from corresponding genomic sequences. Expression constructs encoding the human, chimpanzee, orangutan, cynomolgus, rhesus, marmoset, megabat, dog, pig, cow, elephant and panda IFNL4 proteins were transfected into a hepatic HepG2 cell line stably expressing a luciferase reporter assay under the control of an interferon-stimulated response element (ISRE). The activity of IFNL4 paralogs was evaluated as their ability to induce ISRE-Luc activation and compared to the activity of human IFNL4. We found that the human IFNL4 was the most biologically active. Paralogs separated into three groups: orangutan, cynomolgus monkey and megabat achieved ~70% biological activity; chimpanzee, dog and pig were around 50%; and cow, elephant, marmoset, panda and rhesus were around 20%, compared to human IFNL4. Interestingly, in non-primates the IFNL4 activity was lowest in herbivorous species (panda, cow and elephant) compared to omnivorous species (dog, pig) and megabat (consumes insects and blood). In primates, the strong diversity in IFNL4 biological activity could be attributed just to a few single point amino-acid differences, suggesting functional importance of these variants. In conclusion, we identified and characterized biological activity of IFNL4 paralogs in 11 mammalian species. Further analysis will explore evolutionary forces that shaped the IFNL4 protein sequence and biological activity.

683W

Analysis of the expression of celiac disease associated genes in T-cells. *C. Coleman, E. Quinn, R. Grealy, V. Trimble, R. McManus.* 1.Genomic Research, Institute of Molecular Medicine, Trinity Centre for Health Sciences, St. James' Hospital, Dublin 8, Ireland.

Coeliac disease (CD) is a common, complex and chronic immune-mediated disease affecting the small intestine. It is triggered in genetically sensitive individuals by the ingestion of gluten proteins, which initiate an immune reaction that leads ultimately to the destruction of the normal architecture of the gut wall, most obviously the loss of the villus architecture and the development of crypt hyperplasia. There is a strong inherited component to the disease, demonstrated by the fact that concordance rates in monozygotic twins are at least 75%. HLA-DQA1 and HLA-DQB1 alleles are necessary but not sufficient to cause disease and recent genome-wide association studies (GWAS) have succeeded in identifying almost 40 non-HLA risk loci. For the majority of these non-HLA loci very little is known with regards their potential role in immune function and/or the mechanism by which they might be contributing to disease biology. Given that the activation of T cells in the intestinal mucosa in response to gluten exposure is thought to play a key role in the pathogenesis of the disease, we measured the mRNA levels of disease associated candidate genes specifically in activated and resting CD4+ T cells in a group of coeliac individuals and unaffected controls. Results indicate that a number of genes that have been genetically associated with the disease also show significantly different expression at the mRNA level in CD4+ T cells and further investigation of these genes through the integration of Immunochip derived genetic data and pathway analysis will hopefully lead to a greater understanding of their role in disease biology.

684T

Alternative promoter activation leads to the expression of a novel variant of human lysyl oxidase (*LOX-v2*) that functions as an amine oxidase. *Y. Kim, S. Park, S. Kim.* Department of Biochemistry, Wonkwang University School of Medicine, Iksan, Jeonbuk, South Korea.

Lysyl oxidase (LOX) is a copper-dependent amine oxidase that is responsible for the lysine-mediated cross-links found in the extracellular matrix proteins, such collagen and elastin. Four additional *LOX*-like genes (*LOXL1*, *LOXL2*, *LOXL3*, and *LOXL4*) have been identified in human, each encoding the functional domains of LOX, such as a copper-binding domain, residues for lysyl-tyrosyl quinone (LTQ) and a cytokine receptor-like (CRL) domain. Several novel functions including tumor suppression, tumor progression, cellular senescence, and chemotaxis have been recently attributed to LOX. The presence of LOX paralogues, thus, suggests that the diverse multiple functions of LOX may be derived from differential regulation of the LOX paralogues. In searches of more human *LOX* paralogues, we identified several expressed sequence tag (EST)-clones that showed an alternative exon-intron splice pattern from LOX. Those ESTs corresponded to the *LOX* transcript variant 2 (*LOX-v2*) that was recently reported in the GenBank (accession no. NM_00117812). *LOX-v2* is lack of exon 1 of *LOX*, encoding a 188 amino acid-long polypeptide of 22 kDa. In peroxidase-coupled fluorometric assays, LOX-v2 showed a significant amine oxidase activity toward collagen and elastin. In RT-PCR analysis with human tissues, *LOX-v2* showed distinct tissue specificity from *LOX*. An alternative promoter element present in the intron 1 region of *LOX* was sufficient for the differential transcriptional activation of *LOX-v2*. These findings indicate that the human *LOX* gene encodes two variants, LOX and LOX-v2, both of which function as amine oxidases with distinct tissue specificities from one another.

685F

Real-time PCR for Indel markers detection in cffDNA. *M.D. Santos[1], F.M.F. Nunes[2], A.L. Simões[1].* 1) Department of Genetics, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil; 2) Department of Genetics and Evolution, Federal University of Sao Carlos, Sao Carlos, Sao Paulo, Brazil.

Fetal DNA present in pregnant women plasma is fragmented and in lower concentrations than maternal DNA. Therefore, prenatal detection of fetal genotype require methods based on genetic markers detectable by analysis of small segments of the genome, as well as being able to avoid competition between maternal and fetal DNA primer binding site. In this context, we developed a methodology based on real-time PCR for insertion/deletion polymorphism (Indel) detection in cell-free fetal DNA (cffDNA) present in pregnant women plasma.Three Indel loci (MID1386, MID818 and MID856) were chosen, for which three primers were designed. Primers flanking the insertion region were used in standard PCR to confirm mothers (n=150) and their newborns (n=150) genotype. The primer complementary to the insertion segment and a complementary flanking primer were used in real time PCR in presence of SYBR Green I followed by melting curve analysis to detect the insertion polymorphism in DNA samples extracted from maternal plasma. Melting curve analysis indicated the melting peak at 71.3°C, 81.5°C and 73.8°C, corresponding to the presence of insertion sequence of loci MID1386, MID818 and MID856, respectively, in plasma of pregnant women homozygous for deletion polymorphism, whose newborns were identified as heterozygotes by conventional PCR. The real-time PCR method presented in this study was able to detect fetal DNA present in maternal plasma by avoiding competition for primer binding site. Increasing number of these markers will allow its use in non-invasive prenatal identification and diagnosis when is no possible use STRs markers. Acknowledgement: we thank Zilá Luz Paulino Simões, PhD for Applied Biosystems 7500 Real-Time PCR System grant.

686W

Haplotyping single nucleotide polymorphisms and distinguishing deletion carrier status. *D. Maar, S. Tzonev, N. Klitgord, S. Cooper, G. Karlin-Neumann, K. Hamby, J. Regan.* Bio-Rad Laboratories, Pleasanton, CA.

Haplotype information is routinely used to make critical medical decisions, such as HLA allele matching for successful tissue transplantation. With increasing frequency, other examples are emerging where cis-acting sites rather than single sites alone appear to be required to produce a certain phenotype. Examples include: a β-globin locus associated with less severe sickle cell disease, an IL10 promoter region associated with a lower incidence of graft-versus-host disease, and an ApoE locus that when configured in cis affects lipid metabolism. However, the lack of tools to haplotype certain alleles has slowed the discovery of additional associations. To speed discovery, the HapMap Consortium initiated the HapMap Project in 2002 to map roughly 3 million SNPs. Project contributors used pedigrees, allele-specific PCR (AS-PCR), long-range sequencing, and somatic cell hybrids to assist in mapping these SNPs. These methodologies have limitations, namely: late-onset diseases cannot be reasonably investigated because pedigree analysis typically depends on linkage disequilibrium of microsatellite and SNP markers across several generations; AS-PCR is hampered by the inefficiencies of long-range PCR and is prone to false positives; long-range sequencing is limited to targets within ~3000 bp of each other; and somatic cell hybridization is costly and time consuming. Here, we introduce droplet digital PCR (ddPCR) for digital linkage analysis, which is a low cost, rapid methodology to haplotype heterogeneous loci that are up to at least 30 kb apart. The technique is independent from PCR efficiency, does not rely on long-range PCR, and does not require pedigree analysis. We establish the effect sample preparation has on extending the limits of this approach. We also show that this approach can be used to map tagging SNPs within adjacent haplotype blocks, allowing for the spanning of unstable hot-spot regions known to frequently undergo homologous recombination. Lastly, we demonstrate this technique can effectively determine whether 2 copy genes are arranged in cis or trans, a useful technique to identify disease carriers. Such a technique will prove valuable in validating haplotypes derived from whole-genome association studies.

687T

Deep Mutational Scanning to Assess How Mutations Impact Protein Aggregation. *D.M. Fowler.* Genome Sciences, University of Washington, Seattle, WA.

Protein aggregation plays a critical role in many common diseases including Alzheimer's and Parkinson's, each of which afflicts tens of millions of individuals. These diseases remain largely intractable, with grim, inexorable clinical courses. The protein aggregates found in these diseases are called amyloid and have a highly organized, fibrous structure known as the cross-β sheet. The exact role of amyloid formation in amyloid diseases is under active debate, but the importance of amyloid is highlighted by the existence of early onset, familial forms of these diseases in which mutations increase the aggregation propensity of the offending protein. To better understand protein aggregation, we are leveraging deep mutational scanning, a method that uses high-throughput DNA sequencing to enable the functional characterization of hundreds of thousands of mutants of a protein simultaneously. To generate these large-scale mutagenesis data for aggregation-prone proteins, we are using an assay that links growth of the budding yeast S. cerevisiae to the aggregation of a plasmid-encoded protein expressed as a dihydrofolate reductase (DHFR) fusion. Aggregation of the fusion protein inactivates DHFR and results in a reduced growth rate in the presence of methotrexate, a DHFR inhibitor. We have demonstrated differential growth of yeast carrying aggregation-enhancing versus aggregation-reducing mutations of both Aβ, which aggregates in Alzheimer's disease, and α-synuclein, which aggregates in Parkinson's disease. Screening of libraries of mutants of each of these proteins will yield aggregation propensity maps, detailing the impact of nearly all possible single mutations and many double mutations on aggregation. These maps will reveal the underlying physicochemical features responsible for amyloidogenesis in these proteins, yielding insight into the mechanism of aggregation and the structure of the aggregates in vivo. Finally, because these data describe the aggregation propensity of each possible single mutation in these aggregation-prone proteins, they could be used to interpret the significance of variants found in individuals' genomes.

688F

Allele level sequencing and phasing of full-length HLA class I and II genes using SMRT® sequencing technology. *S. Ranade[1], J. Chin[1], B. Bowman[1], K. Eng[1], S. Suzuki[2], Y. Ozaki[2], T. Shiina[2].* 1) Pacific Biosciences, Menlo Park, CA; 2) Department of Molecular Life Sciences, Tokai University School of Medicine, Isehara, Kanagawa, Japan.

MHC gene family is comprised of three classes of genes, which are instrumental in determining donor-recipient compatibility for organ transplant as well as susceptibility to autoimmune diseases via cross-reacting immunization. Specifically Class I HLA-A, -B, -C, and Class II HLA-DR, -DQ and -DP genes are considered medically important for genetic investigation to determine histocompatibility. Highly polymorphic in nature, thousands of alleles of these HLA genes have been implicated in disease resistance and susceptibility, and the importance of full-length HLA gene sequencing for genotyping, detection of null alleles and phasing is being widely acknowledged. Although DNA-sequencing-based HLA genotyping is now routinely done, only 7% of the HLA genes have been characterized by allele-level sequencing, while 93% are still defined by partial sequences only. The gold-standard Sanger sequencing technology, though widely used, is unable to generate phased reads from heterozygous alleles. Second-generation, high-throughput clonal sequencing methods, although better at heterozygous allele detection, are inadequate at generating full-length haploid gene sequences from enhancer promoter to 3'UTR along with phasing information. The best way to overcome these challenges is to sequence these genes with a technology that is clonal in nature and has the longest possible read lengths. We have employed Single Molecule Real-Time (SMRT®) sequencing technology from Pacific Biosciences for sequencing full-length HLA class I and II genes. PCR systems were developed to amplify entire HLA genes ranging between 4.6 kb and 9.7 kb. For HLA genes that were too long, like HLA-DRB1 and -DPB1 (13–21 kb), the PCR regions were divided into two parts (enhancer-promoter to exon 2, and exon 2 to 3'-UTR for HLA-DRB1; and enhancer-promoter to intron 2, and intron 1 to 3'-UTR for HLA-DPB1). The long-range PCR amplicons were converted into SMRTbell™ libraries, and long reads ranging from ~3,500 to 20,000 bases were generated on the PacBio® RS. Full-length consensus reads were obtained with correct phasing information and accurate genotypes for all the alleles of each of the HLA genes based on the zygosity and presence of pseudo genes. The resolving power of SMRT Sequencing for simultaneously genotyping, determination of phasing information, and detection of novel alleles makes it a uniquely suitable DNA-based method for unambiguous allele-level characterization of HLA genes.

689W

The Role Of MDR1 C3435T and C1236T Single Nucleotide Polymorphisms in Male Infertility. *S. Aydos[1], A. Karadag[1], T. Ozkan[1], B. Altinok[1], M. Bunsuz[1], S. Heidargholizadeh[1], K. Aydos[2], A. Sunguroglu[1].* 1) Ankara University, Faculty of Medicine, Department of Medical Biology Ankara, Turkey; 2) Ankara University, Faculty of Medicine, Reproductive Health Research Center Ankara, Turkey.

Infertility is a common problem, affecting one in six couples.In 30% of infertile couples, the male factor is a major cause due to defective sperm quality.However, the factors responsible for defective sperm quality remain largely unknown.The multi-drug resistance1 (MDR1) gene encoding a Pglycoprotein, which has a role in active transport of various substrates, including xenobiotics, and thus has a protective function in various tissues and organs. In the present case-control study, we investigated the effect of MDR1 gene C3435T and C1236T SNPs and on male infertility in Turkish population.The study was performed on 192 patients with infertil and 102 healthy control.The genotyping of C3435T and C1236T SNPs was done by PCR-RFLP.In statistical analysis were tested with chisquare test and SHEsis program. 101 patients were compared with sperm parameters. Frequencies of the C and T alleles of the 1236 locus were found to be 61.2(n=235) and 38.8(n=149)% in the infertil patient, similar to the 59.8 and 40.2% found in the healthy control, respectively.The C allele was detected more frequently in the patient.The frequencies of MDR1 1236 CC,CT and TT genotypes were 34.4%,53.6%,12% in the patient; those in the control were 34.3%,51%,14.7%, consecutively. However, the observed genotype frequencies did not show significant difference in either group(P>0.05).Frequencies of the C and T alleles of the 3435 locus were found to be 56.2, 43.8 % in the infertil patient, similar to the 51%, 49% found in the healthy control, respectively.The frequencies of MDR1 3435 CC,CT and TT genotypes were 27.6%,57.3%,15.1% in the patient; those in the control were 26.5%, 49%,24.5%, consecutively.However,the observed allel and genotype frequencies did not show significant difference in either group (P>0.05).Each of the four possible haplotypes was noted in both infertil patient and control. When the frequency distributions of estimated haplotypes were compared between the patient and control, the frequency of the T-T haplotype was found to be significantly higher in the control than in infertil patients (P=0.006).There was no correlation between sperm parameters and genotypes.Our findings show that two SNPs do not play a role in the genetic susceptibility to male infertility, controversial to Polish population.But T-T haplotypes may be a protective factor for the fertility,because this haplotype was found to be statistically lower in infertile patients than in control.

690T

Interactions between Epigenetic and Genetic Signatures in the Asthmatic Airway. *J. Nicodemus-Johnson[1,5], R.A. Myers[1,5], D.K. Hogarth[2], J. Sudi[2], J.F. McConville[2], E.T. Naureckas[2], A.I. Sperling[2], J. Solway[2], J.A. Krishnan[2,4], S.R. White[2], D.L. Nicolae[1,2,3], Y. Gilad[1], C. Ober[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL; 4) Current Address: Department of Medicine, University of Illinois Hospital & Health Sciences System, Chicago IL; 5) co-first author.

A large number of genes are differentially expressed in airway epithelial cells of asthmatic and non-asthmatic individuals, yet neither the epigenetic mechanisms underlying these difference nor epigenetic interactions with asthma-risk alleles have previously been investigated. To address this gap in knowledge, we obtained genome-wide methylation profiles (450K array), gene expression by RNA sequencing and genetic variation (SNPs) in freshly isolated airway epithelial cells from 26 asthmatic and 20 non-asthmatic subjects. We identified differentially methylated regions (DMRs) and characterized their locations with respect to nearby genes, and assessed the influence of SNPs on nearby CpG methylation levels and gene expression. We identified 276 DMRs at a false discovery rate (FDR) of 5%. DMRs near genes are enriched for asthma relevant pathways in the airways such as focal adhesion, cytokine-cytokine receptor interactions, and calcium signaling (FDR < 5%). We identified methylation quantitative trait loci (meQTLs) for nearly half of all DMRs, some of which were associated with asthma in published genome-wide association studies. For example, we identified a significant interaction effect between the asthma-associated TSLP promoter SNP (rs1837253) and asthma status on methylation patterns: the asthma risk allele (T) is associated with decreased methylation in asthmatics and increased methylation in non-asthmatics (interaction p=0.010). In contrast, an asthma associated SNP at the HLA-DQ locus was a meQTL in both asthmatics ($p=5.6\times10^{-9}$) and non-asthmatics ($p=1.4\times10^{-10}$). However, the same SNP was associated with HLA-DQB1 transcript abundance in asthmatic subjects only (interaction p=0.027). These examples provide evidence for two distinct functional mechanisms by which previously reported genetic associations with asthma can be explained by variation in methylation patterns. Collectively, these data highlight the importance of epigenetic variation in promoting asthma phenotypes in the lung, both independently and in conjunction with local genetic variation. This work was supported by NIH grant AI095230.

691F

Investigation of Cytotoxic T-lymphocyte-associated Protein 4 Gene Polymorphisms in Systemic Lupus Erythematosus. *H. Yang[1,2], T. Chen[3], C. Lee[1,2,4], S. Chang[1], W. Chen[1], W. Lin[1], C. Lin[1], Y. Lee[1,5,6,7,8].* 1) Dept. of Medical Research, Mackay Memorial Hospital, New Taipei City, Taiwan; 2) Department of Nursing, Mackay Medicine, Nursing and Management College, Taipei, Taiwan; 3) Dept. of Rheumatology, Mackay Memorial Hospital, Taipei, Taiwan; 4) Department of Internal Medicine, Mackay Memorial Hospital, Taipei, Taiwan; 5) Dept. of Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan; 6) Institute of Biomedical Science, Mackay Medical College, New Taipei City, Taiwan; 7) Dept. of Medicine, Mackay Medical College, New Taipei City, Taiwan; 8) Dept. of Pediatrics, Taipei Medical University, Taipei, Taiwan.

Systemic Lupus Erythematosus (SLE) is a chronic inflammatory, multisystemic, and autoimmune-related disease. The prevalence of SLE in Taiwan is around 37.0/100.000 persons and associated with gender, especially in childbearing age. SLE characterized by T and B cell hyperactivation, autoantibody production, and immune complex deposition. Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) is expressed on CD4+ and CD8+ activated both T and B cells. Protein expression of CTLA4 was increased in T cells from patients with SLE. Numerous studies have been found in association between CTLA4 polymorphisms and SLE in different ethnic groups. However, large-scale study of CTLA4 polymorphisms in SLE is still elusive in Taiwan. For this case-control cross-sectional study among Taiwanese, 283 patients with SLE and 920 controls were enrolled. Genotyping of -318 C/T, +49 A/G, and CT60 A/G single nucleotide polymorphisms (SNPs) was performed by PCR-RFLP and TaqMan assays. The genotype, allele, and haplotype frequencies were calculated by direct counting or with Haploview 4.2 software. Genotype, allele, and haplotype frequencies of the CTLA4 SNPs studied were equally distributed in SLE patients and controls. No significant associations between SLE and these 3 SNPs were observed. Our data suggest that CTLA4 -318 C/T, +49 A/G, and CT60 A/G SNPs do not confer increased susceptibility to SLE.

692W

Large scale identification of alternative polyadenylation through next generation sequencing. *C.M. Lopes-Ramos[1,2], A.A. Camargo[1], P.A.F. Galante[1], R.B. Parmigiani[1].* 1) Molecular Oncology Center, IEP, Hospital Sírio-Libanes, São Paulo, Brazil; 2) Fundação Antonio Prudente, AC Camargo Cancer Center, São Paulo, Brazil.

Most eukaryotic mRNAs acquire an uncoded polyA tail at their 3' ends during maturation in a process called polyadenylation (polyA). This process involves the cleavage of mRNA at a specific site and the subsequent addition of adenosine residues. Cleavage position is defined by recognition of specific sequences, called polyA signals. More than half of mammalian genes have multiple polyA signals, which can lead to the formation of transcript variants with different 3' untranslated regions (UTRs) or even different coding regions. Alternative polyA may influence the location, stability and transport of transcripts, in tissue or disease-specific manner. This is a result of many cis-acting elements involved in post-transcriptional regulation located within the 3'UTRs, such as miRNA binding sites. Here we studied alternative polyA events in two colon cancer cell lines (HCT116 and SW480) through next generation sequencing. We developed an original cDNA library protocol to enrich for the transcripts' 3' ends, enabling the preferential sequencing of the 3'UTRs and identification of polyA sites. Using a SOLiD sequencer (Life Technologies), more than 50 million sequences were generated for each cell line. Alternative polyA generating transcripts shorter than the reference gene (RefSeq with the longest 3'UTR) was verified for more than 6000 genes, corresponding to 30% of genes expressed in these cells. Besides the above-mentioned effects, evidences of alternative polyA functional impact may be demonstrated by the loss of miRNA target sites. In fact, 2400 polyA variants have lost at least one conserved miRNA target site, which represent about 30% of all identified variants. Preferential usage of shorter polyA variants allows escaping from the inhibitory regulation of miRNAs on the expression of such target genes. We have also evaluated miRNA expression profile of these two cell lines and identified more than 50 miRNAs being differently expressed. miRNA expression profile and polyA variants expression might be tightly associated. Therefore, identification of inverse correlation of miRNA and polyA variants expression is underway. This innovative approach increased our capacity of identification and expression analysis of polyA variants for several genes. By focusing on variants that lose binding sites for miRNAs, it may rise possibilities to better understand how this phenomenon regulates different cellular processes in physiological and pathological situations.

693T

Identification of DNA editing in retrotransposons of diverse genomes.
B.A. Knisbacher, E.Y. Levanon. The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan , Israel.

Retrotransposons comprise a large fraction of mammalian genomes. Their dynamics contribute to genomic plasticity and enhance evolution. Yet, retrotransposons must be kept in check to retain genomic stability and avoid detrimental mutagenesis. The APOBECs, a family of cytidine deaminases, take that responsibility and restrict mobilization of retrotransposons in the genome. Through their ability to bind and edit DNA they can cause deleterious hypermutation in nascent retrotransposon DNA, right after reverse transcription, which will typically cause its degradation. However, in some cases, the retrotransposons can complete mobilization despite being hypermutated. Such an event results in the insertion of a unique retrotransposon sequence, thus increasing genomic diversity and the probability of developing a novel functional unit at this genomic locus. In this study, we computationally screened retrotransposon sequences in >80 genomes (the UCSC reference genomes), for edited integrants. By generating pairwise alignments of retrotransposons we revealed that DNA editing is abundant in many lineages and especially in mammals. In total, these genomes contain tens of thousands of edited sequences harboring hundreds of thousands of edited sites. Our analyses of the edited elements and editing rates, including comparative genomics, gives valuable insights into the evolutionary impact of DNA editing and the role of retrotransposon restrictors in distinct species.

694F

Association between estrogen receptor and intraocular pressure in mice.
F. Mabuchi[1], R. Yamagishi[2], K. Kashiwagi[1], M. Aihara[3]. 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Ophthalmology, Univ Tokyo, Bunkyo-ku, Tokyo, Japan; 3) Shirato eye clinic, Shinjuku-ku, Tokyo, Japan.

Purpose: To evaluate intraocular pressure (IOP) in transgenic mice with a knockout of estrogen receptor alpha (*ESR1*) or estrogen receptor beta (*ESR2*) gene. **Methods:** Heterozygous B6.129P2-*Esr1*^{tm1Ksk}/J (*ESR1* gene knockout) or B6.129P2-*Esr2*^{tm1Unc}/J (*ESR2* gene knockout) mice, and corresponding wild-type mice were anesthetized. A fluid-filled glass microneedle connected to a pressure transducer was then inserted through the cornea into the anterior chamber to measure IOP. All measurements were made between 9:00 PM and 11:00 PM. The IOP of fifteen *ESR1* $^{+/-}$ and wild-type *ESR1* $^{+/+}$ mice, and IOP of nine *ESR2* $^{+/-}$ and wild-type *ESR2* $^{+/+}$ mice were measured at 6 to 8 weeks after birth. **Results:** The IOPs of the *ESR1* $^{+/-}$ and wild-type *ESR1* $^{+/+}$ mice were 20.5 ± 1.0 mmHg (mean ± standard deviation) and 20.3 ± 1.2 mmHg respectively, and there was no statistically significant difference (P = 0.81, Mann-Whitney U test) between them. In contrast, the IOPs of the *ESR2* $^{+/-}$ and wild-type *ESR2* $^{+/+}$ mice were 18.3 ± 1.2 and 19.9 ± 1.6 mmHg respectively, and there was a statistically significant difference (P = 0.04, Mann-Whitney U test) between them. **Conclusion:** These results suggest that the *ESR2* is associated with IOP regulation.

695W

CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. *E.J. Hollox[1], L. Odenthal-Hesse[1], J. Bowdrey[1], A. Habtewold[2,3], E. Ngaimisi[2,4], G. Yimer[2,3], W. Amogne[5,6], S. Mugusi[7], O. Minzi[4], E. Makonnen[3], M. Janabi[8], F. Mugusi[8], G. Aderaye[5], R. Hardwick[1], B. Fu[9], M. Viskaduraki[10], F. Yang[9], E. Aklillu[2].* 1) Department of Genetics, University of Leicester, Leicester, United Kingdom; 2) Department of Clinical Pharmacology, Karolinska Institutet, Stockholm, Sweden; 3) Department of Pharmacology, Addis Ababa University, Addis Ababa, Ethiopia; 4) Unit of Pharmacology, School of Pharmacy, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 5) Internal Medicine, Addis Ababa University, Addis Ababa, Ethiopia; 6) Institution of Medicine, Unit of Infectious Diseases, Karolinska Institutet, Karolinska University Hospital, Huddinge, Sweden; 7) Department of Internal Medicine, Muhimbili National Hospital, Dar es Salaam, Tanzania; 8) Department of Internal Medicine, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 9) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 10) College of Medicine, Biological Sciences and Psychology, University of Leicester, Leicester, United Kingdom.

The role of genome copy number variation of the CCL3L1 gene, encoding MIP1α, in contributing to the host variation in susceptibility and response to HIV infection is controversial. Here we analyse a sub-Saharan African cohort from Tanzania and Ethiopia, two countries with a high prevalence of HIV-1 and a high co-morbidity of HIV with tuberculosis. We use a form of quantitative PCR called the paralogue ratio test, we determine CCL3L1 gene copy number in 1134 individuals and validate our copy number typing using array comparative genomic hybridisation and fiber-FISH. We find no significant association of CCL3L1 gene copy number with HIV load in antiretroviral-naïve patients prior to initiation of combination highly active anti-retroviral therapy. However, we find a significant association of low CCL3L1 gene copy number with improved immune reconstitution following initiation of highly active anti-retroviral therapy (p=0.012), replicating a previous study. Our work supports a role for CCL3L1 copy number in immune constitution following antiretroviral therapy in HIV, and suggests that the MIP1α-CCR5 axis might be targeted to aid immune reconstitution.

696W

Role of translocator protein (TSPO) in antipsychotic-induced weight gain. *J.G. Pouget[1,2], V.F. Gonçalves[1], A.K. Tiwari[1], J.A. Lieberman[3], H.Y. Meltzer[4], D.J. Müller[1,5], J.L. Kennedy[1,5].* 1) Pharmacogenetics Research Clinic, Neuroscience Department, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada; 3) Department of Psychiatry, College of Physicians and Surgeons, Columbia University and the New York State Psychiatric Institute, New York City, NY, USA; 4) Northwestern University Feinberg School of Medicine, Chicago, IL, USA; 5) Department of Psychiatry, University of Toronto, Toronto, ON, Canada.

Background: Weight gain is a common side effect of second-generation antipsychotics (SGAs), contributing to patient non-compliance and cardiovascular disease. With heritability estimates ~60%, genetic markers may improve prediction of patients at greatest risk of AIWG. Translocator protein (TSPO), part of the mitochondrial permeability transition pore (mtPTP), regulates cytosolic reactive oxygen species (ROS) levels and shows enhanced functioning after SGA treatment. The present study explores the association between genetic variation in TSPO and AIWG. **Methods:** Genomic DNA was obtained from blood samples of schizophrenia patients with weight change observed after up to 14 weeks of antipsychotic treatment (n=235). Eight putative functional SNPs in the TSPO region (rs138926, rs80411, rs6971, rs6973, rs113515, rs138911, rs5759197, rs739092) were genotyped using TaqMan assays (Applied Biosystems), capturing 77% of common variation across TSPO. Pairwise interactions of TSPO markers significantly associated with % weight change from baseline and markers in other mtPTP genes (SLC25A4: rs10024068, rs7660552; VDAC1: rs13169435, rs4279383, rs2288834, rs2066944, rs10491289; HK1: rs16926246, rs7072268) were assessed using model-based multifactor dimensionality reduction. **Results:** No significant association with % weight change was observed for any TSPO SNPs in the total sample. In a sub-analysis of European patients on clozapine or olanzapine (n=82), drugs with the greatest propensity for weight gain, nominal association was observed between % weight change and the rs6971 Thr allele after multiple testing correction (p=0.02). rs6971 Thr/Thr homozygotes showed 4.65% greater weight change from baseline (95% CI: 0.03 - 9.26%, p=0.05) compared to Ala/Ala homozygotes. A significant interaction was observed between TSPO rs6971 and ANT1 rs10024068 ($P_{10\ 000}$=8.0×10^{-3}). **Discussion:** The rs6971 Thr allele may predispose European patients treated with clozapine or olanzapine to AIWG. Previous functional studies indicate rs6971 predicts a large portion of variance in TSPO binding, with the Thr allele conferring lower binding affinity. TSPO rs6971 may interact with rs10024068, a marker 1.1kb upstream of SLC25A4, to influence AIWG. These findings provide preliminary support for the role of the mtPTP in AIWG. Replication, in addition to functional studies clarifying the underlying mechanisms, is required to elucidate the role of these variants in AIWG.

697W

Association between polymorphisms in genes coding for cellular drug transporters and the effectiveness of antiretroviral therapy against HIV-1 among Brazilians. *C.C. Cardoso, J.Z.C. Dias, M.B. Arruda, R.M. Brindeiro, O. Ferreira, A. Tanuri, R.S. Aguiar.* Genetics Department, Biology Institute, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Background: The highly active antiretroviral therapy (HAART) improved the life expectation and prognostic of the HIV+ patients. However, about 10-20% of patients develop HAART failure, mostly due to the emergence of virus mutations associated with drug resistance. Moreover, human genetics polymorphism can also be associated with antiretroviral failure, mainly in genes related to drug absorption, activation and metabolism. Here we investigate the association between 40 single nucleotide polymorphisms (SNPs) in the genes SLC22A1, SLC22A2, SLC22A3 (organic cation ransporters 1-3) and ABCB1 (P-Glycoprotein) and HAART effectiveness. Methods: We have conducted a case-control study including 237 HIV+ individuals (117 cases of HAART failure and 120 controls), from Curitiba and Porto Alegre cities (Brazil). The case definition was based on the reduced HIV-1 viral load to undetectable levels in the first-line therapy. All patients were treated for at least 6 months. Human SNPs were genotyped using SnapShot, TaqMan and Sanger sequencing. All samples were also genotyped for additional 32 ancestry informative SNPs. Pairwise linkage disequilibrium patterns were estimated by the r2 coefficient. The frequencies of each SNP and haplotypes in cases and controls were compared by logistic regression models adjusted for the covariates age, genetic ancestry and therapy scheme. Results: Logistic regression models showed an association between the SNPs rs2229109 and rs6961419 in ABCB1 gene and therapeutic failure. The presence of allele A for the SNP rs2229109 was associated with an increasing risk of HAART failure with OR values of 7.33 (95% CI: 1.85-29; p = 0.001). Moreover, patients sharing alleles rs2229109A/rs6961419G were more susceptible to develop HAART failure (OR=8.04; 95% CI: 2.02-32; p = 0.003). Our results are in concordance with functional data that associated the substitution S400N codified by this SNP (rs2229109) with higher extrusion levels of drugs mediated by P-Glycoprotein. Conclusion: Our data describe a clear association between ABCB1 gene and HAART failure, suggesting that polymorphisms in this gene may affect intracellular drug concentrations and influence treatment effectiveness. Studies of host genetics are crucial to predict more effective treatment for HIV+ patients.

698W

Differential effects at candidate SNP loci on reduction in plasma LDL cholesterol, apolipoprotein B, LDL particle number, and mean LDL particle size with statin therapy.
A.Y. Chu[1], B.J. Barratt[2], B. Ding[3], F. Nyberg[3,4], S. Mora[1,5], P.M Ridker[1,5], D.I. Chasman[1,6]. 1) Division of Preventive Medicine, Brigham & Women's Hospital, Boston, MA USA; 2) Personalized Healthcare and Biomarkers, AstraZeneca Research and Development, Alderley Park, United Kingdom; 3) Global Epidemiology, AstraZeneca Research and Development, Molndal, Sweden; 4) Unit of Occupational and Environmental Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 5) Division of Cardiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA USA; 6) Division of Genetics, Brigham and Women's Hosp, Boston, MA USA.

Statins differentially reduce LDL cholesterol (LDL-C), apolipoprotein B (apoB), total number of LDL particles (LDL-P) and mean LDL particle size (LZ). Variations in the reduction of these measures and possibly their relationship to cardiovascular risk reduction may be under genetic control. We performed a candidate SNP analysis for association with change in LDL-C, apoB, LDL-P and LZ after 12 months of statin therapy, focusing on 156 SNPs previously associated with either statin-induced reduction in LDL-C (6 SNPs) or untreated serum lipid levels from metabochip analysis by the Global Lipids Consortium (150 SNPs). The study population included 3,534 statin-allocated participants of European ancestry in JUPITER (NCT00239681), a randomized, placebo-controlled trial of rosuvastatin (20 mg/d) for primary prevention of cardiovascular disease. Residuals for all LDL change measures were calculated adjusted for age, sex, region and population stratification and further transformed by inverse-quantile normalization. Genetic associations were assessed by linear regression applying the standard additive assumption for estimating effects on quantile normalized lipoprotein measures (β_q). Of the 156 candidate SNPs, 42 were nominally associated with at least one of the four LDL-change measures ($p<0.05$) and could be clustered into 4 groups depending on whether the association was primarily with change in LDL-C, apoB, LDL-P or LZ. A Bonferroni corrected threshold of $p<8.3e-5$ (=0.05/156x4 tests) identified novel associations for variants at *APOA1-A5* for change in LZ (rs964184; β_q=0.17, p= 4.1e−5), and at *CILP2* for change in apoB (rs10401969; β_q=0.17, p=7.2e−5). The *APOA1-A5* SNP was associated only with change in LZ (associations with LDL-C, apoB and LDL-P p>0.05; all β_q different from LZ β_q, p<0.05 for all three comparisons). The SNP at *CILP2* was primarily associated with change in apoB, but also was nominally associated with change in LDL-P (β_q=0.13, p=8.0e-3; pairwise comparison p>0.05). Additionally, variants in *ABCG2* and *LPA*, previously identified for change in LDL-C in JUPITER (Circ Genetics 2012;5:257) were associated with change in apoB and LDL-P. None of these SNPs were associated with change in the four LDL-associated measures in the placebo arm (n=3512). Differential genetic effects of statin-induced change in LDL-C, apoB, LDL-P and LZ based on candidate SNP analysis suggest distinct pathways underlying the change of these LDL-related measures with statins.

699W

A hypothesis driven association analysis of nuclear-encoded mitochondrial genes with antipsychotic-induced weight gain in schizophrenia subjects.
V.F. Gonçalves[1, 2], C. Zai[1, 2], A. Tiwari[1], A. Derkach[3], H. Meltzer[4], J. Lieberman[5], A.D. Paterson[6], L. Sun[3, 7], D. Mueller[1, 2], J.L. Kennedy[1, 2]. 1) Department of Psychiatry, University of Toronto, Toronto, Canada; 2) Neuroscience Section, Centre for Addiction and Mental Health, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Toronto, Canada; 4) Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, USA; 5) Department of Psychiatry, College of Physicians and Surgeons, Columbia University and the New York State Psychiatric Institute, New York City, USA; 6) Program in Genetics and Genomic Biology, Hospital for Sick Children, Toronto, Canada; 7) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

Antipsychotic-induced weight gain (AIWG) is an important phenotype that often leads to obesity and metabolic syndrome. We are proposing that mitochondrial system might play a role in its development since mitochondria is main source of energy for neurons and control energy homeostasis in the hypothalamus. In our study, we investigated the hypothesis that nuclear-encoded mitochondrial genes, particularly those with altered gene expression or involved in the oxidative stress, mitochondrial biogenesis, inflammation and apoptosis, would be associated with AIWG. In total, we selected 28 genes and analyzed 60 SNPs, most of them classified as functional or regulatory elements, in schizophrenia subjects (N=164), treated with atypical medications up to 14 weeks. Single-SNP genetic association was tested using linear regression with percentage of weight gain from baseline as the dependent variable and treatment duration and baseline body weight as covariates. Assuming MAF of 0.15, we had more than 80% power to detect a mean difference of 2.4% between carriers and non-carriers of the risk genotype in the additive model. Several multi-SNP analyses were also carried out, including haplotype analysis, stepwise linear regression, and gene-gene interactions. The statistical strength of our biological hypothesis was measured by comparing the sum of the observed association evidence across all 60 SNPs with the value expected under the null based on a phenotype permutation method (10,000 permutation replicates). We observed a significant association between rs6435326 in the NDUFS1 gene and percentage of weight gain (N= 150, Pcorrected=0.02). The haplotype carrying the risk alleles for rs6435326 and two other SNPs in NDUFS1 was also significantly associated with weight-gain (%) (P=0.005). In addition, we observed a significant interaction between the TT risk genotype of rs6435326 in NDUFS1 and AG genotype of rs3762883 in COX18 (Pcorrected=0.001). Finally, permutation-based test showed that the set of 60 SNPs from the 28 nuclear-encoded mitochondrial genes selected based on our hypothesis, collectively, was associated with weight gain (P=0.02). To the best of our knowledge, this is the first study to explore genetic variation in the mitochondrial genes in the context of AIWG. This study provides evidence implicating mitochondrial genes to be involved in the regulation of the energy homeostasis and body weight in schizophrenia subjects under atypical antipsychotic treatment.

700W

PAX4 genetic variant is associated with therapeutic effect of repaglinide in the Chinese type 2 diabetes patients.
C. Hu, M. Chen, R. Zhang, Y. Bao, K. Xiang, W. Jia. Shanghai Diabetes Inst, Shanghai Jiao Tong University, Shanghai, Shanghai, China.

A recent genome-wide association study (GWAS) identified a novel diabetes-associated locus near PAX4 (rs10229583) in Chinese population and other populations. However, whether this single-nucleotide polymorphism (SNP) influence the therapeutic effects of oral antihyperglycemic drugs has not been reported. The aim of this study was to investigate the association of PAX4 variants rs10229583 with therapeutic effects of repaglinide or rosiglitazone in patients with type 2 diabetes. A total of 209 newly diagnosed type 2 diabetes patients were recruited and treated with repaglinide or rosiglitazone randomly for 48 weeks (104 and 105 patients, respectively). In the repaglinide cohort, individuals who were GG homozygotes of rs102229683 showed lower fasting plasma glucose , 2h glucose and HbA1c levels than the A allele carriers. Moreover, carriers of GG genotype exhibited significantly higher cumulative attainment rates of target fasting plasma levels (Plog-rank = 0.0009) than the A allele carriers, and this difference was still significant (P = 0.0037) even after adjusting for age, gender, and body mass index (BMI) at baseline. Besides, significant linear relationships were detected between the number of rs10229583 A alleles and increments in fasting insulin levels (P= 0.0426, adjusted for age, gender, dosage, and BMI at baseline). However, no effect of rs10229583 on the efficacy of rosiglitazone was found in our study. In conclusion, the PAX4 variant rs10229583 was associated with therapeutic effect of repaglinide in Chinese patients with type 2 diabetes.

701W

Multi-ethnic cytochrome P450-2D6 (CYP2D6) allele frequency profiling and full gene single-molecule real-time (SMRT®) sequencing. *S. Martis[1], Y. Yang[1], A. Gaedigk[2], R. Sebra[3], R. Vijzelaar[1], R. Kornreich[1], R.J. Desnick[1], S.A. Scott[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Division of Clinical Pharmacology & Therapeutic Innovation, University of Missouri-Kansas City, Kansas City, MO; 3) MRC Holland, Willem Schoutenstraat 6, Amsterdam, The Netherlands.

The polymorphic cytochrome P450-2D6 (CYP2D6) isoenzyme is involved in the oxidative metabolism of approximately 25% of commonly used medications, underscoring its importance in human drug metabolism. Common variant CYP2D6 alleles have been implicated in interindividual drug response variability and adverse reactions; however, the CYP2D6 gene is difficult to interrogate due to high sequence homology with its neighboring CYP2D7P1 pseudogene. To determine the multi-ethnic frequencies of an extensive panel of CYP2D6 variant alleles, *2 - *12, *14, *15, *17, *29, *35, *41, and the gene duplication were genotyped among 464 healthy adult African-American, Asian, Caucasian, Hispanic, Ashkenazi, and Sephardic Jewish individuals using the xTAG CYP2D6 Kit (Luminex Molecular Diagnostics). The frequencies of functional, reduced function, and non-functional alleles in the tested populations ranged from 0.46-0.76, 0.11-0.43, and 0.11-0.21, respectively. In addition, multiplex ligation-dependent probe amplification (MLPA) with probes for CYP2D6 exons 1, 4, 6 and 9 was performed on all samples, which validated the deletion (*5) and duplication genotyping. However, discrepant copy number results between genotyping and MLPA were observed predominantly in the Asian population, which likely was indicative of the CYP2D6*36-*10 tandem allele as it correlated with genotype detection of *10 and the increased MLPA-detected copy number in these samples was restricted to the exon 1, 4, and 6 probes. Additionally, in an effort to devise a unique full gene sequencing strategy that could identify novel CYP2D6 alleles, establish the phase of genotyped variants, and specifically characterize the duplicated CYP2D6 copy when present, long-read third-generation single-molecule real-time (SMRT®) sequencing (Pacific Biosciences) of 5.0 kb CYP2D6 full gene amplicons was performed on 12 publically available DNA samples with previously reported discrepant CYP2D6 genotypes. Importantly, in contrast to Sanger and short-read second-generation sequencing, SMRT® sequencing can determine the phase of CYP2D6 diplotypes in these samples, including the specific identity of the duplicated copies when present. Taken together, these data identify the broad allelic spectrum of the CYP2D6 gene in a racially diverse and multiethnic population and support the use of long-range SMRT® sequencing for full gene and phased CYP2D6 characterization, including novel allele discovery.

702W

Association of Cyclooxygenase-2 genetic variant with cardiovascular disease. *S.A. Ross[1], J.W. Eikelboom[1], S.S. Anand[1], N. Eriksson[2], H. Gerstein[1,3,4], S.R. Mehta[1], S.J. Connolly[1], L. Rose[5], P.M. Ridker[5,6], L. Wallentin[2], D. Chasman[5,6], S. Yusuf[1], G. Pare[1].* 1) Population Health Research Institute, Hamilton Health Sciences and Departments of Medicine, Epidemiology, Pathology McMaster University, Hamilton, ON, Canada; 2) Uppsala Clinical Research Center and Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Department of Medicine, McMaster University, Hamilton, ON, Canada; 4) Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; 5) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 6) Harvard Medical School, Boston, Massachusetts, USA.

Background: Cyclooxygenase (COX) enzymes convert arachidonic acid to prostaglandin H2 leading to the production of biologically active prostanoids, prostacyclin and thromboxane, which are believed to play key roles in atherothrombosis. The non-selective COX inhibitor aspirin is associated with a decreased risk of cardiovascular disease (CVD) whereas selective COX-2 inhibitors have been associated with an increased risk. A genetic variant (rs20417) of the PTGS2 gene, encoding for COX-2, has been associated with decreased COX-2 activity and a decreased CVD risk. However, this genetic association and the role of COX-2 in CVD remains controversial. Methods: The association between rs20417 and major adverse cardiovascular outcomes (non-fatal myocardial infarction, stroke or cardiovascular death) was prospectively explored in 49,233 subjects (ACTIVE-A, CURE, epiDREAM/DREAM, ONTARGET, RE-LY, and WGHS) and the effect of potentially modifiable risk factors on the genetic association was further explored in 4,465 non-fatal myocardial infarction cases and 4,898 controls from the INTERHEART study. We also examined the effects of rs20417 on urinary thromboxane and prostacyclin metabolite concentrations in 119 healthy individuals. Results: Carriage of the rs20417 minor allele was associated with a decreased risk of major CVD outcomes (OR=0.78, 95% CI: 0.70 - 0.87; $P=1.2\times10^{-5}$). The genetic effect was significantly stronger in aspirin users (OR: 0.74, 95% CI: 0.64-0.84; $P=1.20\times10^{-5}$, N=22,441) than non-users (OR: 0.87, 95% CI: 0.72-1.06; P=0.16, N=26,719) with an interaction p-value of 0.0041. Carriers had significantly lower urinary levels of thromboxane (97.0 vs. 125.5 ng/mmol creatinine; P=0.02) and prostacyclin (3336.0 vs. 4702.0 ng/mmol creatinine; P=0.01) metabolites as compared to noncarriers. Conclusion: The rs20417 polymorphism is associated with a reduced risk of major cardiovascular events and lower levels of thromboxane and prostacyclin. Our results suggest that a genetic decrease in COX-2 activity may be beneficial with respect to CVD risk and may vary in effect size depending on ASA use.

703W

Influence of CYP4F2 on dose, anticoagulation control and risk of hemorrhage among African American and European American warfarin users. *A. Shendre[1], T.M. Beasley[2], D.A. Nickerson[3], N.A. Limdi[4, 1].* 1) Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 2) Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, AL; 3) Genome Sciences, School of Medicine, University of Washington, Seattle, WA; 4) Neurology, University of Alabama at Birmingham, Birmingham, AL.

The management of warfarin therapy is complicated by a wide variation in dose and response across patients. A significant portion of variability in dose is explained by single nucleotide polymorphisms (SNPs) in CYP2C9 and VKORC1 across racial groups. CYP4F2 (rs2108622; p.V433M) explains variability in dose among European Americans but its influence on warfarin dose, anticoagulation control, and risk of hemorrhage has yet to be fully examined, especially in African Americans. Herein we assess the influence of CYP4F2 on stable warfarin dose, anticoagulation control (percent time in target range; PTTR) and risk of major hemorrhage in 649 European and 430 African Americans patients in a prospective warfarin pharmacogenetics study. We evaluated the influence of the CYP2F2 genotypes on the 3 outcomes of interest using an additive model with adjustment for clinical factors, co-medications, and established genetic predictors (i.e., VKORC1; CYP2C9). For warfarin dose and anticoagulation control (PTTR) as outcomes, multivariable linear regression modeling was performed. The association of CYP4F2 with risk of hemorrhage was examined using Cox proportional hazards regression.The frequency of the minor (A) allele was higher in European (52%) compared to African (16%) Americans. Compared to the GG genotype, the CYP4F2 A allele was associated with a 7.0% higher (p=0.03) warfarin dose [5.5% higher (p=0.08) for AG and 13.4% higher (p= 0.03) for AA genotype] in European Americans. CYP4F2 did not significantly influence warfarin dose in African Americans [0.1% higher (p=0.98) for AG and 13.6% higher (p=0.6) for AA genotype]. PTTR was not influenced by CYP4F2 for either European American (p=0.59) or African American patients (p=0.78).The incidence of hemorrhage was lowest among patients with the AA genotype (1.05/100pyrs) compared to the AG (8.1/100pyrs) and GG (6.8/100pyrs) genotype. Patients with AA genotype had a lower risk of hemorrhage [HR: 0.17 (0.02-1.37), p=0.09] compared to patients with GG or AG genotype. Incorporation of CYP4F2 genotype can provide additional improvement in warfarin dosing among European Americans. Additionally, CYP4F2 AA genotype appears to provide a protective effect against the risk of hemorrhage but further research is needed to confirm these associations and to establish the utility of CYP4F2 in the care of a racially diverse patient population.

704W

A mechanism for docetaxel induced neutropenia: The role of Cwc27. *T. Wiltshire[1], C.S. Benton[1], F. Muhale[2], O. Suzuki[1], A. Frick[1], O.J. Trask[2], R. Thomas[2], S. Cai[3], H.L. McLeod[1].* 1) School of Pharmacy, University of North Carolina, Chapel Hill, NC; 2) The Hamner Institutes for Health Sciences, Research Triangle Park, NC; 3) Washington University School of Medicine.

Chemotherapy-induced neutropenia (CIN) often results in dose reduction or treatment delay, negatively impacting response rates and overall survival for cancer patients. Previous studies have suggested that genetic factors play a role in CIN, and genetic information could be used to better identify patients who are likely to develop CIN prior to drug and dose selection. To identify genes that mediate docetaxal-induced neutropenia, we conducted a genome-wide association analysis in twenty-one genetically diverse mouse inbred strains. After a 14-day treatment with docetaxel, we found a significant inter-strain difference in the development of neutropenia and this phenotypic difference was not correlated to docetaxel serum levels or AUC. Using genomewide-association analysis, we found a locus on Chr 13 that was significantly linked to docetaxel-induced neutropenia (−logP >3.0). Within this locus is the spliceosome associated protein homolog, Cwc27 gene. We observed a 2-fold reduction in IC50 values following Cwc27 shRNA knockdown cells compared to control (shRNA-Cwc27 = 19.5 nM, eGFP = 38.7 nM, and shRNA scramble = 41.2). Furthermore, shRNA knockdown of Cwc27 resulted in greater signal intensity for α tubulin, caspase 3, and phosphohistone H3, which suggests that the increased cell loss observed following Cwc27 knockdown was likely due to dysregulated microtubule assembly and organization that prevented cells from undergoing mitosis. These results along with the data from the FACs analysis of bone marrow samples indicated that strains with severe neutropenia had a greater than two-fold reduction in the numbers of progenitor cells compared to strains with moderate neutropenia. These data imply that the impact of docetaxel was not specifically in loss of neutrophils, but in the ability to replenish them after drug treatment. Altogether, our findings suggest that Cwc27 mediates risk to CIN through its role in cellular differentiation and consequently, hematopoietic stem cell mobilization and neutrophil production.

705W

ABCB1 rs3842 polymorphism affects tacrolimus pharmacokinetics during the first week after liver transplantation in Chinese. *J. Shi[1], C. Zhang[1], D. Chen[2], J. Fan[2], K. Zhang[1], Z. Peng[2], W. Huang[1,3].* 1) Department of Genetics, Chinese National Human Genome Center at Shanghai, Shanghai,China; 2) Department of General Surgery, Shanghai First People's Hospital, Shanghai Jiao Tong University, Shanghai, China; 3) Rui Jin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China.

Immunosuppressive drug, tacrolimus, is widely used in solid organ transplantation to prevent allograft rejection. However, it has a narrow therapeutic index with high inter-individual variations in its pharmacokinetics, which makes it difficult to establish an empirical dosage regimen in organ transplant recipients. Tacrolimus is a substrate for the multidrug efflux transporter P-glycoprotein, encoded by ABCB1 gene. A considerable number of reports have studied the correlation between the ABCB1 gene polymorphisms and the metabolism of tacrolimus. However, the obtained results are conflicting. The reason for these conflicting results might be the fact that the studies were mostly focused on 1236C>T (rs1128503), 2677G>T/A (rs2032582) and 3435C>T (rs1045642). However, these polymorphisms are not the only polymorphisms capable of influencing the P-glycoprotein function. Thus, we selected ten variations with minor allele frequency great than 3% in Han Chinese and examined genotypes for both recipient and donor in 115 liver transplantation patients, to clarify the influence of these genetic variants on tacrolimus dose requirements after liver transplantation. The recipient ABCB1 variant rs3842 significantly influenced the mean tacrolimus trough blood concentrations ($P=2.6\times10^{-4}$), and the mean dose-adjusted trough concentrations ($P=8.8\times10^{-3}$) during the first week post transplantation. A significantly higher trough blood concentration was observed for recipients with the homozygous variant rs3842 GG genotype compared to those with the homozygous AA genotype. This association remained significant after adjusting for age, sex, clinical outcome parameters and information of combined drugs. None of the donar ABCB1 polymorphism showed correlation with tacrolimus pharmacokinetics. The extensively studied 3435C>T, 2677G>T/A and 1236C>T in both donor and recipient showed no correlation with tacrolimus pharmacokinetic. This study indicates, for the first time, that recipient ABCB1 polymorphism rs3842 is associated with tacrolimus dose requirements limited to the early post-transplant period. The result highlights the importance of screening the comprehensive polymorphisms in ABCB1 gene region and examining the impact of polymorphisms in different post transplant time, which might provided new clues of the association of ABCB1 and tacrolimus pharmacokinetic.

706W

Effect of APOE genotype on response to donepezil and placebo in subjects with mild-to-moderate Alzheimer's dementia. *D.P. King, J.F. Waring, Q. Tang, W.Z. Robieson, U. Das, J. Dubow, S. Dutta, G.J. Marek, L.M. Gault.* AbbVie Inc., North Chicago, IL.

Background: Previous studies examining the influence of the apolipoprotein E (APOE)-ε4 allele on the clinical response to acetylcholinesterase treatment in Alzheimer's dementia patients have yielded inconsistent results. Data on the donepezil response from 3 studies in subjects with mild-to-moderate Alzheimer's dementia were pooled to examine this relationship. Methods: Data from 3 multinational, randomized clinical studies with a 12-week treatment duration, similar study design and a statistically significant response to donepezil were pooled. Patients with mild-to-moderate Alzheimer's dementia (Mini-Mental Status Examination [MMSE] score of 10-24, inclusive) provided DNA during study screening. APOE-ε4 carrier status was determined using pyrosequencing. The change from baseline to final observation on the 13-item ADAS-Cog total score in the intent-to-treat population was analyzed using ANCOVA with terms of age, treatment, study site, APOE-ε4 allele status, using baseline ADAS-Cog score as a covariate. Results: Placebo (N=170) and donepezil (N=165) patients had a mean (SD) age of 72 (8.5) years and a mean (SD) baseline MMSE of 19 (3.9). Although the interaction between APOE-ε4 carrier status and treatment was not significant (P=0.61), donepezil significantly improved ADAS-Cog scores versus placebo in both the APOE-ε4 carriers (donepezil [N=81], placebo [N=94], LS mean difference from placebo = −2.34, P=0.01) and noncarriers (donepezil [N=84], placebo [N=76], LS mean difference from placebo = −1.71, P= 0.05). The baseline to final change in ADAS-Cog scores did not differ significantly (P=0.89) among donepezil patients in terms of APOE-ε4 allele copy number: 0 APOE-ε4 alleles (LS mean [SE] change from baseline = −4.09 [0.67], N=84); 1 allele (−2.97 [0.73], N=72); 2 alleles (−2.74 [1.98], N=9). Of note, the magnitude of the donepezil treatment response was similar in APOE-ε4 carriers and noncarriers; however, the carrier group exhibited a smaller placebo response than the noncarrier group (LS mean difference baseline to final = −0.60 versus −2.38, P=0.05). Conclusions: There was no significant effect of APOE-ε4 genotype on the donepezil response in these mild-to-moderate Alzheimer's dementia patients. APOE-ε4 carriers in the placebo group exhibited a smaller change from baseline on the ADAS-Cog than noncarriers. Thus, APOE-ε4 genotype status may be useful for randomization in future clinical trials. Support: AbbVie.

707W

The Role of Multidrug Resistance-1 (MDR1) Variants in Response to Atorvastatin among Jordanians. *K.H. Alzoubi[1], O.F. Khabour[2], S.A. Alazzam[1].* 1) Clinical Pharmacy, Jordan University of Science and Technology, Irbid, Jordan; 2) Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid, Jordan.

The MDR1 gene encodes for P-glycoprotein (P-gp), which is an efflux transporter at the cell membrane. The P-gp has wide substrate specificity for multiple medications including the lipid lowering drug, atorvastatin. In this study, we investigated the possible association between three common MDR1 gene polymorphisms (G2677T, C3435T, and C1236T), and the lipid lowering effect of atorvastatin among Jordanians. Lipid and liproproteins were measured in blood samples collected from patients (n = 201) at baseline and during atorvastatin treatment. MDR1 polymorphisms were genotyped using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Both the TT genotype of G2677T and the TT genotype of the C3435T polymorphisms were associated with lower levels of low-density lipoproteins (LDL) after atorvasatin treatment. However, the effects of atorvastatin on the levels of total cholesterol (Tchol), triglycerides (TG), and high-density lipoprotein (HDL), were not correlated with any of the genotypes in both polymorphisms. Finally, the C1236T polymorphism was not associated with the lipid lowering effect of atorvastatin. In conclusion, the MDR1 gene polymorphisms G2677T, and C3435T, but not C1236T were associated with the lipid lowering effect of atorvastatin among Jordanians.

708W

Association of *ABCB1*, *SLC15A1*, *CES1* and *NEU2* gene polymorphisms with side effects by oseltamivir in a Mexican population. *M. Bermudez de Leon Sr[1,2], R.N. Gonzalez-Ríos[1,2,4], R.A. Leyva-Parra[1], E. González-González[1,4], A. Alvarado-Díaz[1,4], O.E. Vázquez-Monsiváis[2], V.L. Mata-Tijerina[2], B.L. Escobedo-Guajardo[2], M.J. Currás-Tuala[2], M.E. Aguado-Barrera[2], B. Silva-Ramírez[3], J.M. Alcocer-González[4], A.M. Salinas-Martínez[5].* 1) Dept of Mol Biol, Centro de Investigación Biomédica del Noreste. Instituto Mexicano del Seguro Social, 64720 Monterrey, Nuevo León, Mexico; 2) Laboratory of Molecular Diagnostics, Centro de Investigación Biomédica del Noreste. Instituto Mexicano del Seguro Social, 64720 Monterrey, Nuevo León, Mexico; 3) Dept of Immunogenetics, Centro de Investigación Biomédica del Noreste. Instituto Mexicano del Seguro Social, 64720 Monterrey, Nuevo León, Mexico; 4) School of Biological Sciences, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León, Mexico; 5) Epidemiology Research and Health Services Unit, Instituto Mexicano del Seguro Social, 64000 Monterrey, Nuevo León, Mexico.

Oseltamivir is a neuraminidase inhibitor extensively used during the latest pandemic influenza to prevent the release of progeny virions and thereby limit the spread of infection. Oseltamivir phosphate is a prodrug that is absorbed by Peptide transporter PEPT1 and effluxed by P-glycoprotein. Once inside of the cell, oseltavimir phosphate is converted by human carboxilesterase 1 to oseltamivir carboxilate, which is the active form to inhibit influenza virus neuraminidase. However, oseltamivir also has an inhibitory effect on human sialidases, that are important in varios cellular functions including lysosomal catabolism. There are reports about deaths and neuropsychiatric events in Japanese population with the use of oseltamivir, suggesting that this drug could inhibit also to human sialidases in a similar way that viral sialidase. Then, the aim of this study was to determine the single nucleotide polymorphisms (SNPs) for PEPT1(*SLC15A1* gene), P-glycoprotein (*ABCB1* gene), carboxilesterase 1 (*CES1* gene) and sialidase (*NEU2* gene) in >700 Mexican patients with oseltamivir therapy, and these data were correlated with side effects reported between 2010-2012. The SNPs evaluated were Gly185Val (rs1128501), Ser893Thr/Ala (rs2032582) and Ile1145Ile (rs1045642) for *ABCB1* gene; Ser117Asn (rs2297322) for *SLC15A1* gene; Gly143Glu and Arg199His (rs71647871 and rs2307243, respectively) for *CES1* gene; and Arg41Gln (rs2233385) for *NEU2* gene. Clinical data of each patient were obtained from institutional electronic files. This work was approved by the Ethics Committee of Mexican Social Security Institute. We found that eight percent of patients showed side effect as depression, anxiety, seizures, hallucinations among others. Genetic and allelic frequencies are presented and associated with the side effects of oseltamivir in a Mexican population. The authors thank Consejo Nacional de Ciencia y Tecnologia for financial support (Grant number SALUD-2011-1-162243).

709W

PGRN Network-wide Project: Transcriptome Analysis of Pharmacogenes in Human Tissues. *E.R. Gamazon[1], C.E. French[2], S.W. Yee[3], A. Chhibber[3], X. Qin[4], E. Theusch[5], A. Webb[6], A. Konkashbaev[1], S. Weiss[7], S.E. Brenner[2], S.E. Scherer[4], N.J. Cox[1], K.M. Giacomini[3], A. on behalf of PGRN RNASeq Group[3].* 1) University of Chicago, Chicago, IL; 2) University of California, Berkeley, CA; 3) University of California, San Francisco, CA; 4) Baylor College of Medicine, Houston, TX; 5) Children's Hospital Oakland Research Institute, Oakland, CA; 6) Ohio State University, Columbus, OH; 7) Harvard School of Public Health, Boston, MA.

Gene expression variation impacts many common disorders and pharmacological traits; however, the nature and extent of this variation remains poorly understood. The NIH Pharmacogenomics Research Network (PGRN) Network-wide RNA-seq project aims to create a community resource containing quantitative information on known and novel isoforms of genes involved in therapeutic and adverse drug response (pharmacogenes, see http://www.pharmgkb.org/search/annotatedGene/). Using 160 samples from 6 major tissues of pharmacologic importance (liver, kidney, adipose, heart, lymphoblastoid cell lines [LCLs], and brain), some with extensive pharmacogenomic phenotyping, we performed transcriptome sequencing. The data were analyzed for expression quantification, junction analysis, and transcript reconstruction. We utilized the JuncBASE pipeline developed by members of our consortium to identify and classify splicing events. In samples from heart, kidney, liver and adipose tissues, similar numbers of transcripts and genes were detected; however, notable differences in expression levels of important pharmacogenes across the various tissues were observed. For example, as expected, many CYP enzymes (e.g., CYP2A7 and CYP2D6) were highly expressed in the liver and showed low levels of expression in other tissues. Other important drug metabolizing enzymes such as DPYD and PPARD showed more balanced gene expression patterns across the tissues. We uncovered substantial variation in both annotated and novel splicing events. For example, LDLR--a major target of the statins—showed markedly different splice variant patterns across the various tissues. In addition, given the importance of LCLs as a pre-clinical model for human genetic studies, we systematically investigated differential expression and splicing between LCLs and the other tissues. These studies provide mechanistic insights into pharmacogenomic findings and facilitate an understanding of the factors that lead to inter-individual differences in drug response.

710W

Systematical functional characterization of CYP2D6 alleles in the Chinese Han population. *S. Qin, Q. Xu, Z. Wu, L. He.* Bio-X Institutes, Shanghai Jiaotong University, Shanghai, Shanghai, China.

Cytochrome P450 2D6 (CYP2D6) plays a crucial role in the metabolism of approximately 30% of drugs presently on the market and CYP2D6 gene polymorphisms exhibit high individual variability in catalytic activity. Based on the database of CYP2D6 gene polymorphisms in the Chinese Han population established by our group, we functionally characterized CYP2D6 alleles in this population, including four novel alleles CYP2D6*75 (g.4046G>A, p.441R>H), MU1 (g.100C>T + 2467G>A. p.34P>S + 231P>L), MU2 (g.100C>T + 2851C>T. p.34 P>S + 296R>C), MU3 (g.2467G>A + 2851C>T. p.231P>L + 296R>C) and CYP2D6*2, CYP2D6*10, CYP2D6*14. And we have successfully established the phenotype profile of CYP2D6 in the population. CYP2D6 proteins of wild-type (CYP2D6.1) and the seven variants were heterologously expressed in yeast cells and the kinetic parameters (Km, Kcat, Vmax,CLint and Kcat/Km) for debrisoquine 4-hydroxylation were determined. The data suggests the enzyme activity of CYP2D6.75 has been decreased than that exhibited by CYP2D6.1 , indicating R441 plays an important role in catalytic reaction. The kinetic parameters of CYP2D6.2 were much the same with CYP2D6.1. As a most prevalence variant in Chinese population, CYP2D6.10 showed decreased enzyme activity. We found that four alleles MU1, MU2, CYP2D6.10 and CYP2D6.14 which all harbor 100C>T site exhibited similar activity, suggesting P34 is crucial for CYP2D6 enzyme activity and catalytic efficiency. This is the first study to conduct systematic phenotype profile analysis of CYP2D6 alleles in the Chinese Han population. These findings might be useful for optimizing pharmacotherapy and the design of personalized medicine.

711W

A pharmacogenomic study of inhaled short-acting beta-agonist response in African American individuals identifies SPATA13-AS1. *B. Padhukasahasram[1], J.J. Yang[2], A.M. Levin[2], M. Yang[1], E.G. Burchard[3,4], R. Kumar[5], P. Kwok[6,7], D.E. Lanfear[1,8], L.K. Williams[1,8].* 1) Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI; 2) Department of Public Health Sciences, Henry Ford Health System, Detroit, MI; 3) Departments of Medicine, University of California, San Francisco, CA; 4) Department of Bioengineering & Therapeutic Sciences, University of California, San Francisco, CA; 5) Department of Pediatrics, The Ann and Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, IL; 6) Department of Dermatology, University of California, San Francisco, CA; 7) Cardiovascular Research Institute, University of California, San Francisco, CA; 8) Department of Medicine, Henry Ford Health System, Detroit, MI.

Inhaled short-acting beta2-agonists (SABAs) are commonly used in patients with asthma to rapidly reverse airway obstruction and improve acute symptoms. However, treatment response to SABA medication is highly variable and is likely to be genetically influenced. We performed a genome wide association study of SABA medication response in 328 healthy, non-asthmatic African Americans and replicated these findings in 1073 individuals with asthma and 149 without asthma. Healthy, non-asthmatic individuals were used for discovery as SABA response might be less obscured by unrelated factors associated with lung disease. A linear mixed model approach was first used for SNP associations, and results were then combined to generate gene-based associations. We also assessed whether genes associated with SABA response were related to the amount of SABA use among those with asthma. Gene-based association tests identified SPATA13-AS1, an antisense RNA encoding segment within the SPATA13 gene as being significantly associated with SABA response. In replication, this gene was also associated with SABA response among 1073 African American individuals with asthma ($p = 0.011$). The same association was also replicated in an additional 149 healthy African American individuals ($p = 0.027$). Lastly, SPATA13-AS1 was also associated with annual SABA medication use among individuals with asthma ($p = 0.047$). SPATA13 is a recently discovered protein that is likely to be involved in airway smooth muscle contraction-relaxation through its interactions with Rho family GTPases such as RhoA, Rac1 and Cdc42. Elucidating the precise mechanism of action of this protein may reveal new SABA response pathways as well as targets for future asthma therapeutics.

712W

Genome-wide analysis of Methotrexate pharmacogenomics in rheumatoid arthritis reveals novel risk variants and leads for *TYMS* regulation. *B.K. Thelma[1], S. Senapati[1], S. Singh[1], M. Das[1,2], A. Kumar[3], R. Gupta[4], U. Kumar[5], S. Jain[6], R.C. Juyal[2].* 1) Dept Genetics, Univ Delhi, South Campus, New Delhi, Delhi, India; 2) National Institute of Immunology, New Delhi, India; 3) Department of Rheumatology, Fortis Flt. Lt. Rajan Dhall Hospital, New Delhi, India; 4) Division of Rheumatology & Clinical Immunology, Medanta Bone & Joint Institute, Medanta - The Medicity, Gurgaon, India; 5) Department of Rheumatology, All India Institute of Medical Sciences, New Delhi, India; 6) Department of Physics, University of Delhi, Delhi, India.

Objective: Methotrexate (MTX) is the drug of first choice for rheumatoid arthritis (RA) treatment but is effective only in around 60% of the patients. Further, early initiation of MTX in a brief critical window period of around 6 months is crucial for the effective treatment. Therefore, identification of predictive markers is critical for the choice of early and appropriate treatment regimen. To date, genetic determinants underlying MTX response have remained elusive despite extensive candidate gene association studies. In the present study, we used genome-wide genotype data to identify potential risk variants associated with MTX (non)response in a north Indian RA cohort. **Methods:** Genome-wide genotyping data from Illumina 660w quad array for a total of 457 RA patients (297 good [DAS 28-3 < 3.2] and 160 poor [DAS 28-3 >5.1] responders) on MTX monotherapy were tested for association using additive model. Other tools including support vector machine (SVM) and genome-wide pathway analysis were used to identify additional risk variants and pathways. All risk loci were imputed to fine-map the association signals and identify causal variant(s) of therapeutic/diagnostic relevance. **Results:** Nine novel suggestive loci from GWAS ($p \leq 5 \times 10^{-5}$) and two from SVM analysis were associated with MTX (non)response. Association of published candidate genes namely *DHFR* (p=0.014), *FPGS* (p=0.035) and *TYMS* (p=0.005) and purine and nucleotide metabolism pathways were reaffirmed. Imputation followed by bioinformatic analysis indicated possible interaction between two reversely oriented overlapping genes namely *ENOSF1* and *TYMS* at post-transcriptional level. **Conclusion:** In this first ever genome-wide analysis on MTX treatment response in RA patients, 11 new risk loci were identified. Further, *TYMS* expression at post-transcriptional level seems probably regulated through an antisense-RNA involving the 6bp *ins/del* marker in the overlapping segment at 3' UTR of *TYMS-ENOSF1*, a finding with impending pharmacogenetic applications.

713W

Using Neuroimaging Endophenotypes to Identify Molecular Markers for Treatment Response to Major Depressive Disorder. *T. Carrillo-Roa[1], C.A. Lareau[2], C.L. McGrath[3], B.W. Dunlop[3], M.E Kelley[4], E.B. Binder[3], H.S Mayberg[5].* 1) Max Planck Institute for Psychiatry , Munich, Germany; 2) Department of Mathematics, University of Tulsa; 3) Department of Psychiatry and Behavioral Sciences, School of Medicine, Emory University; 4) Department of Biostatistics and Bioinformatics, School of Public Health, Emory University; 5) Department of Psychiatry and Behavioral Sciences and Department of Neurology, School of Medicine, Emory University.

Major depressive disorder (MDD) is a prevalent disease with high rates of treatment resistant and non-remission. However biological measures to guide optimal treatment have been lacking. Recently, our group described that resting state brain activity patters (BAPs) of specific brain regions (i.e. right insula, right inferior temporal cortex, left amygdala, left premotor cortex, right motor cortex and precuneus) can predict differential response to either escitalopram (sCIT) or psychotherapy. The aim of this study was to identify molecular markers that associate with these BAPs, in the hope to identify predictive measures that are more easily obtained in clinical practice than neuroimaging measures. Patients were recruited at Emory University and randomized at baseline to either 12 weeks sCIT, or 16 sessions of cognitive behavioral therapy (CBT). Pre-treatment BAPs of the six brain regions of interest (ROIs) were used as neuroimaging endophenotypes for this study. In peripheral blood DNA drawn at baseline genome-wide SNP genotypes (Illumina OmniExpress) and DNA methylation patterns (Illumina HM 450K) were measured. Genome-wide univariate and multivariate association analyses including all possible ROIs combinations were conducted in 76 MDD patients. Tests for association between methylation status at ~485,000 CpG sites and BAPs were performed for each of the ROIs. We observed genome-wide significant association of rs34383296, ($p = 9.4 \times 10^{-9}$) in a multivariate analysis that included the right insula, left amygdala and left premotor cortex. Univariate analyses did not reveal genome-wide significant associations. The associated variant lies in a gene dense region on chromosome 9 within the NDOR1 gene locus and it is an eQTL for ARRDC1, a gene ~400kb downstream and is related to arrestin-mediated internalization of cell surface receptors. No genome-wide significant association of DNA-methylation status was observed with any of the ROIs. Our data suggest that using quantitative neuroimaging endophenotypes and genomic approaches may be able to identify markers to guide individualized depression therapy choices in clinical routine. Further analysis will test surrounding CpG sites for the associated SNP as mQTLs as well as imputation of additional variants in the locus and tests for association with treatment response in independent samples.

714W

Genome-wide association study loci are enriched for clinically relevant drugs targets for common human diseases. *R. Li[1], V. Forgetta[2], O. Yu[3], Z. Dastani[4], M. Lathrop[5], J.B. Richards[1,6]*. 1) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, Centre Clinical Epidemiology, Jewish General Hospital, Montreal, Quebec, Canada; 2) Departments of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada; 3) Department of Medicine, McGill University, Montreal, Quebec, Canada; 4) Departments of Epidemiology, Biostatistics and Occupational Health. Jewish General Hospital, Lady Davis Institute, McGill University Montreal, Quebec, Canada; 5) McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada; 6) Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

While many genetic loci have been recently mapped to common human disease it is not yet apparent if this information will aid in drug development. We reasoned that if GWAS loci identified such targets then these loci would be enriched for the targets of drugs that are already on the market. To test this hypothesis, we determined enrichment of GWAS loci for clinically relevant drugs for multiple common human diseases. We selected easily definable, representative, common diseases, with relatively high heritabilities for which large-scale GWAS meta-analyses had been published. These included, type 2 diabetes, hypercholesterolemia (LDL), osteoporosis, Crohn's disease, hypertension, psoriasis and asthma. We also included an additional disease, COPD, whose current therapies are largely aimed at symptom control. Pipeline was used to identify marketed drugs and their targets and this list was refined to include only drugs accepted by relevant American clinical guidelines. The NHGRI GWAS Catalog was used to collect all genome-wide significant SNPs for each disease and genes within 100kb or 500kb were identified. The nearest network neighbours for each drug target were identified using STRING. Finally, since several immune-mediated diseases have promising interleukin-based therapies in development we assessed whether drugs in phase III of development were enriched at GWAS loci. On average, there were 9 drug targets per disease that achieved guideline-level acceptance. Assessing all genes within 500kb of GWAS SNPs and the nearest network neighbours of drug targets, 80% of guideline-accepted LDL drug targets were identified. The proportion of such targets identified for osteoporosis was 80%, 71% for type 2 diabetes, 43% for Crohn's, 22% for psoriasis, 15% for hypertension, 9% for asthma and 0% for COPD. Consideration of Phase III drugs improved these proportions to 56% for Crohn's, 45% for psoriasis and 29% for asthma. Genes within 100kb of GWAS SNPs were strongly enriched for guideline-accepted drug targets for some diseases (eg. 40-fold [95% CI: 14-118] for LDL and 35-fold [95%CI: 12-104] for osteoporosis). GWAS loci for some common diseases identify a large proportion of drug targets that are targeted by clinically relevant medicines. This implies that other clinically relevant drug targets exist among the novel loci identified from GWAS.

715W

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

An anticonvulsant, carbamazepine (CBZ), is known to show incidences of cutaneous adverse drug reactions (cADRs) including Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and drug-induced hypersensitivity syndrome (DIHS). Previously, to identify a gene(s) susceptible to CBZ-induced cADRs, we conducted a genome-wide association study (GWAS) and following HLA typing in 61 cases and 376 patients who showed no cADRs by administration of CBZ (CBZ-tolerant controls), and found that HLA-A*31:01 was present in 60.7% (37/61) of the patients with CBZ-induced cADRs, but in only 12.5% (47/376) of the CBZ-tolerant controls (odds ratio (OR) = 10.8, 95% confidence interval (CI) = 5.9–19.6, $P = 3.64 \times 10^{-15}$). To confirm associations of HLA-A alleles other than HLA-A*31:01 with CBZ-induced cADRs, we pick-upped a paired HLA-A allele of HLA-A*31:01 using case-control subjects carrying at least one HLA-A*31:01 allele, which consisted of 56 CBZ-induced cADR cases and 54 CBZ-tolerant controls. There was no HLA-A allele that reached significant association after Bonferroni correction with $P < 4.17 \times 10^{-3}$ (0.05/12 HLA-A alleles). To detect additional genetic marker for CBZ-induced cADRs, we performed genome-wide imputation using the GWAS data and genotype data from 1000 Genomes Project as reference panels by MaCH-Admix software. For a conditional logistic regression analysis of the imputed genotype data, we selected rs1633021 as the covariant with the firmest association observed in the analyzed GWAS peak of association. After the conditional analysis, there were 22 loci that reached GWAS-level significant association ($P < 5 \times 10^{-8}$). The replication analysis using independent population are ongoing.

716W

Whole genome association study identifies novel antidepressant response loci for the treatment of obsessive-compulsive disorder with selective serotonin re-uptake inhibitors. *H.D. QIN[1], Y. Wang[2], M.A. Grados[2], M.A. Riddle[2], B.D. Greenberg[4], J.A. Knowles[5], A.J. Fyer[6], J.T. McCracken[7], D.L. Murphy[8], S.A. Rasmussen[4], B. Cullen[2], J. Piacentini[7], D. Geller[3], D. Pauls[3], E. Stewart[10], O.J. Bienvenu[2], Y. Chen[9], F.S. Goes[2], B. Maher[2], J.F. Samuels[2], G. Nestadt[2], Y.Y. Shugart[1].* 1) Unit of Statistical Genomics, Intramural Research Program, Division of Intramural Research Program, National Institute of Mental Health, NIH, Bethesda, MD 20892, USA; 2) Department of Psychiatry and Behavioral Sciences, School of Medicine, Johns Hopkins University, Baltimore, Maryland; 3) Departments of Psychiatry and Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA; 4) Department of Psychiatry and Human Behavior, Brown Medical School, Butler Hospital; 5) Department of Psychiatry, Keck Medical School, University of Southern California; 6) College of Physicians and Surgeons at Columbia University; 7) Department of Psychiatry and Biobehavioral Sciences, School of Medicine, University of California, Los Angeles; 8) Laboratory of Clinical Science, NIMH, NIH, Bethesda; 9) Department of Medicine, University of Maryland School of Medicine; 10) Department of Psychiatry, University of British Columbia, A3-118, West 28th Avenue, Vancouver, BC, Canada V5Z 4H4.

Selective serotonin reuptake inhibitors (SSRIs) are first line medications for the treatment of obsessive-compulsive disorder (OCD). Although SSRIs are currently the most frequently used drug therapy for OCD, approximately 30% of OCD patients show limited or no response to these medications, and >7% cannot tolerate side effects. Genetic predictors for OCD treatment responsiveness are in demand. We hypothesized that genetic variations in genes expressed in human brain influence SSRI response. To test this hypothesis, we conducted an association study on SSRI drug effect in 1,773 OCD cases from a family-based GWAS study. We used a powerful Quasi-Likelihood Score Test, namely, MQLStest, to conduct association test correcting for the relatedness coefficients (based on identity-by-descent). A sex-, age- and cognitive behavioral therapy (CBT)- adjusted logistic model was used to evaluate effect size. The significant variants ($P < 10^{-4}$) with large effect-size (odds ratio ≥1.75) for drug response are reported. We identified 15 significant loci (PMQLS_Robust $<10^{-4}$) with large effect-size (adjusted odds ratio ≥1.75) for OCD patients' response to SSRIs. The top two significant SNPs are rs17162912 and rs9957281, with $P = 3.44\times10^{-7}$, OR = 2.91 (95%CI 1.94-4.36) and $P = 1.43\times10^{-5}$, OR = 2.08 (1.36-3.2), respectively. rs17162912 is intergenic variation near DISP1 gene on 1q41-q42, a microdeletion region related to mental retardation. rs9957281 is located within a newly identified obsessive-compulsive disorder susceptibility gene DLGAP1, a member of the neuronal postsynaptic density complex. The 13 other significant variants are located in six genes, namely LINC00256B, CDYL2, PAFAH1B1 (LIS-1), TACC1, CSMD1, and GAS2. Interestingly, all these genes are expressed in human brain, In particular, DLGAP1 and CSMD1 have the most biological relevance to SSRI response based upon the current literature, suggesting these top ranked variants or genes may play an important role in SSRI drug-response in OCD patients. The importance of identification of drug response loci is for the development of 'personalized medical treatment' of OCD patients treated with SSRIs. The potential results would provide also new targets for developing novel drugs for the treatment of non-responders. Further well-designed case-control studies with large sample size and using next-generation sequencing are needed to explore the role of causal exonic variations or rare CNVs of the most significant genes.

717W

Integrative genome modeling reveals common genetic architecture of neuropathy resulting from distinct environmental exposures. *H.E. Wheeler[1], C. Wing[1], M. Komatsu[1], S. Delaney[1], E.R. Gamazon[1], C. Rodriguez-Antona[2], N.J. Cox[1], M.E. Dolan[1].* 1) Dept Medicine, University of Chicago, Chicago, IL; 2) Human Cancer Genetics Programme, Spanish National Cancer Research Center, Madrid, Spain.

Chemotherapy-induced peripheral neuropathy is the major dose-limiting toxicity for several mechanistically distinct anticancer drugs. Diabetic neuropathy affects a large proportion of diabetes patients due to prolonged exposure to high levels of glucose. Our goal is to elucidate the molecular genetic mechanisms underlying such neuropathies by integrating results from patient and lymphoblastoid cell line (LCL) genome-wide association studies (GWAS) with functional results from iCell Neurons derived from human induced pluripotent stem cells (iPSCs). When comparing modestly sized pharmacogenomic GWAS from patients and LCLs treated with the same drug, SNPs rarely overlap at stringent thresholds such as $P < 10^{-6}$, but significant overlaps of SNPs at more relaxed thresholds determined by enrichment analysis through random sampling are possible. Under this cumulative hypothesis, large numbers of common variants with small effects account for substantial heritability. For example, we observed an enrichment of paclitaxel-induced LCL (n = 247) cytotoxicity SNPs in the peripheral sensory neuropathy-associated SNPs from ovarian and lung cancer patients (n = 143) treated with paclitaxel and carboplatin (empirical $P = 0.034$). Interestingly, we also observed an enrichment of paclitaxel cytotoxicity SNPs in diabetic neuropathy-associated SNPs from patients (n = 1651) in the GoKinD cohort (empirical $P < 0.001$). These enrichments demonstrate that susceptibilities to increased cytotoxicity in LCLs and increased chemotherapy- and diabetes-induced neuropathy in patients likely have some genetic mechanisms in common. We are using iPSC-derived iCell Neurons to develop models for functional screens of candidate targets of interest from these enrichment analyses. Upon treatment of iCell Neurons with increasing concentrations of paclitaxel (0.001-100 μM) for 72 h, we identified a reproducible 3–5 μm (12–14%) decrease in cell median neurite process length and a 13–19 μm (10–13%) decrease in cell total neurite outgrowth per order of magnitude increase in drug. Paclitaxel binds to beta-tubulin to exert its cytotoxic effect. As a proof of concept, we have shown that decreased expression of the beta-tubulin isotype *TUBB2A* by siRNA transfection causes decreased median neurite process length (interaction $P < 10^{-3}$) and decreased total neurite outgrowth (interaction $P < 10^{-9}$) of iCell Neurons 48 h post-paclitaxel treatment. This work was supported by NIH grants U01GM61393 and F32CA165823.

718W

Genetic variants associated with elevated triglyceride levels in patients with genotype 2/3 chronic hepatitis C treated with cyclophilin inhibitor alisporivir. *Y. Li[1], M. Healey[1], M. Waldvogel[2], N. Hartmann[2], G. Nabel[1], L. Li[1], F. Staedtler[2], W. Zhang[3], C.A. Brass[4], N.V. Naoumov[5], K.J. Johnson[1], B. Li[1].* 1) Novartis Institutes for Biomedical Research, Cambridge, MA, USA; 2) Novartis Institutes for Biomedical Research, Basel, Switzerland; 3) Novartis Pharmaceuticals Corporation, Shanghai, China; 4) Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA; 5) Novartis Pharmaceuticals Corporation, Basel, Switzerland.

Background: Host-targeting antiviral alisporivir (ALV) inhibits cyclophilin A that is essential for HCV replication. In the phase IIb VITAL-1 study involving treatment-naïve genotype 2/3 patients with chronic hepatitis C, ~81% patients achieved SVR24 with IFN-free or IFN-containing ALV treatment with or without ribavirin. In a proportion (8.8%) of patients treated with ALV IFN-free regimens, maximum fasting triglyceride (TG) level of >400mg/dL was detected over the course of 24 weeks of treatment. The goal of this investigation was to evaluate the contribution of genetic variants to elevated TG levels observed in some patients in this study. Methods: DNA samples were obtained from 196 patients who consented to pharmacogenetic assessment and received ALV-containing treatment. Samples were genotyped using the Illumina Omni5Exome array and TaqMan allelic discrimination assays for APOE genotyping. Linear regression was performed to evaluate association between genotype and log-transformed maximum on-treatment TG level. The analysis was adjusted for race, log-transformed baseline TG, ALV exposure and presence of pegIFN treatment. Results: Targeted analysis of APOE genotype found that the APOE ε2 carriers had higher levels of TG than non-carriers, in patients treated with ALV (p = 0.035). However such a trend was not identified for ε4 carriers versus non-carriers (p = 0.12). Genome-wide association analysis (GWAS) with maximum fasting TG level while on ALV treatment identified one SNP in CNTNAP4 gene (in all races combined analysis; p= 2.63×10^{-9}) and one SNP in THSD7B gene (in Caucasians only; p=1.12×10^{-8}) to be statistically associated with elevated level of TG, after adjusting for multiple testing, although functional relevance of these genes (CNTNAP4 and THSD7B) to level of triglyceride is unclear. Conclusions: Overall, these findings suggest a potential role of genetic variation contributing to the elevated TG level observed in some patients treated with ALV. Targeted analysis revealed a trend of association between elevated TG level and the presence of the APOE ε2 allele in patients treated with ALV. Other genetic variations have also been identified from GWAS analysis to be associated with level of TG while treated with ALV-containing therapy without clear functional interpretation. Testing of these candidate variants in independent studies is needed for validation.

719W

Variants of the cysteinyl leukotriene 1 and 2 genes are additively associated with atopy in a founder population. *M. Thompson[1], J. Stankova[2], V. Capra[3].* 1) Lab Medicine, Banting Inst, Univ Toronto, Toronto, ON, Canada; 2) University of Sherbrook, Sherbrook, QC, Canada; 3) University of Milan, Milan, Italy.

The cysteinyl leukotriene receptor 1 (cysLT1) and 2 (cysLT2) genes have been investigated because they are functionally and pharmacologically implicated in the atopy phenotype affecting many asthma patients. In a founder population, we reported that the G300S variant of the cysLT1 receptor gene and the M201V variant of the CysLT2 receptor gene are implicated in atopic asthma. Here we discuss the statistical association of both variants with the atopy phenotype - a phenomenon suggesting that the interaction of cysLT1 and cyLT2 gene variants gives rise to atopy in the population we studied. The functional interaction of cysLT1 and cysLT2 proteins within the cell may represent a mechanism for the etiology of atopy in some individuals with atopic asthma.

720W

Identification of expression quantitative trait locus associated with drug biotransformation. *H.-C. Yang[1,2], C.-W. Chen[1].* 1) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; 2) School of Public Health, National Defense Medical Center, Taipei, Taiwan.

Scientific evidence has shown that drug response may vary with genetic background of populations. Pharmacogenomics association studies without accounting for population stratification and/or population admixture will cause false positive findings. Ancestry informative markers which exhibit different genetic distributions and reflect genetic ancestry in populations can be applied to adjust for population stratification and/or population admixture in genetics and pharmacogenomics association studies. This study aims to identify the ancestry informative markers which regulate gene expression and cause a down-stream consequence, differential drug responses in populations. We analyzed single nucleotide polymorphism and gene expression data of 210 independent samples in the HapMap II Project and identified a large number of cis- and trans-acting eQTL for each population by a partial least square method. Then we used our developed BIASLESS (Biomarkers Identification and Samples Subdivision) software (http://www.stat.sinica.edu.tw/hsinchou/genetics/prediction/BIASLESS.htm) [Yang et al., BMC Genomics, 2012, v13, 346] to identify ancestry informative eQTL that they are able to classify samples from different populations with a high testing accuracy in a cross-validation procedure. The identified eQTL were correlated with single nucleotide polymorphisms on drug biotransformation genes. We found that the differential allelic distributions of cytochrome P425 enzymes, which are the essential enzymes involving in drug metabolism, in populations were found. The results provide an insight into the complex relationship of genetic ancestry, gene regulation and pharmacoresponse.

721W

Genome-wide association study of combined paclitaxel and carboplatin treatment-induced severe neutropenia/leucopenia for patients in Biobank Japan. *S. Low[1], S. Chung[2,3], A. Takahashi[1], T. Mushiroda[4], M. Kubo[5], Y. Nakamura[2,3].* 1) Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Tokyo, Japan; 2) Department of Medicine, The University of Chicago, Chicago, IL; 3) Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 4) Laboratory for Pharmacogenomics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 5) Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan.

Chemotherapeutic agents are notoriously known to have narrow therapeutic range that often resulted in life-threatening toxicity. Hence, it is clinically important to identify the patients who are at high risk for severe toxicity through pharmacogenomics approach. In this study, a genome-wide association study was performed with cancer patients who administered combined paclitaxel and carboplatin treatment to identify genetic variants that are associated with the risk of severe neutropenia/leucopenia in the Japanese population. A total of 477 patients' DNA were recruited from the Biobank Japan with 166 patients who did not develop any adverse drug reactions, 161 develop mild (grade1/2) and 150 develop severe (grade3/4) neutropenia/leucopenia after administering the combined treatment. All of these samples were genotyped using Illumina OmniExpress BeadChip™. After sample and SNP quality control, Two genetic loci, rs12310399 (P=2.46×10^{-7}, OR=1.85, 95% CI =1.33-2.58) at *FGD6* gene and rs10785877 (P=7.38×10^{-7}, OR=2.58, 95% CI =1.77-3.77) near *RXRA* gene, were identified to be suggestively associated with paclitaxel and carboplatin treatment induced severe neutropenia/leucopenia when genome-association study was performed with patients who do not develop any adverse drug reaction versus those who develop severe neutropenia/leucopenia. Weighted genetic risk score analysis using six SNPs with P<1.0×10^{-5} identified that individuals who carry the highest range of genetic risk score possess 188 times (95% CI =36.1-979) higher risk to develop severe neutropenia/leucopenia compare to patients who belong to the lowest range of genetic risk score. Interestingly, individuals who developed grade 1/2 (mild neutropenia/leucopenia) were found to show intermediate risk scores between patients with severe neutropenia/leucopenia and those without any adverse reactions. Although we failed to identify genetic variants that surpassed the genome-wide significance level (P≤5.0×10^{-8}) through GWASs probably due to insufficient statistical power and complex clinical features, we were able to shortlist some of the suggestive associated loci. The current study is at the relatively preliminary state, but could highlight the complexity and problematic issues in retrospective pharmacogenomics studies. However, we hope that verification of these genetic variants through local and international collaborations could improve the clinical outcome of cancer patients.

722W

Expression of Mx1, OAS1, PKR and TP53 - interferon stimulated genes during treatment of hepatitis C patients. *B. Swiatek[1], I. Bereszynska[2], D. Januszkiewicz-Lewandowska[1,2,3], J. Wysocki[2], I. Mozer-Lisewska[2], A. Kowala-Piaskowska[2], J. Rembowska[1], J. Nowak[1].* 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) University of Medical Sciences, Poznań, Poland; 3) Department of Medical Diagnostics, Poznań, Poland.

Interferon stimulated genes (ISGs) such as Mx1, OAS1, PKR play a key role in antiviral responses against HCV infection. Moreover, it is suggested that ISGs pre-activation is associated with anti-HCV treatment failure. Also, it was observed that interferon stimulates transcription of TP53 gene. The aim of this prospective study was to examine the association between Mx1, OAS1, PKR (EIF2AK2) and TP53 expression and response to pegylated interferon and ribavirin treatment in 35 chronic hepatitis C (CHC) patients. Viral load was determined using one-step quantitative RT-PCR. Analyses were performed before as well as at 4 and 12 week of treatment.Rapid Virological Response (RVR) and complete Early Virological Response (cEVR) was achieved by 13 (37.1%) and 10 (28.6%) patients, respectively. 12 (34.3%) did not response to pegIFN and ribavirin combination treatment during 12 weeks (Primary Non-Response, PNR; less than 2 log10 decrease in viral titer after 12 weeks of therapy). The mean baseline viral load was comparable in RVR, cEVR and PNR group (6.7, 7.3 and 3.5×104 IU/ml, respectively). Expression of classical ISGs (Mx1, OAS1, PKR), but not TP53 increased during CHC treatment. The expression was low in RVRs and higher in cEVRs and PNRs before therapy and increased noticeable in cEVRs and PNRs and poorly in RVRs at week 4. The espression of studied ISGs was stable or poorly decreased in RVRs, was stable or poorly increased in cEVRs and noticeable decreased in PNRs group between week 4 and 12 of therapy. These results indicate that exogenous IFN can stimulate transcription of Mx1, OAS1, PKR in PBL in chronic hepatitis C patients. Moreover, increase in ISGs expression at week 4 of CHC therapy might depend on baseline level and next changes of HCV presence. In addition, level of ISGs tested might predict the outcome of combination CHC treatment. It looks like, pre-activation of the endogenous interferon system is associated with RVR and thereby with high likelihood of achieving SVR. It is also possible that treatment failure during first 12 weeks of anti-HCV therapy (PNR) may be related to noticeable decrease in ISGs expression between week 4 and 12. Lastly, it can be suggested that there is no association between TP53 expression during interferon treatment of CHC patients. This research was supported by National Science Centre grant no 2011/01/B/N26/04258.

723W

Analysis of the CYP3A4, CYP3A5, CYP2C9, CYP2C19 and CYP2D6 polymorphism in Nayarit population: Involvement in drug metabolism. *LE. Wong-Ley, AB. Martinez Rizo, JB. Velázquez Fernández, V. Mondragón Jaimes, AY. Bueno Durán, A. Zambrano Parra, FJ. Medrano Valenzuela.* Dept Gen, SSN-UAN, Tepic, Mexico.

Know that there is individual variability in the response to specific drugs given by genetic polymorphisms involved in metabolism. Drugs used in the treatment of multiple diseases have a different therapeutic range, with the possibility of death from toxicity, possibility of adverse effects that may limit the dose administered. The dose intensity is an important determinant of response to treatment. Many of the drugs used in the clinic are substrates of the CYP3A4, CYP3A5, CYP2D6, CYP2C9 and CYP2C19 polymorphism. They have been reported in different expression of these proteins in combination with polymorphisms in genes encoding them, so it is likely that these polymorphisms are involved in the response to these drugs. In Mexico there are few studies of the allelic frequencies of these genes and their correlation with the treatment of various diseases. The objective the present study is to know the association between the CYP3A4 *1B, CYP3A5 *3, CYP3A5 *6, CYP2D6 *4, CYP2C9 *2, CYP2C9 *3 and CYP2C19 *2 polymorphisms and the response to therapy of various drugs in Nayarit, a West Mexican population. They were included in the study 200 healthy patients. All patients signed an informed consent and a questionnaire was answered. DNA was obtained from lymphocytes and followed the identification by PCR and enzymatic digestion of CYP3A4 *1B, CYP3A5 *3, CYP3A5 *6, CYP2D6 *4, CYP2C9 *2, CYP2C9 *3 and CYP2C19 *2 variants. All genes studied are in Hardy-Weinberg equilibrium. The allelic frequencies of CYP3A4 * 1B, CYP3A5 * 3, CYP3A5 * 6, CYP2D6 * 4, CYP2C9 * 2, CYP2C9 * 3 and CYP2C19 * 2 genes in our state have relatively high frequencies compared to other populations, such as Caucasians, Asian or African, indicates a strong relationship between therapeutic response shown in our population and the polymorphisms studied. It is important to establish individualized treatments for better therapeutic management of diseases. We believe that this will be possible through better understanding of the physiology of both, the disease and the patient, thus allowing for personalized treatments.

724W

Can IL28B polymorphism identify patients who achieve early virological response during treatment of hepatitis C? *J. Nowak[1], B. Swiatek[1], J. Wysocki[2], I. Bereszynska[2], A. Kowala-Piaskowska[2], I. Mozer-Lisewska[2], J. Rembowska[1], D. Januszkiewicz-Lewandowska[1,2,3].* 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) University of Medical Sciences, Poznań, Poland; 3) Department of Medical Diagnostics, Poznan, Poland.

Large number of factors affecting response to hepatitis C therapy makes it difficult to optimize. Identification of molecular markers playing role in predicting anti-HCV treatment outcome before its beginning would facilitate therapeutic decision-making and therapy optimizing. Recently IL28B rs12979860 (C/T) polymorphism has been identified as a strong predictor of Sustained Virological Response (SVR) in chronic hepatitis C (CHC) patients treated with pegylated interferon and ribavirin (pegIFN+RBV). Moreover, monitoring viral kinetics, particularly during the first few weeks of therapy, can help identify patients with high chances of treatment success. The aim of this prospective study was to examine the association between SNP rs12979860(C/T) in the IL28B gene and on-treatment virological response (week 4. and 12.) in CHC patients. The study consisted of 35 patients who were treated with pegylated interferon α2a/2b and ribavirin. To determine treatment effects, serum HCV-RNA was measured on the first day of therapy and then at 4 and 12 week of treatment by one-step quantitative RT-PCR. DNA, isolated from peripheral blood lymphocytes, was used for IL28B rs12979860 (C/T) genotyping by High Resolution Melting method.13 patients (37.1%) became HCV RNA negative at week 4 (RVR-Rapid Virological Response) and 10 (28.6%) at week 12 (cEVR-complete Early Virological Response). 12 patients (34.3%) did not achieve virological response until 12 weeks (PNR-Primary Non-Response, less than 2 log10 decrease in viral titer after 12 weeks of treatment). The mean baseline viral load was comparable - 6.69×104 IU/ml vs 7.32×104 IU/ml vs 3.51×104 IU/ml in RVR, cEVR and PNR group, respectively. The rs12979860 CC, CT and TT genotypes found in 8 (22.9%), 23 (65.7%), 4 (11.4%) patients, respectively. Among patients with CC genotype, 75% achieved RVR and 25% achieved cEVR. Among CT genotype RVR, cEVR and PNR were observed in 30.5%, 30.5% and 39% of patients, respectively. 25% of patients with the genotype TT achieved cEVR and 75% achieved PNR. Favorable CC was not observed in PNR and unfavorable TT genotype was not observed in RVR group.The initial results confirm that IL28B rs12979860 C/T polymorphism may identify those CHC patients, treated with pegIFN and ribavirin, who achieve virological response during early phase of treatment and thereby who are more likely to achieve SVR. This research was supported by National Science Centre grant no 2011/01/B/N26/04258.

725W

Effect of genetically tailored statin therapy on health behaviors and outcomes: A pilot study in the primary care setting. *J.H. Li[1], S.V. Joy[2], S.B. Haga[1], L.A. Orlando[1], W.E. Kraus[1], G.S. Ginsburg[1], D. Voora[1].* 1) Duke University Medical Center, Durham, NC; 2) University of Colorado Denver Anschutz Medical Campus, Aurora, CO.

Background: Despite the cardiovascular benefits of statins, long-term adherence is often limited by real or perceived side effects. The **5* variant in the hepatic transporter *SLCO1B1* is a risk factor for myopathy, a common side effect. The risk conferred by **5* is statin-specific: greatest with simvastatin and atorvastatin (S/A) and least with pravastatin and rosuvastatin (P/R). We hypothesized that providing *SLCO1B1*5* genotype guided statin therapy (GGST) would be associated with improved patient adherence, provider behavior, and laboratory outcomes in primary care patients who were initially nonadherent to statins in a pilot study. **Methods:** Patients (n=58) and their providers received **5* genotype results and GGST recommendations pushed to them via the electronic medical record (EMR). Noncarriers were not at increased risk of side effects and were reassured to restart any statin, while carriers had a higher risk of side effects from S/A and were recommended therapy with P/R. The primary outcome was the change in patients' perceived need for statins and concerns about risks from baseline to 4 months, measured by a validated survey. Concurrent controls (n=59) receiving standard of care were gathered from the same clinic for comparison of secondary outcomes: 1) proportion with new statin prescriptions written by their providers, 2) change in low-density lipoprotein cholesterol (LDLc), 3) patient-reported statin usage over 1 year. **Results:** The largest changes were in the 'need for statin to prevent sickness' (pre: 2.9±0.9 vs. post: 3.3±0.9, p<0.001) and the 'concern for statin to disrupt life' (pre: 3.2±1.4 vs. post: 2.8±1.2, p=0.006). Overall, GGST patients expressed a trend toward higher necessity (p=0.2) and lower concern (p=0.2) for statins at 4 months compared to baseline. Compared to controls, GGST patients had a higher proportion of new statin prescriptions (55% vs. 20%, p<0.001), greater change in LDLc (-17.0±48.3 vs. 6.3±37.8, p=0.03), and higher patient-reported statin use (45% vs. 15%, p<0.001) during follow-up. **Conclusions:** Delivery of *SLCO1B1*5* genotype results coupled with genotype guided recommendations to tailor statin therapy is feasible via the EMR in the primary care setting. This novel intervention improved patients' perceptions of statins and was associated with physician and patient behaviors that promoted higher statin adherence and lower LDLc. The impact of **5* genotyping should be further explored in a larger randomized control trial.

726W

Incidental pharmacogenetic variants identified by massively parallel sequencing in the ClinSeq® study. *D. Ng[1], J.J. Johnston[1], K.L. Lewis[1], S.G. Gonsalves[1], W. Newman[2], P.D. Stenson[3], D.N. Cooper[3], M. Vemulapalli[4], J.C. Mullikin[4,5], L.G. Biesecker[1,4], NIH Intramural Sequencing Center.* 1) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Manchester Centre for Genomic Medicine, St. Mary's Hospital, Manchester, UK; 3) Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK; 4) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Massively parallel sequencing can identify incidental, medically relevant genetic variants in asymptomatic individuals. This represents a major paradigm shift in medicine, from the treatment of manifest disease to predictive medicine aimed at monitoring asymptomatic individuals with disease susceptibility. One of the aims of ClinSeq® is to study the analysis and return of incidental, clinically relevant genetic variants (secondary findings). As part of our ongoing endeavor to study secondary variants, we selected 20 genes associated with drug metabolism or transport for analysis. Whole exome sequences (WES) from 951 ClinSeq® participants were annotated for variants in 20 drug metabolism genes selected from the Affymetrix DMET plus array using the Human Gene Mutation Database (HGMD®) cDNA/protein reference. Of the 327 variants targeted by the Affymetrix DMET plus chip, 70.6% (231/327) had sequence coverage in ≥80% (761–951) and 20.5% (67/327) were covered in <80% (1–759) of participants. WES identified 360 variant positions in the 20 genes. Ninety-two of the pharmacogenetic variants interrogated by the Affymetrix chip were identified by WES in one or more ClinSeq® participants. Two hundred sixty-eight variant positions were novel. Twenty-four of these variants were predicted to cause a loss of function (frameshift, nonsense, splice). The remainder of the 244 coding changes were missense variants without a reported phenotype. The 92 variants identified by WES, and targeted by the Affymetrix chip were categorized in HGMD® as follows: functional polymorphism (n=66), disease-associated polymorphism (n=8), disease-associated polymorphism with additional supporting functional evidence (n=11) and disease-causing mutation (n=7). Variants with potential medical relevance include VKORC1 p.Asp36Tyr reported to predispose to warfarin resistance (n=5 individuals), SLCO1B1 p.Val174Ala associated with higher basal cholesterol (n=20 individuals) and CYP2C8 p.Ile264Met associated with an increased risk of coronary artery disease among smokers (n=67 individuals). Three hundred and sixty variants were identified in 20 genes associated with drug metabolism. WES identified insertion/deletion variants, but not promoter or deep intronic SNPs. Ninety-two variants have published data suggestive of a role in drug metabolism. Further research is needed to determine which of these variants have clinical utility thereby meriting the return of genetic results to participants.

727W

High Throughput Scaling and Performance Assessment of Gene Panel Capture with Molecular Inversion Probes. *L.S. Felker, A. Gordon, M.B. Beightol, T. Shaffer, D.R. Crosslin, J.D. Smith, D.A. Nickerson.* Genome Sciences, University of Washington, Seattle, WA.

Molecular Inversion Probes (MIPs) are a customizable, scalable and affordable method to perform massively parallel target capture and sequencing with a minimal amount of genomic DNA input. With the intention of streamlining and scaling a MIP capture platform to process thousands of samples with any given panel of gene candidates, we characterized the capture efficiency of a 10-gene panel of pharmacologically significant genes that contain a spectrum of challenging capture targets; highly paralogous, repetitive and GC-rich (i.e., CYP2D6) to unique sequence (i.e., COMT). All MIPs in our 10 gene panel were pooled at equimolar ratios and captured at an initial genomic DNA target to MIP ratio of 1:200. Due to the complexity of the gene panel, we observed non-uniform MIP capture across our targets requiring rebalance of the MIP pool to increase the likelihood of capture for inefficient MIPs. To eliminate the need for time-consuming successive pool rebalancing for individual samples, we assessed capture repeatability in two ways: intra-sample (using 12 replicates of a single sample) and inter-sample (across four HapMap trios that represent diverse ancestry and CYP2D6 genotype). Pairwise comparisons between capture reactions from both repeatability assessments show high correlation. Utilizing the normalized MIP capture events, we developed a method to rebalance the 10 gene panel MIP pool to compensate for poorly performing probes. We observed that MIP performance correlates with the nature of the gene complexity and the sample ancestry and diversity. Additionally, we included a 96 'fingerprint' target MIP pool to serve as a capture efficiency control, for sample identity confirmation and to provide quality control monitoring for sample contamination. In summary, we have optimized the MIP capture platform by eliminating successive pool rebalancing, simplifying reagents, and incorporating robotics for making and rebalancing probe pools to increase throughput.

728W

Pharmacogenetic Genotyping of Clinical Specimens using Next-Generation Sequencing. *S.M. Jacobs-Helber, T. Reynolds.* American International Biotechnology Services, LLC, Richmond, VA.

An outcome of the era of personalized medicine has been more frequent genotyping of Cytochrome P450 and other risk factor genes of patient samples to assist physicians in drug dosing as well as the determination of risk based on genetic factors. Such analysis requires high throughput methods capable of multiplex analysis that can generate results for patients in a timely manner. AIBiotech has developed a multiplex pharmacogenomics panel using Next Generation Sequencing technology that assists physicians in the treatment of their patients. The Personalized Medicine Panel (PMP) consists of 10 genes including the Cytochrome p450 genes CYP2C19, CYP2C9(with VKORC1), CYP2D6, CYP3A4, and CYP3A5, as well as the cardiac risk factors Factor II (prothrombin) Factor V Leiden, APOE (for cardiovascular risk) and MTHFR. The AIBiotech PMP delivers genotype data on all 10 markers, giving clinicians actionable data to assist in the therapeutic treatment of their patients and reduce the risk of adverse events. After only a year of testing AIBioTech has processed more than 20,000 patient samples on this platform. Using data generated from these samples, AIBiotech will present genotyping frequencies on all markers for >10,000 deidentified clinical specimens tested. Comparisons will be made between published frequencies from smaller studies in four different ethnic populations (African-American, Caucasian, Hispanic, and Asian). The data generated from this study will generate new insight in to the ethnic frequency of mutations that affect drug dosing and management.

729W

Development of a scoring tool to prioritize clinical pharmacogenomic testing. *S. Manzi, Clinical Pharmacogenomics Oversight Committee.* Research Connection, Boston Children's Hospital, Boston, MA.

A scoring tool was developed to assist in the process of prioritizing drug/gene pairs for clinical implementation of pharmacogenomic testing. The provision of clinical interpretation of genotype data in the electronic medical record (EMR) along with decision support at the point of ordering and dispensing is complex and requires resources to build and implement. Once the capability to provide pharmacogenomic testing with EMR decision support becomes publicized, an increasing number of requests for testing for a specific service or patient population are filed. Competing demands for IT support, balancing the needs of a single patient population for whom the information will be critical against the needs of many for whom the information will be important, providing a comprehensive test, and ensuring cost containment are some of the many factors that we incorporated into the tool. Additionally, factors such as commercial availability of the test, published knowledge base tools and guidelines and the projected utilization were also included. The tool was designed from review of the available literature and revised based upon input from our Clinical Pharmacogenomics Service Oversight committee members. We will demonstrate a use case for the tool via an example of prioritizing between TPMT genotyping for mercaptopurine use in Acute Lymphoblastic Leukemia (ALL) and cyp2D6 genotyping for opioid drugs in acute pain management. The overall score for TPMT (thiopurine S-methyltransferase) genotyping was 27 points compared with a score of 20 points for cyp2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6) genotyping. The major dissenting areas included delayed toxicity (present for thiopurines, not present for opioids), overlapping toxicities with concomitant agents (present for thiopurines, not present for opioids), frequency of use (limited for thiopurines, extensive for opioids) and the commercial availability of a comprehensive test at the time of evaluation (present for TPMT, not present for cyp2D6). To our knowledge, no other decision tool designed to assist with prioritization of movement of drug/gene pair data from research to the clinical realm has been published. We will demonstrate the tool and scores relevant to our program.

730W

Assessment of the predictive effect of genetic variation in key genes associated with drug therapy of cardiovascular diseases in the Azores Islands (Portugal). *L. Mota-Vieira[1,2,3], M.S. Melo[1,4], L. Balanco[1,4], R. Cabral[1], C.C. Branco[1,2,3], T. Pereirinha[1].* 1) Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo of Ponta Delgada, EPE, Azores, Portugal; 2) Instituto Gulbenkian de Ciência, Lisbon, Portugal; 3) BioFig, Lisbon, Portugal; 4) The University of the Azores, Ponta Delgada, Portugal.

New understandings of individual genetic variants are increasingly reshaping, in various aspects, the drug treatment of cardiovascular diseases (CVD). Three highly prescribed drugs - clopidogrel (platelet aggregation inhibitor), warfarin (anticoagulant) and simvastatin (cholesterol-lowering drug) - are examples with strong genetic evidence. Here, we investigate the predictive effect of relevant pharmacogenes - *CYP2C19* (clopidogrel bleeding risk), *CYP2C9* and *VKORC1* (warfarin bleeding risk), and *SLCO1B1* (simvastatin-induced myopathy risk, from myalgia to rhabdomyolysis) - associated with drug response in Azoreans. This population could benefit from these pharmacogenetic data, since CVD are the first cause of mortality and morbidity in Azores Islands. According to SNP information available on PharmGKB, Pubmed, dbSNP and SNPedia databases, DNA of 170 blood donors was genotyped by TaqMan® genotyping assays (Applied Biosystems). Our results demonstrate that in Azoreans the influence of the *CYP2C19* loss-of-function alleles (**2* and **3*) on clopidogrel response is only due to the **2* allele. Therefore, clinicians should prescribe a non-*CYPC19* dependent thienopyridine, such as prasugrel or ticagrelor, in homozygous (1.9%) and prefer, if possible, the use of these alternative medications in heterozygous (21.8%). These individuals correspond to the poor and intermediate clopidogrel responders, respectively. Regarding warfarin, the joint analysis of *CYP2C9* (**2* and **3*) and *VKORC1*3* revealed that 82.4% of Azoreans will need intermediate or low drug doses, guided by a clinical algorithm, if treatment is started. In what concerns the *SLCO1B1* c.521T>C SNP, known to cause high plasma levels of simvastatin, three genotypes were observed: TT (individuals with two functional alleles, 70.6%), TC (one functional allele plus one reduced-function allele, 27.6%) and CC (two reduced-function alleles, 1.8%). Thus, clinicians should be alert to simvastatin-induced myopathy in patients with intermediate and high risk (TC and CC, respectively) and, whenever needed, proceed to dose adjustment or prescribe another drug in order to achieve desired lipid levels. In conclusion, this population study constitutes an initial step in how clinicians can use the patient's genetic makeup to provide a better and safer CVD therapy. (lmotavieira@hdes.pt; funded by the Government of the Azores).

731W

Sequence-based characterization of ADME variation in two Northeast Asian and 19 1000 Genomes Project Populations. *D.H. Hovelson[1], Z. Xue[2], M. Zawistowski[1], M.G. Ehm[2], S. Stocker[2], A. Gross[2], I.J. Jang[3], I. Ieiri[4], J.E. Lee[5], L.R. Cardon[2], S.L. Chissoe[2], G. Abecasis[1], M.R. Nelson[2].* 1) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, NC; Sydney, AUS; Stevenage, UK; 3) Seoul National University, Seoul, Korea; 4) Kyushu University, Fukuoka, Japan; 5) DNA Link, Inc., Seoul, Korea.

Identifying factors which influence the pharmacokinetics of a drug (the body's exposure to the drug) and selecting optimal doses is a crucial part of drug development. Variation in genes that regulate drug absorption, distribution, metabolism and excretion (ADME) can contribute to variability in drug pharmacokinetics; in addition, the frequency of variants in ADME genes may differ between ethnic groups. Despite this, there have been few systematic studies to identify and characterize the extent of variation in ADME genes in different ethnic groups. To address this, we called single nucleotide variants (SNVs), short insertion-deletions (indels) and gene copy number variants (CNVs) in exome sequence data from healthy Northeast Asian (NEA) subjects (Korean N=126; Japanese N=125) and subjects from 19 different ethnic groups in the 1000 Genomes Project (1000G; N=1181). Average depth across the core (N=38) and extended (N=266) ADME genes (as defined by PharmaADME.org) is ~60X and ~130X for the NEA and 1000G subjects, respectively. Approximately 75% of ADME coding bases were well-captured (sequenced to an average depth >20X) in the combined sample (NEA+1000G). NEA subjects carried on average approximately 180 nonsynonymous (NS; 15% predicted to be deleterious) and 3 loss-of-function SNVs across all ADME coding regions. On average, each NEA subject carried ~9 NS ADME coding SNVs currently uncharacterized in public databases of functional ADME variation, of which 90% are predicted to be deleterious. NEA subjects carried on average ~1 frameshift indel in ADME coding regions and exhibit evidence of potentially novel disruptive gene deletions or duplications in UGT and CYP gene families. Restricting focus to well-captured ADME regions across the combined sample, NEA subjects carried slightly fewer NS variants (~135 SNVs per person) than 1000G subjects (Asian (ASN): ~136; European (EUR); ~141; African (AFR): ~174). Quantities of uncharacterized NS ADME variation predicted to be deleterious were similar in the NEA, ASN and EUR subjects (2.0, 2.2 and 2.0 per person, respectively) and higher for the AFR subjects (4.5 per person). Follow-up experiments to validate uncharacterized ADME variation (including SNVs, indels, and CNVs) are being performed. Ultimately, these results will help enhance understanding of ADME gene variation across ethnic groups which has not been captured in existing genotyping panels and which warrants further functional characterization.

732W

The *CYP2C1917 Variant is not Independently Associated with Clopidogrel Response.** *J. Lewis[1, 2], S. Stephens[1,2], R. Horenstein[1,2], J. O'Connell[1,2], K. Ryan[1], C. Peer[3], W. Figg[3], S. Spencer[4], M. Pacanowski[5], B. Mitchell[1,2], A. Shuldiner[1,2,6].* 1) Endocrinology, Diabetes & Nutrition, University of Maryland, Baltimore, Baltimore, MD; 2) Program in Personalized and Genomic Medicine, University of Maryland, Baltimore, Baltimore, MD; 3) Clinical Pharmacology Program, National Cancer Institute, Bethesda, MD; 4) Applied and Developmental Research, SAIC-Frederick Inc., National Cancer Institute, Frederick, MD; 5) U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Office of Clinical Pharmacology, Silver Spring, MD; 6) Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD.

Background: Cytochrome P450 2C19 (CYP2C19) is the principle enzyme responsible for converting clopidogrel into its active metabolite and genetic variants have been identified, most notably *CYP2C19**2 and *CYP2C19**17, that are believed to alter its activity/expression. We evaluated whether the consequences of the *CYP2C19**2 and *CYP2C19**17 variants on clopidogrel response were independent of each other or genetically linked through linkage disequilibrium (LD). **Methods:** We genotyped the *CYP2C19**2 and *CYP2C19**17 variants in 621 members of the Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study and evaluated the effects of these polymorphisms singly then jointly, taking into account LD, on clopidogrel prodrug level, clopidogrel active metabolite level, and ADP-stimulated platelet aggregation pre- and post-clopidogrel exposure. **Results:** The *CYP2C19**2 and *CYP2C19**17 variants were in LD (lD'l=1.0; r2=0.07). In association analyses that did and did not account for the effects of *CYP2C19**17, *CYP2C19**2 was strongly associated with levels of clopidogrel active metabolite (beta=-5.24, P=3.0×10^{-9} and beta=-5.36, P=3.3×10^{-14}, respectively) and post-treatment ADP-stimulated platelet aggregation (beta=7.55, P=2.9×10^{-16} and beta=7.51, P=7.0×10^{-15}, respectively). In contrast, *CYP2C19**17 was associated with clopidogrel active metabolite levels and ADP-stimulated platelet aggregation before (beta=1.57, P=0.04 and beta=-1.98, P=0.01, respectively) but not after (beta=0.40, P=0.59 and beta=-0.13, P=0.69, respectively) adjustment for the *CYP2C19**2 variant. Stratified analyses of *CYP2C19**2/*CYP2C19**17 genotype combinations revealed that *CYP2C19**2, and not *CYP2C19**17, was the primary determinant in altering clopidogrel response. **Conclusions:** Our results suggest that *CYP2C19**17 has a small (if any) effect on clopidogrel-related traits and that the observed effect of this variant is due to LD with the *CYP2C19**2 loss-of-function variant.

733W

A pharmacogenomic approach to targeted medicine. *A. Albayrak, J.D. Hogan, A.P. Peter, B.R. Braun, F.D.G. Schacherer.* Pharmacogenetics, BIOBASE International, Beverly, MA.

The advent of next-generation sequencing technology enables comprehensive screening of an individual's genomic variations. However, due to the millions of variants discovered in each patient, the interpretation of this data poses a challenge, particularly in a clinical context. To address this challenge, we have developed a manually curated database of pharmacogenomic variants from the scientific literature. The aim of the database is to provide a comprehensive resource for all variants that have been reported as significant in a pharmacogenomic context in human studies. The database is designed to provide rich information as evidence for these associations, including information on the exact genomic location and sequence changes, resulting phenotype, drugs administered, patient population, study design, disease context, statistical significance, and a link to the reference from which the information was taken. The database is available through an online search interface, or for download for integration into existing tools and data processing pipelines. To make the database useful for exome or whole genome screening, we have developed algorithms that allow matching of the entries in the database against a sampled subject's variants, taking into account that in many cases, haplotypes need to be matched, exact nucleotide changes must be considered, and complex star alleles must be resolved properly. The online user interface enables the database to be easily searched by drug, disease, gene, haplotype or variation, and also provides information on SNPs that are in linkage disequilibrium with reported pharmacogenomic variants. Detailed reports on each of these data types are presented for in-depth review.

734W

Exome sequencing of asthmatics with extreme corticosteroid response. *Q.L. Duan[1], M.H. Cho[1], R. Kelly[1], W. Qiu[1], E.R. Mardis[4], E.R. Bleecker[5], D.A. Meyers[5], G.A. Hawkins[5], S.P. Peters[5], J.J. Lima[6], K.G. Tantisira[1,2], S.T. Weiss[1,3].* 1) Channing Div Network Med, Brigham & Women's Hosp, Boston, MA; 2) Pulmonary Division, Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 3) Partners Center for Personalized Genetic Medicine, Partners Health Care, Boston, MA; 4) The Genome Institute at Washington University School of Medicine, St. Louis, MO; 5) Center for Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC; 6) Center for Pharmacogenomics and Translational Research, Nemours Children's Clinic, Jacksonville, FL.

Corticosteroids (CS) are potent, anti-inflammatory drugs used for the treatment of numerous diseases including asthma, which affects over 300 million individuals worldwide and 20 million Americans. However, great inter-individual variability in response to this class of drugs has been described, whereby up to 25% of patients are non-responsive. Previous pharmacogenetic studies of CS response focused primarily on the contribution of common polymorphisms, whereas increasing evidence suggests that rare variants are important genetic determinants of complex traits and confer larger effect sizes. We hypothesize that novel rare variants, identified through next-generation sequencing of whole exomes, contribute to the heterogeneity of CS response in asthma patients. To test this, we selected 196 individuals including equal numbers of poor and good responders to inhaled CS, measured as ≤0% and ≥13% change in FEV_1 (forced expiratory volume in one second) following 4 to 8 weeks of therapy. Exome sequencing was performed using the NimbleGen SeqCap EZ v2.0 array on the Illumina Hi-Seq 2000. Sequences were aligned using BWA, and variants called using Samtools and VarScan. We identified 403,373 single nucleotide variants (SNVs), of which 347,802 passed quality control in 188 individuals (91 poor and 97 good responders). Of these, 75,206 coded for non-synonymous, splice, or stop variants, the majority of which were rare (60,725 or 81% with MAF < 0.05). Our primary analysis in SKAT-O was a per-gene burden test of these rare variants, weighted by MAF, and adjusted for gender, age, and height. Our top genes (P-values between 2.8 to 9.2×10^{-4}) were *FBXO6*, *CD163*, *ZDBF2*, *ARHGEF16*, *PCK1*, *FOXK1*, *CASC5*. Although none of these were exome-wide significant after correction for the number of genes tested, several of the top loci are differentially expressed in response to dexamethasone (a corticosteroid) in the lymphoblastoid cell lines of asthmatics. Furthermore, *CD163* and *PCK1*, are excellent biological candidates that are known to be regulated by glucocorticoids. In conclusion, exome sequencing of extreme CS response in asthmatics may identify novel loci that contain rare variants that could predict a CS response phenotype. A secondary analysis using Illumina Exome Chip data in nearly 400 asthma trios is under way.

735W

Two-Hit Models of Disease Risk in Drug-Induced Liver Injury. *J.I. Goldstein[1,2], E.T. Lim[1,2,3], H. Huang[1,2], S. Raychaudhuri[2,4,5,6], C. Stevens[2], P.I.W. de Bakker[2,7,8], M.I. Lucena[9], A.K. Daly[10], M.R. Nelson[11], A. Holden[12], B.M. Neale[1,2], M.J. Daly[1,2], International Severe Adverse Events Consortium.* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 5) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA; 6) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA; 7) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 8) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 9) Hospital Universitario Virgen de la Victoria, Malaga University, IBIMA and Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain; 10) Institute of Cellular Medicine, Newcastle University Medical School, Newcastle Upon Tyne, UK; 11) Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, NC; 12) International Severe Adverse Events Consortium, Ltd., Chicago, IL.

Genome-wide association studies for drug-induced liver injury (DILI) have found strong associations with large effect sizes to classical human leukocyte antigen (HLA) alleles. However, the frequencies of the risk alleles in the treated population are three orders of magnitude higher than the prevalence of DILI (approximately 8/100,000), suggesting a role for other genes in DILI. It has been proposed that part of the missing heritability in complex diseases can arise from non-additive genetics, such as epistasis. Therefore, we hypothesize that there might be a two-hit model involved in conferring risk for DILI. To test this hypothesis, we genotyped samples with two DILI phenotypes (flucloxacillin DILI: n=69; co-amoxiclav DILI: n=124) and ancestry-matched controls (n=3159) on the Illumina HumanExome array. When we imputed the classical HLA alleles and performed single-variant association tests, we obtained similar results to what had previously been reported [flucloxacillin: HLA-B*57:01 (p= 3e-35, OR=50); co-amoxiclav: HLA-DQB1*06:02/HLA-DRB1*15:01 (p=3e-18, OR=3.4)]. However, single-variant tests for rare variants did not yield any significant associations, suggesting we are underpowered to detect rare risk variants with moderate effect sizes. To test the two-hit model of disease, we calculated binomial probabilities of observing an excess of cases with rare alleles at two unlinked genes and then calculated empirical p-values using permutation testing to control for population stratification. We also implemented a conditional scanning approach by incorporating what is known about the biology of DILI from the literature. For example, Monshi et al. showed flucloxacillin covalently binds to lysine residues in albumin and three of these residues occur in amino acid sequences that are predicted to bind to HLA-B*57:01 (Hepatology, 2013). Therefore, we looked for sets of variants that mimicked this scenario in other genes expressed in the liver and found three variants that look promising for flucloxacillin DILI in *CYP4V2*, *FMO3*, and *CD68*. In this work, we demonstrate novel approaches for testing a two-hit model in a complex disease and present results that demonstrate the utility of testing such a model in pharmacogenetic studies.

736W

Pharmacogenomics of glucuronidation in American Indians: The Strong Heart Family Study. *PE. Melton[1], N. Franceschini[2], K. Haack[3], C. Bizon[4], ET. Lee[5], JG. Umans[6], LG. Best[7], SA. Cole[3], LA. Almasy[3].* 1) Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, Western Australia, Australia; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC; 5) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 6) Medstar Health Research Institute, Hyattsville, MD; 7) Missouri Breaks Research Industries, Inc, Timber Lake, SD.

Glucuronidation is a phase II (conjugation) process facilitating both biliary and renal clearance of both drugs and endogenous substances. Bilirubin serves as an endogenous glucuronidation biomarker, as its levels are affected by the UDP glucuronisyl transferase (UGT) and organic anion transporter (OAT) products encoded by the genes *UGT1A1* and *SLCO1B1*, respectively. Polymorphisms in these genes are also known to impact clearance of several cardiovascular (CVD), and antineoplastic drugs. Previous research on genes influencing bilirubin in American Indians from the Strong Heart Family Study (SFHS) detected a QTL (LOD=6.61) on chromosome 2q near the *UGT1A1* gene, but linkage conditional on an associated repeat promoter polymorphism, rs5839491, did not completely explain the linkage signal. Genetic testing of this particular variant is used routinely to guide dosing of irinotecan, an antineoplastic agent. To elucidate additional functional variants influencing glucuronidation and bilirubin metabolism, we conducted whole exome sequencing (Illumina TruSeq Exome Enrichment Kit) in SHFS participants from Arizona (n=47) and Oklahoma (n=47). Genetic diversity was maximized by selecting founders from distant relatives (Arizona) or by principal component analysis (OK). Average sequence coverage was 36x and 263,378 variants passed quality control. Using a priori evidence we identified 398 variants in 31 genes from the UGT family and 7 genes involved in bilirubin metabolism. Of these variants, 83 were non-synonymous and 8 were novel. These included the *UGT1A1* exon variant, rs4148323, which was detected in 4% of sequenced SFHS participants and was significantly associated with total bilirubin (p=0.0004). This SNP is known to influence bilirubin metabolism and is associated with increased irinotecan toxicity in East Asian populations but is absent in Europeans. We also detected the exonic *SLCO1B1* SNP rs4149056 in 4% of these SHFS participants, which is known to impact statin metabolism, effect, and toxicity. This variant was not significantly associated with total bilirubin (p=0.66) in these SHFS participants. These results demonstrate the efficacy of whole exome sequencing to identify novel pharmacogenetic markers in American Indians and may have implications for recommended testing to decrease adverse drug events in this population.

737W

Primary action of clozapine exposure on activation of SREBP-controlled lipogenic gene expression may explain benefit and detriment.

J.S.A. de With[1,2], T. Wang[2], E. Strengman[2], S. de Jong[2], W.G. Staal[3], R.A. Ophoff[2]. 1) Donders Institute for Brain, Cognition and Behavior, Department of Cognitive Neuroscience, Radboud University Medical Center Nijmegen, Utrecht, Netherlands; 2) UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, California, USA; 3) Karakter, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands.

Clozapine is an atypical antipsychotic drug with superior efficacy in the treatment of treatment resistant schizophrenia. However, its use is complicated by metabolic adverse effects, weight gain and clozapine-induced agranulocytosis (CIA), a potential lethal adverse effect occurring in approximately 1% of patients. This study aimed to investigate the mechanism behind the (adverse) effects of clozapine by studying gene expression profiles and the genetic background of response variation in clozapine exposed lymphoblast cell lines. We performed two experiments to study the molecular basis of clozapine exposure at the cellular level. First, we established gene expression profiles of four lymphoblast cell lines that were exposed to a range of concentrations of clozapine (2µM to 100µM) for 24h. Secondly, we exposed 90 lymphoblast cell lines (HapMap CEU trios) to a range of concentrations of clozapine (2µM to 140µM) for 48h to create survival curves. The survival of cells for each cell line is used as quantitative trait measure for genetic mapping. Available gene expression data of these cell lines will be used to gain insight in molecular mechanisms behind differences in cell survival in response to clozapine exposure. We observed significant gene expression changes after exposure to clozapine in lymphoblast cell lines. Using a network approach we found a gene co-expression module significantly associated with clozapine exposure. Gene enrichment analyses showed very strong activation of Sterol Regulatory Element Binding Protein (*SREBP*) together with *SREBP* target genes involved in sterol biosynthesis. Cell survival curves of the 90 cell lines showed individual differences suggestive of genetic factors playing a role in clozapine response, which may be important for identification of agranulocytosis susceptibility loci; genetic analysis is ongoing. The observed gene expression upregulation of genes from the *SREBP* pathway is a recurrent theme in different cell types including neuronal, liver and and blood-derived lymphoblast cells. The consistent finding of activation of the *SREBP* system in different cell types is suggestive of a systematic and primary lipogenic molecular underlying mechanism. Accordingly, this atypical APD could have beneficial effects on myelination in the brain and simultaneously cause metabolic side effects. The results of the cytotoxicity assay could elucidate possible mechanisms behind CIA, enabling prediction of CIA susceptibility.

738W

A Phenome-Wide Association Study of ADME Core Variants in an EMR-linked Biobank.

M.T. Oetjens[1], J.C. Denny[2], N.B. Gillani[3], M. Herrera[1], L. Olson[1], H.H. Dilks[1], D.M. Richardson[1], E. Bowton[4], D.M. Roden[3,5,6,7], D.C. Crawford[1]. 1) Center for Human Genetics Research and Departments of Molecular Physiology and Biophysics; 2) Biomedical Informatics; 3) Department of Pharmacology; 4) Office of Research; 5) Department of Medicine; 6) Oates Institute for Experimental Therapeutics; 7) Office of Personalized Medicine, Vanderbilt University, Nashville, TN.

Many of the genes that participate in drug metabolism harbor deleterious variants at high frequencies compared with the gene networks of other disease pathways. For instance, a functional copy of the CYP2D6, a PharmGKB designated Very Important Pharmacogene, is absent in 7% of individuals from European-descent populations. Other common alleles impacting drug metabolism are found in a continuum between abolished function and gene amplification. The high frequency of functional variants in CYP2D6 and other pharmacogenes is a partial explanation to the stronger effect sizes detected in pharmacogenetic studies designed to identify associations between genes and drug response when compared to those of complex disease. Yet the effect of many of the deleterious variants in pharmacogenes have not been explored in the context of pleiotropy because they are often found in regions of the genome, that for technological reasons, are difficult to genotype accurately and are often missing from fixed-content genotyping platforms. As part of the Vanderbilt Electronic Systems for Pharmacogenomics Assessment (VESPA) 'Grand Opportunities' stimulus grant from the NIH, ~10,000 DNA samples linked to de-identified electronic medical records (EMR) have been genotyped on the ADME Core Panel, a genotyping platform that with specialized assays that accurately target 184 of functional variants across 34 VIPs. The demographics of our cohort are as follows: 83% of the individuals are of European descent, 12% are of African descent, 49% are female, and the median BMI is 27.4kg/m^2. We are applying a reverse genetics methodology, termed phenome-wide association study (PheWAS), to explore of pleiotropic relationships and undiscovered mechanisms of action of these clinically important genes. Our analysis includes a broad range of phenotypes that can be derived from ICD9 code-derived phenotypes, ranging from chronic diseases to environmental effects and/or injury.

739W

Swedegene: Genome-Wide Association Studies of Adverse Drug Reactions.

M. Wadelius[1], N. Eriksson[2], Q. Ying-Yue[3], E. Eliasson[4], H. Melhus[1], P. Hallberg[1]. 1) Department of Medical Sciences, Clinical Pharmacology, Uppsala University, Uppsala, Sweden; 2) Uppsala Clinical Research Center and Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) Medical Products Agency, Uppsala, Sweden; 4) Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.

Background: Adverse drug reactions (ADRs) are common reasons for hospitalisation of adults. Most ADRs are dose-dependent and pharmacologically predictable (type A). Others are non-predictable and have no known pharmacological cause (type B). They are on the other hand often serious and sometimes lead to withdrawal of the drug from the market. The knowledge about possible genetic causes of type B reactions is small. **Objective:** Swedegene wishes to discover genetic and clinical factors that predispose patients to ADRs with the aim to prevent them in the future. **Method:** We are assembling a biobank at Uppsala University (www.swedegene.se) by contacting patients with ADRs reported to the register at the Swedish Medical Products Agency. Consenting patients are interviewed concerning their medical and drug history, and a blood sample is drawn. Supplementary information is collected from their medical records. We have initially chosen to collect drug-induced liver reactions, severe skin reactions, agranulocytosis, long QT syndrome/torsade de pointes, CNS toxicity, neuropsychiatric reactions, statin myopathy, fluoroquinolone-induced tendon rupture and phototoxicity, angioedema and cough related to ACE-inhibitors or angiotensin receptor blockers, osteonecrosis due to bisphosphonates, weight gain due to neuroleptics and antidepressants, and hyponatremia due to antidepressants and antiepileptic drugs. When a sufficient number is collected, genetic analyses are performed, mainly in the form of genome-wide association studies (GWAS). We have access to 5000 unrelated population controls with GWAS data through collaboration with the Swedish Twin registry. **Results:** Swedegene started in 2010 and currently has >1100 cases with DNA and complete phenotypes. We have promising GWAS results on Swedish cases with drug-induced agranulocytosis and on cases with angioedema and cough related to ACE-inhibitors or angiotensin receptor blockers. However, we need more samples for conclusive results. **Conclusion:** Serious ADRs are often too rare to study in a single country, even using a nationwide register. We therefore collaborate with researchers abroad and invite new collaborators interested in collecting cases with the aim to detect genetic and clinical risk factors for ADRs.

740W

Pharmacogenetics at NCBI.

A.J. Malheiro, W. Rubinstein, B. Kattman, J. Lee, D. Maglott, V. Hem, M. Ovetsky, G. Song, K. Katz, C. Wallin, R. Villamarin, J. Ostell. National Center for Biotechnology Information (NCBI), NLM, NIH, Bethesda, MD.

Medical professionals are beginning to practice precision medicine by ordering pharmacogenetic tests to guide drug selection and dosing. The National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH) offers free access to comprehensive pharmacogenetics information via user-friendly websites tailored to clinicians and researchers. We will focus on the Genetic Testing Registry (GTR) and ClinVar. The GTR is a centralized, international registry of genetic test information covering clinical and research tests for Mendelian disorders and drug responses. The GTR is a primary resource for locating detailed, practical information about pharmacogenetic tests and the indications for which tests are performed (www.ncbi.nlm.nih.gov/gtr). As of June 2013, laboratories have voluntarily provided information for 14 pharmacological responses (e.g., abacavir, clopidogrel, warfarin). Registered tests describe target patient population, indications, gene/variant targets, and ordering details. NCBI staff work with authoritative groups including the Pharmacogenomics Knowledge Base (PharmGKB), and the Clinical Pharmacogenetics Implementation Consortium (CPIC). Customized links are provided to practical information in PharmGKB such as dosing guidelines and clinical annotations of gene variants. When available, succinct, curated synopses of drug responses by CPIC are presented. Practice guidelines from CPIC, the Pharmacogenetics Working Group of the Royal Dutch Association for the Advancement of Pharmacy (KNMP), and other authorities facilitate the translation of genetic test results into clinical actions for specific drugs. GTR also links to the FDA labels from NLM's DailyMed. The lack of structured, open-access, peer-reviewed, comprehensive gene-drug information is a deterrent to clinical implementation. To fill this gap, NCBI collaborates with CPIC and other experts to create Medical Genetics Summaries, available from NCBI's Bookshelf (http://www.ncbi.nlm.nih.gov/books/NBK61999/). These summaries display in GTR condition and test pages and are supplemented with links to PubMed and OMIM. GTR also provides information about the genes involved in the drug response and variants that affect drug metabolism. Clinicians and researchers can access ClinVar (www.ncbi.nlm.nih.gov/clinvar) for a comprehensive dataset of clinically relevant variants observed in humans, their phenotypic relationships, current interpretations, and supporting evidence.

741W

AlleleTyper Software: a flexible application for mapping SNP genotype and CNV data patterns to pharmacogenetic allele nomenclature. *T. Hartshorne, N. Mehmet, E. Shelton, H. Leong.* Genetic Analysis, Life Technologies, South San Francisco, CA.

Pharmacogenetic (PGx) studies require genetic testing of individuals for multiple variants in drug metabolism enzyme and transporter genes. For phenotype interpretation purposes, genotyping results must be translated to 'star allele' nomenclature. Star alleles are haplotype patterns that have been defined at the gene level and, in many cases, associated with protein activity. Genetic variants within a haplotype can include SNPs, InDels, and copy number variations. Knowing the combination of variants within a given haplotype, and the diploid content in an individual, is of key importance for studying drug metabolism, drug response and adverse drug reactions. To facilitate the translation of results for individuals genotyped in studies using TaqMan™ Drug Metabolism Genotyping Assays and TaqMan™ Copy Number Assays, we developed a web-based flexible software tool called AlleleTyper™. This software uses genetic pattern information in user-defined translation tables to map sample genotyping data to star allele nomenclature. Typically, haplotype information in public resources such as the Cytochrome P450 Allele Nomenclature or PharmGKB databases is used to create a translator for gene alleles of interest. A translator can include one or more genes. A monoallelic translation table is first prepared that contains the star allele/haplotype pattern information for each TaqMan assay in a study. This is imported into AlleleTyper, which automatically converts it to a biallelic translator containing diploid genetic pattern information. The biallelic translator is reviewed and edited, if needed. Subsequently, a study is created in AlleleTyper, the biallelic translator is imported, and then genotype results that have been exported from TaqMan Genotyper Software and/or CopyCaller Software are also imported. AlleleTyper matches the genetic information in these files to the patterns in the biallelic translator and reports the star allele genotypes determined for each individual. The software includes error reporting for missing data, unmatched patterns, etc. A review of the software workflow and features will be presented, along with data analysis examples. AlleleTyper Software greatly facilitates PGx study data analysis, particularly for high throughput studies. This software is also flexible enough to be used for other genotyping applications requiring translation of data from multiple TaqMan assays, including triallelic SNP data analysis and blood genotyping.

742W

Genetic ancestry modifies effects of naltrexone on smoking cessation in African Americans: an analysis of a randomized controlled trial. *A.P. Bress[1], C. Wing[2], A. King[3], R. Kittles[4].* 1) Pharmacy Practice, University of Illinois at Chicago, IL; 2) Health Policy and Administration, University of Illinois at Chicago; 3) Department of Psychiatry & Behavioral Neuroscience, University of Chicago; 4) Department of Medicine, University of Illinois at Chicago.

Nearly 1/5 Americans smoke cigarettes and 443,000 Americans die of smoking every year. Several pharmacotherapies substantially increase smoking cessation rates. However, treatment effects vary across ethnic groups, especially African Americans (AA). To determine if genetic background explains these differences between AA and European Americans, we studied the effects of naltrexone therapy on smoking cessation rates among AA patients with different levels of west African (WA) genetic ancestry. We used data from a previously published randomized, double blind trial of naltrexone vs. placebo for smoking cessation in 315 smokers from Chicago. Ninety five participants were AA and provided a DNA sample. We genotyped 105 autosomal DNA ancestry informative markers using the SEQUENOM iPLEX MassArray platform. Individual ancestry was estimated for all subjects using STRUCTURE. We used logistic regression models to estimate the effects of naltrexone on 4-week quit rates for AAs with high and low WA ancestry. The models adjusted for baseline clinical and demographic characteristics and confidence intervals and p-values were estimated using the bootstrap. In the total pooled sample of AAs, 4-week quit rates were not statistically different between naltrexone (41%) and placebo (33%) groups (p=0.44). However the pooled analysis masks treatment effect heterogeneity by WA ancestry. The median WA ancestry among the 95 AAs in our sample was 80%, so we defined subjects above the median as high WA ancestry and those below median as low WA ancestry. Among those AAs with low WA ancestry, quit rates were higher in the naltrexone group than the placebo group (59% vs. 29%) so that the treatment effect for those AAs with low WA ancestry was 30 percentage points (95%CI -0.02 to 0.58, p=0.04). In contrast, for those AAs with high WA ancestry, quit rates were actually lower in the naltrexone group (26 vs. 39%) so that the treatment effect was -13 percentage points (95%CI -0.41 to 0.13 p= 0.33). The difference in naltrexone treatment effects on quit rates across high and low WA ancestry groups was statistically significant (p=0.04). Naltrexone efficacy on quit rates varies across AA subjects with different levels of WA ancestry. Our data suggest naltrexone is effective for AAs with low WA ancestry and is not effective in AAs with high WA ancestry. These results suggest that genetic factors may partially explain racial differences in treatment effects.

743W

Assessment of viability of human lymphocytes exposed to ionizing radiation and curcumin. *B. Gonzalez[1], C. Monterrubio[1], H. Pimentel[1], R. Silva[1], L. Rodríguez[1], C. Barba[1], N. Pérez[1], B. Inda[1], J. Vidal[1], E. Flores[1], H. Reyes[1], L. Bobadilla[1,2], A. Corona[1,2].* 1) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Unidad de Citogenética, Serv. Hematol. Oncol. Pediát., Div. Pediatría, Hosp. Civil Dr. Juan I. Menchaca; Guadalajara, Jalisco, Mexico.

Introduction: Curcumin is a polyphenol with antioxidant, anti-inflammatory and antineoplastic properties. In vitro it inhibits tumoral cell proliferation but in contrapart there are reports of cytotoxicity and genotoxicity effects on healthy human lymphocytes. Ionizing radiation induces cell damage and genotoxicity. Cytotoxicity can be assessed on individual cells by trypan blue testing or flow cytometry or assessment on cells poppulation under XTT assay. Statement of Purpose: To evaluate the effect of curcumin on human lymphocytes viability under the effect of ionizing radiation with three viability assessment methods. Methods Used: Peripheral blood lymphocytes from 5 healthy individuals where obtained and cultured for 72hrs, under the next conditions: 1) control group, 2) 25μM of curcumin for 48hrs, 3) 2Gy radiation after 72hrs culture, and 4) curcumin for48hrs with radiation after 72hrs. XTT, trypane blue, CD45+ and 7AAD flow cytometry where done for cell viability assessment. Results: Viability percentages were as follows: XTT testing showed viability of 18.5 for the cells under condition 2, 106.66 for condition 3, 11.23 for condition 4 and 100 for control. Flow cytometry showed viability of 41.84 for condition 2, 60.74 for condition 3, 36.16 for condition 4, and 66.34 on control group. Under trypane blue: condition 2 had viability of 53.07, condition 3 of 63.05, condition 4 of 39.37, while control had a viability of 67.91. Flow cytometry results were significant when comparing group 2 and control (p=0.003), and group 3 and control (p<0.001), the same found under XTT testing (group 2 vs control with p=0.009; group 3 vs control with p<0.001). Trypane blue results analysis only found significant difference between group 3 and control (p=0.022). Conclusions: Our findings show that curcumin under 25μM, for viability is reduced when compared with the control group with 3 methods as shown before. The viability of the group exposed to 2Gy radiation showed no significant difference to control, regarding this, it has been reported that lymphocytes repair acute radiation damage. Flow cytometry and trypane blue did not showed statistically significant differences between them, however, XTT results where different, which is related that the later study evaluates groups of cells, while the first two assess individual cells. Sebastiá and cols. on 2012 reported a cytotoxic effect of curcumin on cultures, which is supported by our results.

744W

Fine-mapping classical HLA variation associated with ankylosing spondylitis susceptibility.
A. Cortes[1], S.L. Pulit[2], P.I.W. de Bakker[2], M.A. Brown[1], International Genetics of Ankylosing Spondylitis Consortium. 1) The University of Queensland Diamantina Institute, Brisbane, QLD, Australia; 2) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands.

Ankylosing spondylitis (AS) is a common, highly heritable, arthritis. 27 susceptibility loci have been identified in and outside the MHC. The HLA-B27 allele is the major genetic risk factor to AS and its role in disease aetiology remains elusive. Potential mechanisms include disorders of the antigenic presentation function of HLA class I proteins, or on abnormal intracellular effects unique to the B27 protein variant. Prior studies have suggested that other HLA-B alleles and MHC genes are involved in AS-susceptibility. In this study we aim to better define the MHC associations and to identify functional and potentially causal variants to shed light on the mechanisms by which the B27 molecule confers risk to disease. We genotyped 7,264 MHC SNPs in 9,069 affected subjects and 13,578 controls of European descent using the Immunochip. Genotyping was followed by imputation of HLA class I and II alleles and amino acid positions of HLA proteins. Association to disease susceptibility was assessed by logistic regression correcting for population structure. Association was observed with SNPs in the HLA-B locus (P<10-320). Analysis of HLA-B alleles revealed other non-B27 alleles affecting susceptibility of moderate effect size (B40:01, OR=1.13; *40:02, OR=1.59; *51:01, OR=1.36; *07:02, OR= 0.77; *57:01, OR=0.73). After controlling for the associated haplotypes in HLA-B we observed independent association signals with SNPs in the HLA-A locus and in HLA-DRB1. Analysis of polymorphic amino acid positions demonstrated that the most significant polymorphisms in the HLA-B and HLA-DPB1 loci were amino acid residues located in the binding pocket of these molecules. We have previously shown that AS is associated with ERAP1 polymorphisms only in HLA-B27 carriers. Additionally, we show that amongst B27 negative cases, ERAP1 variants are AS-associated in B40-positive but not -negative subjects, indicating that ERAP1 variants also interact with B40. This study has identified susceptibility alleles in HLA class I and II loci. The identification of multiple HLA-B alleles affecting susceptibility and the interaction between the alleles HLA-B27 and *40 with ERAP1 (rs30187) further supports antigen presentation as the molecular mechanism by which HLA-B27 confers risk to disease. The presence of associated polymorphism in both HLA class I and II loci suggests that antigen presentation to both CD4+ and CD8+ T lymphocytes are important in AS pathogenesis and/or tissue specificity.

745T

Mapping an Otitis Media Gene within an Indigenous Population: Evidence for Locus Heterogeneity.
E.M. Cutiongco-de la Paz[1,3], G.T. Abes[2,4], M.R.T. Reyes-Quintos[2,4], C.M. Chiong[2,4], M.L. Tantoco[2], M.C. Garcia[4], E.G. Llanes[2,4], T.L. Gloria-Cruz[2,4], A.L. Chan[2,4], J.W. Belmont[5], S.M. Leal[5], R.L.P. Santos-Cortez[5]. 1) Philippine National Ear Institute, University of the Philippines (UP) Manila - National Institutes of Health (NIH), Manila, Philippines; 2) Philippine National Ear Institute, UP Manila - NIH, Manila, Philippines; 3) Department of Pediatrics, UP College of Medicine - Philippine General Hospital (PGH), Manila, Philippines; 4) Department of Otorhinolaryngology, UP College of Medicine - PGH, Manila, Philippines; 5) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

Otitis media (OM) is a major cause of hearing loss across all age groups and continues to incur a huge cost to health systems due to changes in pathogen patterns and antibiotic resistance. Identification of genes that have a role in development of OM is expected to illuminate biologic pathways that may be targeted for treatment. Although strong evidence for genetic susceptibility to OM exists, the identification of causal mutations remains elusive. From an indigenous population that has a high prevalence of OM, 49 individuals that can be connected by a single large pedigree were examined for occurrence of different types of OM. DNA samples from 37 individuals with evidence of current or previous OM and 5 individuals with no history of OM were successfully genotyped using the exome chip. Linkage analysis and haplotype reconstruction were performed. A statistically significant maximum multipoint LOD score of 3.6 was achieved within the 10q21 region, while a maximum multipoint LOD score of 2.8 was achieved at chromosome 5p13. However linkage heterogeneity was detected within the pedigree. Reanalysis of the genotypes with the pedigree broken into smaller families resulted in a maximum multipoint LOD score of 2.1 at chromosomes 10p15-q21 and 9p24-q21. Examination of the haplotypes within each genomic region showed that a haplotype within the 10p region is carried by 30% of individuals with OM, with haplotype segregation within three families. On the other hand, a 9p haplotype is carried by 40% of OM individuals that are from five other families. Additionally a maximum multipoint LOD score between 1.5 and 2.0 was obtained at five other chromosomal regions. DNA samples from selected individuals have been submitted for whole-exome sequencing. The identification of potentially causal genetic variants is best guided by linkage mapping data in order to reduce the search space within the exome and to identify haplotypic regions with incomplete segregation with phenotype which is possibly due to phenocopies or genetic heterogeneity.

746F

Candidate-gene association study of pain crisis rate in sickle cell anemia patients.
G. Galarneau[1,2], G. Lettre[1,2]. 1) Montreal Heart Institute, Montréal, Québec, Canada; 2) Université de Montréal, Montréal, Québec, Canada.

Sickle cell disease (SCD) is a monogenic disease caused by a mutation in the β-globin gene. Pain crisis, causing acute pain in the limbs due to ischemia, is one of the major and most frequent complications of SCD and is responsible for a major fraction of SCD patients hospitalizations. This complication is also the best predictor of overall SCD morbidity. Small studies have hinted that inter-individual variation in the rate of pain crisis in SCD patients is heritable. We aimed to verify if there were any genetic variants associated with SCD-related pain crisis within genes already known to be implicated in pain, based on the Pain Genes Database (LaCroix-Fralish, M.L., Pain, 2007). For our study, we had access to three different genotyping datasets, all originating from the Cooperative Study of Sickle Cell Disease: (i) 1,434 individuals were genotyped on the IBC array, a gene-centric array targeting ~2,100 genes related to heart, lung and blood diseases, (ii) 1,409 individuals were genotyped on the exome chip and (iii) 1,279 were genotyped on the Illumina 610-Quad array. 873 patients were genotyped on the three platforms. To maximize the number of samples included in each analysis, we analyzed the three datasets independently. We concentrated our analysis on sickle cell anemia patients with a minimum follow-up time of a year. Because the clinical definition of pain crisis is variable, we only analyzed patients who never experienced pain crisis (controls) and patients with pain crisis rates above the 90th percentile (cases). We used simulations to quantify how phenotype misclassification within the middle of the pain rate distribution can affect statistical power. SNPs with a minor allele frequency above 5% located less than 50kb from a gene implicated in pain were included in the common variants analysis. We performed a logistic regression using as covariates: fetal hemoglobin levels, hematocrit, sex, age and principal components. The significance threshold of 7×10^{-6} was determined by calculating the overall number of independent signals tested in the three datasets. We also conducted a rare variants analysis with SKAT-O. This analysis included functional variants with a minor allele frequency lower than 5%. No gene showed significant association (1×10^{-4}). Given our limited power, we can only conclude that there are no variants of strong phenotypic effect in genes previously implicated in pain that associate with pain crisis rate in SCD patients.

747W

Genetic Linkage Analysis and Candidate Gene Approach in Inbreed Puerto Rican Families with Congenital Scoliosis. *J.E. Baez[1], J.C. Orengo[1], E. Suarez[2], C. Burgos[3], V. Franceschini[3], A.S. Cornier[1,3,4].* 1) School of Public Health, Ponce School of Medicine and Health Sciences, Ponce 00717-0211, PR; 2) Department of Microbiology, Ponce School of Medicine and Health Sciences, Ponce, PR; 3) Department of Molecualr Medicine, La Concepcion Hospital, San German, PR; 4) Clinical Research Center, San Jorge Children's Hospital, San Juan ,PR.

Scoliosis can be either idiopathic or secondary to congenital vertebral abnormalities. Idiopathic typically occurs in children and adolescents who are otherwise healthy. Congenital is due to anomalous development of the vertebrae (failure of formation and/or segmentation). Incidence is estimated in 1/1,000 to 1/2,000 but the true incidence remains unknown. Etiology is unknown. Age and Range Mean 6 years (range few days of life to 25 years).Genes of the Notch signaling pathways have been identified as responsible, they are involved in various biologic processes, including somatogenesis, and neurogenesis. We have found compound heterozygozity status in two families with 7 affected individuals with single mutations on the MESP2, Hes 7 and DLL1 genes. Homozygozity linkage analysis in three other inbreed families have proved evidence of linkage (LOD scores 1.75 to 2.5) in chromosomes 2 and 14 respectively. Evidence of genetic linkage to these chromosomes is new and may help identifying new candidate genes resposible for vetrebral malformations and congenital scoliosis.Understanding and defining the natural history of specific mutation(s) and the developmental (molecular) mechanisms in vertebral patterning, will aid in the identification of protective factors for normal spinal development and toward the prevention of disfiguring congenital scoliosis.

748T

GENE POLYMORPHISMS AS RISK FACTOR FOR EARLY PRIMARY OSTEOARTHRITIS OF THE KNEE IN ASIAN INDIANS. *S. poornima, k. subramanyam, q. Hasan.* Genetics & Molecular Medicine, Kamineni Hospitals, Hyderabad, India.

Primary Osteoarthritis (OA) also know “Degenerative arthritis” it is a slowly progressive and irreversible pathology, which is considered as a part of the ageing process. It affects all the joints of the body, predominantly affecting large joints. Symptoms of OR are pain, swelling stiffness and limitations of joint movements. It is estimated that ~10 % of the world’s aged population ≥60 years have symptomatic OA. At present therapy for OA is only palliative and includes the use of pain relieving medicines, physical, exercise and joint replacement surgery. The increasing incidence of OA in individuals below the age of 50 years and the fact that several members of a single family are affected suggests a genetic predisposition. Current evidence indicates an important role of gene polymorphisms in the aetiology of complex diseases like diabetes, cardiovascular disease and arthritis. Three candidate gene polymorphisms (i) A28602G (rs73297147) of COL2A1 on chromosome12q (ii) G8206T (rs72772941) of CRTL1gene on chromosome 5q and (iii) A1412C (rs74063383) of CRTM gene on chromosome 1p have been selected for evaluation in the present study. Each of these are considered to have a functional role in the pathogenesis of Osteoarthritis. All three polymorphisms were assessed in 100 primary osteoarthritis patients and 100 age and sex matched controls from the Asian Indian population. Genomic DNA was isolated from peripheral blood using salting out method. Genotyping was done by polymerase chain reaction followed by restriction digestion and gel electrophoresis. Results indicate that the COL2A1 G allele was associated with OA (2.7373 95% CI 1.5136-4.9504 p=0.0009) and the T allele of CRTL1 was also associated with OA (2.4371 95% CI 1.2748-4.646 P=0.0068). While CRTM A1412C polymorphism showed a deviation from Hardy-Weinberg principle with a total absence of CC genotype in both cases and controls. After Yates correction the odds ratio did not showed any association with OA (2.0415 95% 0.6048-6.891 p=0.2502). Muiti -dimensional regression (MDR) analysis for the three gene polymorphisms showed a negative interaction, suggesting that the COL2A1 A 28602 G and CRTL1 G 8206 T polymorphisms can be used as independent markers to assess the risk of primary OA in Asian Indian population.

749F

Prevalence of specific cell adhesion molecule gene polymorphisms in recurrent aphthous stomatitis. *A. Alkhateeb[1,2], J. Karasneh[3], H. Abbadi[1], A. Hassan[4], M. Thornhill[5].* 1) Dept Biotechnology and Genetics, Jordan Univ Science & Tech, Irbid, Jordan; 2) Deparmtment of Molecular Genetics, Shafallah Medical Genetics Center, Doha, Qatar; 3) Department of Oral Medicine and Oral Surgery, Jordan Univ Science & Tech, Irbid, Jordan; 4) Department of Oral Medicine and Oral Diagnosis, College of Dentistry, Al Mustansiriya University, Baghdad, Iraq; 5) Unit of Oral and Maxillofacial Medicine and Surgery, The University of Sheffield, School of Clinical Dentistry, Sheffield, UK.

Introduction: Recurrent aphthous stomatitis (RAS) is a common oral ulcerative condition. At ulcer sites vascular adhesion molecule-1 (VCAM-1), E-selectin and intercellular adhesion molecule-1 (ICAM-1) are strongly expressed on blood vessels, and ICAM-1 is expressed on keratinocytes. Expression of these molecules would promote leukocyte accumulation and invasion of the epithelium. Thus, polymorphisms in these candidate genes might contribute to RAS susceptibility. Aim: To determine if inheritance of specific selectin, ICAM and VCAM gene polymorphisms are associated with RAS susceptibility. Methods: 100 RAS cases and 153 controls were recruited from a Jordanian population. Blood was collected for hematological investigations and genotyping. Six SNPs were genotyped; E-selectin rs5361 and rs1805193, L-selectin, rs2205849, ICAM-1 rs5498, ICAM- 5 rs885743 and VCAM-1 rs1800821. Results: significant association between inheritance of the A allele of the E-selectin rs5361 gene polymorphism and increased susceptibility to RAS (P= 0.006, Pcorr= 0.027). None of the other SNPs showed a significant association. Conclusions: This is the first report to link inheritance of the A allele of the E-selectin rs5361 polymorphism with increased risk of RAS. Further studies in different patient cohorts are needed to confirm the association and functional analyses might clarify the biological significance of the association.

750W

Role of the DIVERSIN gene in neural tube defects in humans. *R. Allache[1,2], P. De Marco[3], E. Merello[3], V. Capra[3], Z. Kibar[2].* 1) Pathology and Cell Biology, University of Montreal, Montreal, Quebec, Canada; 2) Obstetrics and Gynecology, CHU Sainte Justine Research Center and University of Montréal, Montreal, Quebec, Canada; 3) Neurosurgery Department, G. Gaslini Institute, Genova, Italy.

DIVERSIN (DIV) is a core member of the planar cell polarity (PCP) pathway that controls the process of convergent extension (CE) during gastrulation and neural tube closure in vertebrates. It acts as a molecular switch where it activates the non-canonical Wnt/PCP pathway while simultaneously inhibiting the canonical Wnt/β-catenin pathway. The Wnt/PCP pathway has been strongly implicated in the pathogenesis of neural tube defects (NTDs) in animal models and human cohorts. In this study, we analyzed the role of DIV in these malformations by re-sequencing analysis of its open reading frame and exon-intron junctions in a cohort of 450 unrelated NTD patients. We identified 4 rare heterozygous and 1 homozygous missense mutations in DIV that were absent in all controls analyzed. Polyphen and SIFT analyses predicted those mutations to be functionally deleterious. Functional validation of these mutations on both Wnt/PCP and Wnt/β-catenin pathways using gene reporter assays is currently underway. Our study demonstrates that DIV could act as predisposing genetic factors to NTDs in a subset of patients and further expands our knowledge on the role of PCP genes in the pathogenesis of these malformations.

751T

Epistatic interactions between SNPs in PHF11 and IFNG genes increase risk to allergic phenotypes. *AK. Andiappan[1,2], KJ. PUAN[1], WS. Yeo[2], BT. Lee[1], R. Melchiotti[1], M. Poidinger[1], DY. Wang[3], O. Rotzschke*[1], FT. Chew*[2].* 1) Singapore Immunology Network, Singapore, Singapore; 2) Department of Biological Sciences, National University of Singapore; 3) Department of Otolaryngology, National University of Singapore.

Allergic diseases, such as asthma and allergic rhinitis (AR) are complex diseases resulting from interactions between genetic and environmental factors. According to Allergic Rhinitis and its Impact on Asthma (ARIA), asthma and AR affect nearly 300 million and 500 million of the world population respectively. Genetic variants in PHF11 and IFNG have been associated with asthma and other allergic phenotypes in multiple populations.Here we report a case control study to evaluate the association of PHF11 and IFNG SNPs to allergic rhinitis and asthma either through indivual SNPs or through SNP-SNP epistasis using a population of 2880 ethnic Chinese in Singapore with 2163 having atopy, 795 with AR and 718 having asthma. A total of 8 tagSNPs from PHF11 and 3 from IFNG gene were selected based on linkage disequilibrium. We used PLINK software to evaluate single SNP and SNP-SNP interaction in association to phenotypes. Our association results show that none of the SNPs were significantly associated (P<0.05) individually to any of the allergic phenotypes. In contrast however we identified SNP pairs interacting to increase risk significantly to atopy, asthma and AR (table attached). We also used Synergy Factor (SF) to evaluate the strength of interaction, which allows assessment of binary interactions in case-control studies considering both size and significance. The results validated that the SNP pair rs9526569 (PHF11) and rs2069728 (IFNG) increased risk with SF= 1.78 (P=0.01) for atopy, SF= 1.92 (P=0.017) for AR and SF= 2.22 (P= 0.0035) for asthma. We then tested the hypothesis in a pilot study of 174 more individuals with genotypes for rs9526569 (PHF11) and rs2069728 (IFNG) and also IFNG expression in T cells. Preliminary results suggest that the allele combination of T allele from rs9526569 and G allele of rs2069728 resulted in lesser IFNG expression which could possibly associate to increased risk seen in the genetic association.This is supported by a possible biological interpretation where the transcriptional activator PHF11 regulates Th1 cytokine gene expressions such as IFNG which could in turn control risk for allergy.

Table 1: SNP-SNP epistasis using PLINK software

Chromosome 1	Gene 1	SNP1	Chromosome 1	SNP2	Gene 2	OR_INT	STAT	P
							(Chi-sq)	
Atopy								
12	IFNG	rs2069728	13	rs3794381	PHF11	1.592	5.915	0.015
12	IFNG	rs2069728	13	rs9526569	PHF11	1.571	5.721	0.017
Allergic Rhinitis								
12	IFNG	rs2069728	13	rs9526569	PHF11	1.586	4.418	0.036
Asthma								
12	IFNG	rs2069728	13	rs3794381	PHF11	1.742	6.123	0.013
12	IFNG	rs2069728	13	rs8000149	PHF11	1.705	4.849	0.028
12	IFNG	rs2069728	13	rs9526569	PHF11	1.881	7.981	0.005
12	IFNG	rs1861493	13	rs9568232	PHF11	0.7467	4.301	0.038
12	IFNG	rs1861493	13	rs2274276	PHF11	0.699	7.882	0.005
12	IFNG	rs1861493	13	rs7139494	PHF11	1.326	5.763	0.016

.

752F

Associations of FOXO3A and human lifespan in the Long Life Family Study. *H.T. Bae[1], A.B. Newman[3], J.S. Sanders[4], T.T. Perls[2], P. Sebastiani[1].* 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 2) New England Centenarian Study, Section of Geriatrics, Department of Medicine, Boston University School of Medicine, Boston, MA; 3) Center for Aging and Population Health, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 4) Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, PA.

The gene forkhead box O3 (*FOXO3A*) is a homologue of *daf-16*, which is part of the insulin-IGF1 signaling (IIS) pathway that modulates lifespan in model organisms. Willcox et al. first reported three SNPs in *FOXO3A* to be associated with longevity in long-lived Japanese-American men, and these were replicated in several centenarian studies with different genetic backgrounds. The purpose of this work was to better understand the role of *FOXO3A* in longevity by 1) replicating the known associations, 2) trying to identify new variants of this gene associated with longevity, and 3) attempting to build a network model of the relationships among *FOXO3A* alleles, longevity and some of its sub-phenotypes in the Long Life Family Study (LLFS). The association analysis (N=4656) was performed using Cox proportional hazard models for age at death, censored at the last age at contact for living subjects, adjusting for sex and the top 10 principal components. Correlation within-family was modeled using random effects with variance covariance matrix proportional to the kinship matrix. Gene-wide significance threshold was determined by applying Bonferroni correction to the number of independent loci (M_{eff}=20) in *FOXO3A*, estimated by the simpleM method (Gao et al. 2010). In the LLFS, 6 out of 11 known variants replicated (p<0.05), one of which almost reached gene-wide significance (rs12206094, p= 0.0027, HR=0.83), but no novel variants were identified in 113 tested SNPs. Next we tried to dissect the significant associations between independent *FOXO3A* alleles and lifespan by building a network model (N=3136) that included age at enrollment, sex, birth year cohort, and the Health Aging Index (HAI)~a composite score consisting of age-related phenotypes including systolic blood pressure, pulmonary vital capacity, serum creatinine, fasting glucose, and Modified Mini Mental Status Exam score. The network was built by fitting regression models of follow-up time, HAI, age at enrollment, and by identifying statistically significant predictors through a forward search. These models were linked through the common variables to build the joint network model and the network showed that the association of one *FOXO3A* allele with longevity was completely mediated by the HAI, while a different *FOXO3A* allele was still associated with lifespan after adjusting for the HAI. The analysis suggests that different *FOXO3A* alleles may affect exceptional longevity through different age-related mechanisms.

753W

Association of Endothelial nitric oxide synthase (eNOS) gene polymorphisms (-786T>C, 894G>T and 4a/b) with T2D and its complications (CAD and ESRD). *A.J.S. BHANWER, P. RAINA, R. KAUR, K. MATHAROO.* HUMAN GENETICS, GURU NANAK DEV UNIVERSITY, Amritsar, PUNJAB, India.

Type 2 diabetes (T2D) has become a medical catastrophe of worldwide dimension. The chronic hyperglycemia associated with T2D causes long-term damage and dysfunction of different organs especially kidney, eyes, nerves, heart and blood vessels ensuing to a variety of microvascular and macrovascular complications. T2D patients have two to four times increased risk of developing coronary artery disease (CAD) and end stage renal disease (ESRD) as compared to their non diabetic counterparts. Endothelial dysfunction is prevalent in T2D and is believed to be an important susceptibility factor for development of CAD and ESRD among T2D patients. It is primarily due to deregulation of endothelial nitric oxide synthase (eNOS) enzymatic activity and inactivation of nitric oxide (NO) through oxidative stress. In the present study, polymorphisms in the endothelial nitric oxide synthase (eNOS) gene have been evaluated for association with T2D, CAD and ESRD in the population of Punjab. A total of 936 patients (567 with T2D, 169 T2D with CAD and 200 T2D with ESRD) and 592 healthy controls were genotyped for one insertion deletion (4a/b ins/del) and two RFLP based polymorphisms (-786 T>C and 894 G>T). Statistically significant difference for genotype (p=0.001) and allele frequencies (p=0.002) for -786 T>C promoter polymorphism between pooled cases and controls was observed. Genotype and allele frequencies for T2D cases with CAD (p= 0.003, p= 0.006) and allele frequencies for T2D cases with ESRD (p=0.009) were significantly different in comparison to controls. The C allele provided 1.35, 1.52 and 1.45 fold risk to development of T2D [OR=1.35(1.12-1.64 at 95%CI)], CAD [OR=1.52(1.12-2.04 at 95%CI)] and ESRD [OR=1.45(1.09-1.92 at 95%CI)], respectively. T-allele of 894 G>T exonic polymorphism contributed 1.35 fold increased risk to ESRD (p=0.043, OR=1.35(1.01-1.79 at 95%CI). The deletion allele (a) of intronic ins/del polymorphism 4a/b, was also observed to increase the risk of ESRD by 1.4 fold [(p=0.024, OR= 1.39(1.04-1.85 at 95%CI)]. In conclusion, eNOS -786 T>C polymorphism seems to be associated with risk of T2D and its progression to CAD and ESRD complications. eNOS 894 G>T and 4a/b polymorphisms appears to play a significant role in predisposition to development of ESRD but not CAD in T2D patients in the population of Punjab.

754T

Contribution of FTO (rs9939609), PON1 (rs705379) and SLC6A4 (5-HTTLPR) genes with the genetic risk for obesity in children from Yucatan, Mexico. *L. Gonzalez-Herrera, M.J. Lopez-Gonzalez, Z. Carballo-Cardeña, G. Perez-Mendoza, N. Mendez-Dominguez, D. Pinto-Escalante*. Dept Genetica, Univ Autonoma de Yucatan, Merida, Yucatan, Yucatan, Mexico.

Genetic variation plays a major role in determining the susceptibility to the obesogenic environment. Significant associations of the SNP rs9939609-FTO with body mass index (BMI) and with the risk for obesity have been suggested in homozygotes AA. The central role of FTO might be through an effect on cerebrocortical insulin sensitivity; as homozygous have a reduced insulin response in the brain. Moreover, significant associations have been found between 5-HTTLPR polymorphism of the serotonin transporter SLC6A4 and being overweight, hypothesizing that it regulates behavioral and metabolic responses associated with the development of obesity through feeding and satiety. Because obesity is associated with oxidative damage and that PON1 has been found in the interstitial space between unilocular fat cells of the adipose tissue; some PON1 polymorphisms have been associated with obesity. In this study, we evaluated the association of these three polymorphisms with the risk for obesity in children from Yucatan, Mexico; where child obesity is the first cause of morbidity. We included 76 obese children, and 155 non-obese healthy children under a case-control association study. Genotype and allele frequencies between cases and controls were compared using SNPstats software. Genotype and allele frequencies were distributed according to Hardy-Weinberg expectations (p>0.05) in cases and controls, except for rs9939609-FTO in controls (p=0.007). Risk alleles frequencies in the studied population were: 16.45% for the SNP rs9939609-FTO, and 38.70%; for both the SNP rs705379-PON1 and SLC6A4 (5-HTTLPR). Significant differences were found for the heterozygous AT genotype of the SNP rs9939609-FTO (p= 0.11) between cases and controls, suggesting that the heterozygous AT genotype of FTO is a genetic risk factor an associated with child obesity in the population of Yucatan. The SNP rs705379-PON1 did not show any significant difference for genotype or allele frequencies (p>0.05), suggesting a lack of association with the genetic risk for obesity, whereas the short allele S of 5-HTTLPR polymorphism in the SLC6A4 gene, showed significant differences between cases and controls (p= 0.0003); suggesting association with child obesity in the studied population. We demonstrated the contribution of the heterozygous AT genotype of the SNP rs9939609-FTO and of the short allele S of the 5-HTTLPR polymorphism; to the genetic risk for child obesity in the population of Yucatan, Mexico.

755F

Polymorphisms in IL12, NOD1, and TLR pathway genes associated with tuberculosis disease and resistance to *M. tuberculosis* infection. *N.B. Hall[1], R.P. Igo[1], L.L. Malone[3], F. Qiu[1], A. Schnell[1], S. Zalwango[3], T.R. Hawn[4], H. Mayanja-Kizza[5], W.H. Boom[2,3], C.M. Stein[1,3], Tuberculosis Research Unit (TBRU).* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Medicine, Case Western Reserve University, Cleveland, OH; 3) CWRU-Makerere University Research Collaboration, Cleveland, OH; 4) Department of Medicine, University of Washington School of Medicine, Seattle, WA; 5) Makerere University School of Medicine and Mulago Hospital, Kampala, Uganda.

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), remains a major public health threat globally, with a high burden in Sub-Saharan Africa. According to the World Health Organization in 2011, Uganda's TB incidence rate was 193 per 100,000 people, compared to 3.9 per 100,000 in the United States. Human host genetics have been demonstrated to be involved in the pathogenesis of TB, focusing on innate and adaptive immunity genes. Though millions of people are exposed to TB worldwide, not everyone who is exposed becomes infected with Mtb as diagnosed by a positive tuberculin skin test (TST), and of those infected, it is estimated that only about 1 out of 10 individuals develops active tuberculosis disease. The two main phenotypes of interest in TB pathogenesis, therefore, are active TB disease and remaining persistently TST negative (PTST-) despite continued exposure to the bacteria. In this analysis, we used data collected through a large ongoing household contact study in Kampala, Uganda, to identify genetic association with these phenotypes. We focused on 29 genes involved in the TNF, TLR/NLR, and IFNG/IL12 pathways, genotyping 546 haplotype-tagging single nucleotide polymorphisms within these genes. The 890 subjects from 299 households included in this analysis were enrolled between 1995-1999 and 2002-2008. Analysis was conducted using generalized estimation equations, clustering the data by family. For each outcome, TB and PTST-, both the dominant and the additive model were run, with sex and HIV as covariates. We found SNPs in IL12RB2 and NOD2 to be associated with both PTST- and TB. 3 SNPs in IL10, 3 in SLC6A3, and 2 each in NOD1 and NOD2 were found to be associated with PTST-, with IL12A, IL12B, and TLR pathway genes, associated with TB. Among the most significant SNPs associated with TB were IL12RB2 rs2307147 (3.5e-7), and TLR2 rs1816702 (9.4e-7). For PTST- the most significant were TOLLIP rs5743942 (0.012), SLC6A3 rs409588 (0.014), and IL12RB1 rs17852635 (0.019). Though these were not significant after correcting for multiple testing, some of these regions identified are understudied, which provides an impetus to focus future research in these regions.

756W

Mannose-binding lectin (MBL2) gene variations and malaria risk in Indian populations. *A.N. Jha[1], P. Sundaravadivel[1], L. Singh[1], T.P. Velavan[2], K. Thangaraj[1]*. 1) Evolutionary and Medical Genetics, Centre for Cellular and Molecular Biology, Hyderabad, AP, India; 2) Institute of Tropical Medicine, University of Tübingen, Tübingen, Germany.

The pathogenesis of malaria and its severity depends on complex interplay of host genetic make-up, the parasite virulence and transmission dynamics and the host immune responses. MBL2 is a pattern recognition receptor of the innate immunity, which, recognize and react to specific repertoire of carbohydrates on the surface of invading organisms and plays an important role in the course of infectious diseases. In this study, we aim to investigate the association of *MBL2* variants with *P. falciparum* malaria infection in Indian populations. We re-sequenced the complete MIF gene along with 1 kb each of 5' and 3' region in 434 individuals from malaria endemic regions of the Orissa and Chhattisgarh states of India. The subjects comprised of 176 cases of severe malaria, 101 of mild malaria, and 157 ethnically matched asymptomatic controls. Since the entire Indian subcontinent represents malaria endemic region, we extended our study to 830 individuals from 32 socially, linguistically and geographically diverse Indian populations. The *MBL2 -221C* (*X*) allelic variant is associated with increased risk (mild malaria OR:1.9, P^{Corr}=0.0036; severe malaria OR:1.6, P^{Corr}=0.02). The functional variant *MBL2*B* increases the risk (mild malaria OR:2.1, P^{Corr}=0.036; mild vs. severe malaria OR:2.5, P^{Corr}=0.039) and *MBL2*C* increases the risk towards severity (mild vs. severe malaria OR:5.4, P^{Corr}=0.045). The exon1 (*MBL2*D/*B/*C*) increases the risk towards severe malaria (OR:3.4, P^{Corr}= 0.000045). The lower MBL secretor haplotypes were observed more in severe malaria compared to asymptomatic (OR:2.0, P^{Corr}=0.00002). The *MBL2*LYPA* haplotypes confers protection whereas *MBL2*LXPA* increases the malaria risk. Further, the frequency of low MBL2 haplotype varies significantly among Indian populations (0 - 62.5%) and *MBL2* variant *MBL2*B* is the prevalent form of structural variant in most of the Indian populations contrary to MBL2*C in the African populations. Study suggests that the functional *MBL2* variants are significantly associated with malaria pathogenesis in Indian populations. Further, the high and low serum MBL haplotypes are not restricted to a particular geographic region. In same geographic region, populations show varied prevalence of *MBL2* haplotypes. Increased understanding on the genetic basis of disease susceptibility will help to identify the high-risk individuals, their treatment and effective disease management strategies.

757T

Cytotoxic T-lymphocyte-associated protein 4 (CTLA4) gene and Graves disease: Case-control study. *Y. Lee[1,2,3], C. Huang[1,4], W. Ting[1,4], F. Lo[5], C. Lin[6], Y. Wu[7], M. Chien[8], C. Wang[8], W. Chen[1], W. Lin[1], S. Chang[1], C. Lin[1].* 1) Dept Pediatrics & Med Res, Mackay Memorial Hosp, New Taipei City, Taiwan; 2) Department of Biomedical Sciences, Mackay Medical College; 3) Department of School of Medicine, Taipei Medical University; 4) Department of , Mackay Medicine, Nursing and Management College; 5) Department of Pediatrics, Chang Gung Memorial Hospital; 6) Department of Pediatrics, Mackay Memorial Hospital HsinChu Branch; 7) Department of Pediatrics, Changhua Christian Hospital; 8) Department of Endocrinology and Metabolism, Mackay Memorial Hospital, Taipei, Taiwan.

Graves disease (GD) is one of the most common autoimmune diseases characterized by the presence of TSH receptor antibodies. T cell-mediated autoimmunity against thyroid antigens and B cells producing TSH receptor antibodies can be demonstrated. The *CTLA4* gene encodes cytotoxic T-lymphocyte-associated protein 4 which is involved in the control of the proliferation and apoptosis of T lymphocytes. We investigated whether the *CTLA4* gene was associated with GD in Han Chinese. **Materials** The patients were 504 unrelated subjects (217 children, 287 adults). GD was diagnosed on the basis of clinical and laboratory evidence (positive autoantibodies to TSH receptor). The controls were 920 subjects. **Methods** dbSNP rs5742909 (-318C/T), rs231775 (+49A/G), and rs3087243 (CT60A/G) of the *CTLA4* gene were genotyped by PCR-RFLP or TaqMan Allelic Discrimination Assay. **Statistical analysis** We assessed the Hardy-Weinberg equilibrium for the SNPs, estimated the frequencies of haplotypes with an accelerated expectation-maximization algorithm, and tested pairwise linkage disequilibrium (LD) between the SNPs in both patients and controls using Haploview 4.2. Statistical difference in genotype, allele, carrier, and haplotype distributions between patients and controls were assessed by the chi-square test. Odds ratios (OR) and 95% confidence intervals were also calculated. The Bonferroni correction was used for multiple comparisons. A *Pc* value of less than 0.05 was considered statistically significant. **Results** The genotype distributions of the 3 SNPs in patients and controls were in Hardy-Weinberg equilibrium ($P > 0.42$). They were in linkage disequilibrium with each other in controls (D' >0.87). The distributions of genotype, allele, and carrier of rs5742909 were not statistically different between patients and controls. Genotypes A/A and A/G, allele and carrier A of rs231775 significantly reduced the risk of GD. However, genotype G/G, allele and carrier G conferred a significantly increased risk of GD. Genotypes A/A and A/G, allele and carrier A of rs3087243 rendered significant protection against GD, however, genotype G/G, allele and carrier G conferred a significantly increased risk of GD. Haplotype CGG was significantly more frequent in patients than in controls and conferred a rick of GD, however, haplotype CAA was less frequent in patients than in controls and rendered protection against GD. **Conclusion** The *CTLA4* gene was associated with GD in children and adults.

758F

Variation near the *CETP* gene strongly associated with HDL levels shows a sex-specific association with type 2 diabetes in Pima Indians. *A.K. Nair, A. Malhotra, R.L. Hanson, S. Kobes, W.C. Knowler, C. Bogardus, L.J. Baier.* Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Phoenix, AZ 85004.

Recent studies have provided evidence that high density lipoproteins (HDL) regulate plasma glucose level by affecting insulin action, non-insulin mediated glucose uptake and insulin secretion and consequently affect the risk for type 2 diabetes (T2D). A recent longitudinal study reported that subjects who developed T2D had on average 0.14mmol/l lower HDL at baseline and this effect was stronger in women. In Pima Indians, high HDL cholesterol levels had a protective effect against T2D only in women. The objective of the present study was to analyze variants associated with lipid levels in a prior GWAS of 1100 Pima Indians to determine whether any of these variants increase risk for T2D. The strongest finding to date is with a SNP (rs6499863) in the cholesterol ester transfer protein gene (*CETP*). *CETP* is a known genetic and functional locus for HDL levels with higher CETP levels inversely related to HDL levels. Genotyping of this SNP in 3519 full heritage Pima Indian samples informative for T2D and lipid levels confirmed a strong association of the T allele with lower HDL levels ($p = 5.6\times10^{-11}$) and identified an association with higher risk for T2D ($p = 4.6\times10^{-4}$; 1.26[1.11-1.43) with the evidence for T2D association coming only from female subjects (N=1949; $p=2.7\times10^{-5}$; 1.44[1.21-1.71).We replicated the female-specific association with T2D by genotyping a non-overlapping sample of 3901 (2111 female) mixed heritage American Indians (in females, $p=0.036$; 1.25[1.01-1.55]). The strongest evidence for a T2D association came from combining females from both samples (N=4060; $p = 5\times10^{-6}$; 1.36[1.19-1.55]). Further analysis of this variant in individuals with measures of pre-diabetic traits (N=402) identified an association with insulin action in female subjects (N=172, $p=0.008$), where the risk allele for insulin resistance predicted risk for T2D. Although the association with HDL levels independently replicated in the additional samples, where the allele predicting lower HDL levels increased risk for T2D, there was no evidence for sex interaction in the association of genotypes with HDL. These observations are consistent with prior findings that lower HDL concentration was a stronger predictor of future T2D in women as compared to men. rs6499863 maps to the 5' near gene of *CETP* in close proximity to a distant promoter. Our future studies will involve dense mapping in and around this region and studying functional variants in *CETP* for their role in T2D.

759W

Identification of genes that contribute to diabetic nephropathy in type 2 diabetes through family-based association testing of genetic variants detected by targeted exome sequencing. *M.G. Pezzolesi[1,2], J. Jeong[1], A.M. Smiles[1], J. Skupien[1,2], J.C. Mychaleckyj[3], S.S. Rich[3], J.H. Warram[1], A.S. Krolewski[1,2].* 1) Section on Genetics and Epidemiology, Joslin Diabetes Center, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA.

Genome-wide linkage analysis in extended pedigrees from the Joslin Study of Genetics of Nephropathy in Type 2 Diabetes Family Collection previously identified four genomic regions with evidence for linkage with urinary albumin excretion levels (chromosomes 5q, 7q, 21p, and 22q) and two genomic regions linked to variation in renal function (chromosomes 2q and 7p). To identify the susceptibility genes that contribute to these diabetic nephropathy (DN) sub-phenotypes, we performed targeted exome sequencing of all protein-coding genes across these loci in this collection. One-hundred twenty-six DN cases from 42 families with an excess of renal disease (~3 cases per family) were selected for resequencing of the coding region of 361 genes across the six linkage regions using a custom target enrichment library followed by next-generation sequencing. The resulting data achieved greater than 70X on-target read depth and exceeded 20-fold sequence coverage across 70% of the target regions. Multi-sample variant calling using GATK identified a total of 2,442 non-reference variants, including 966 potentially damaging variants. Among these, 14 rare functional variants were significantly over-represented in T2D DN cases relative to the NHLBI's Exome Sequencing Project reference panel. To improve our ability to detect associations at these loci, we are currently expanding our sequencing efforts to all members of the selected families. While these efforts are ongoing, we recently inferred genotypes for the identified potentially damaging variants in the remaining individuals from each family using MERLIN and performed gene-based family association analyses using FBAT. These analyses identified a significant association at the *DEPDC5* gene on chromosome 22 among diabetic family members with proteinuria or end-stage renal disease ($P = 0.016$). The strength of this association improved further when all relatives were analyzed together ($P=3.82\times10^{-3}$). These preliminary data suggest that variants in this gene may contribute to the linkage peak identified at this locus. We anticipate that complete sequencing of the genes across the six linkage regions identified in the Joslin Study of Genetics of Nephropathy in Type 2 Diabetes Family Collection in all members of these families will allow us to establish the genes that contribute to variation in urinary albumin excretion and renal function at these loci.

760T

Imputation of Turkish Population Genotypes Using Immunochip Data and 1000 Genomes Reference Reveals Behçet's Disease Association of SNPs in the *EGR2* Locus.
E.F. Remmers[1], M. Takeuchi[1], D. Ustek[2], N. Mizuki[3], I. Tugal-Tutkun[4], E. Seyahi[5], Y. zyazgan[5], A. Gül[4], D.L. Kastner[1]. 1) Inflammatory Disease Sect, NHGRI/Med Genetics Branch, Bethesda, MD; 2) Dept. of Genetics, Institute of Experimental Medicine, Istanbul University, Istanbul; 3) Dept. Ophthalmology and Visual Science, Yokohama City University, Yokohama; 4) Istanbul Faculty of Medicine, Istanbul University, Istanbul; 5) Cerrahpaşa Faculty of Medicine, Dept., Istanbul University, Istanbul.

Behçet's disease is a genetically complex multisystem inflammatory disease with common occurrence of orogenital ulcers, uveitis, and skin lesions. It can also involve inflammation of the gastrointestinal, pulmonary, musculoskeletal, cardiovascular and neurological systems. The disease is common among modern-day populations who live along the ancient silk trade routes and is a leading cause of blindness in these countries. We recently completed a genome-wide association study of Behçet's disease in the Turkish population with a discovery collection of 1209 cases and 1278 controls genotyped for 311,459 SNPs for which 799,465 autosomal SNPs were imputed using a reference panel of 96 Turkish samples genotyped for this purpose. With independent replication data and meta-analysis the study identified disease-associated SNPs that exceeded genome-wide significance at genomic regions encompassing the major histocompatibility complex, *IL10*, *IL23R*, *CCR1*, *STAT4*, and the natural killer complex gene cluster on chromosome 12p13.2. In addition, variants in *ERAP1* acted recessively and predominantly exerted their effects in individuals with the disease-associated *HLA-B*51* type. We surmised that even with imputation, the 311K genotyped discovery collection SNPs do not fully inform the variation present in the population and predicted that additional loci could be identified by denser genotyping and by imputation using 1000 Genomes haplotypes as a reference. We genotyped 2014 cases and 1826 controls from Turkey using the Illumina Immunochip, which provides dense coverage for 186 immunologically relevant genes, and evaluated the loci for disease association. We found a chromosome 10 region encompassing *ZNF36*5, *ADO*, and *EGR2* contained many SNPs with near genome-wide significance. We therefore used IMPUTE2 to impute additional SNP genotypes with the 1000 Genomes reference haplotypes. The imputation revealed 2 disease-associated SNPs with $P < 5 \times 10^{-8}$. These SNPs clustered 5' of the *EGR2* gene (early growth response 2), which encodes a transcription factor, E3 SUMO-protein ligase. Defects in this gene are responsible for several Charcot-Marie-Tooth disease types and variants 5' of the gene are associated with increased risk of Ewing sarcoma. Recent work has shown roles of EGR2 and EGR3 in controlling inflammation and in promoting T and B cell antigen receptor signaling, suggesting a contribution of *EGR2* variants to the hyperinflammatory state of Behçet's disease.

761F

Interrogating the *PBX-WNT-TP63* pathway in human nonsyndromic cleft lip/ palate.
R. Silva[1,2,3], A. Letra[1,2,3], L. Maili[3], J.B. Mulliken[4], S. Slifer[5], S.H. Blanton[5], J.T. Hecht[2,3]. 1) Department of Endodontics, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 2) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 3) Department of Pediatrics, Pediatric Research Center, University of Texas Health Science Center Medical School at Houston, Houston, TX; 4) Children's Hospital of Boston, Boston, MA; 5) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Nonsyndromic cleft lip/palate (NSCL/P) is the most common human craniofacial birth anomaly and results from defects during morphogenesis of craniofacial structures. Mutations in genes involved in the morphogenesis and patterning of the craniofacial structures, such as the Wnt and Bmp gene families, *Irf6*, and *p63*, lead to CL/P in mice and humans. Recently, a new regulatory module comprised of *Pbx-Wnt-p63-Irf6* was suggested to control facial morphogenesis in mice. Loss of *Pbx* genes in mice resulted in a fully penetrant CL/P phenotype and perturbed Wnt signaling and *p63* function. Here, we investigated if variations in *PBX1, PBX2, WNT3, WNT9B* and *TP63* genes and their interactions were associated with NSCL/P in humans. Twenty-two SNPs flanking and within selected genes were genotyped in our NSCL/P datasets consisting of simplex and multiplex families, including 513 white, nonHispanic (WNH) American (primarily western European) families and 316 Hispanic families from Texas. Family-based association analyses and association in the presence of linkage were performed for individual SNPs in the families, stratified by ethnicity and family history of NSCL/P. Gene-gene interaction calculations were performed for several genes previously associated with NSCL/P. Evidence of association was found for *WNT3* rs7216231 with NSCL/P in the overall Hispanic dataset (p=0.0007). Associations were also observed for *PBX1* rs1044263 (p=0.002) and *PBX2* rs176095 (p=0.0005) in Hispanic multiplex families. Haplotype analyses showed the association of several marker haplotypes with NSCL/P in both datasets. Evidence of gene-gene interaction was found between *TP63* rs1515490 and *WNT3A* rs752107 (P=0.00005). In summary, we provide the first evidence for a role of *PBX1* and *PBX2* and continuing support for the role of *WNT* and *TP63* genes in the etiology of human NSCL/P. Additional studies across multiple populations will determine the possible involvement of these genes in human NSCL/P.

762W

Genetic heterogeneity of midline facial defects with ocular hypertelorism.
M. Simioni[1], E.L. Freitas[1,2], I. Lopes-Cendes[1], V.L. Gil-da-Silva-Lopes[1]. 1) Department of Medical Genetics, FCM/UNICAMP, Campinas, Brazil; 2) Department of Genetics and Evolutionary Biology, Bioscience Institute, University of São Paulo (USP), São Paulo, Brazil.

Midline Facial Defects with Ocular Hypertelorim (MFDH) is a rare and heterogeneous association of features, with a wide spectrum of abnormalities. Most common ones are ocular hypertelorism, broadening of the nasal root and orofacial clefting, which suggest the involvement of genes related to regulation of embryological development of the frontonasal process. Structural and functional anomalies of the central nervous system (CNS) are also frequently in MFDH. Regarding the etiology of this condition, mutations in ALX1, ALX3 and ALX4 genes were described in some families with frontonasal dysplasia, the most recognizable form of MDFH. SHH, FGF8 and PAX3 genes participate of pattern and growth controlling during developmental facial primordia and central nervous system processes. Mutations in these genes have been associated with other craniofacial disorders. Also, few cases of MFDH with chromosomal abnormalities have been reported. We performed array-CGH and investigated mutations in SHH, FGF8 and PAX3 genes by direct sequencing and in 14 patients with MFDH. Recognized monogenic disorders and well-known syndromic conditions were excluded. Array-CGH detected chromosomal abnormalities in almost 15% of sample: a deletion of 6.2 Mb at 2q36.1-q36.3; a deletion at 9p24.1-9p24.3 (6.36 Mb) and a duplication at 20p13 (14.83 Mb). In first case, PAX3 gene was located between the deleted region. Despite of several single nucleotide polymorphisms (SNPs), no pathogenic mutations were found, which could be related to the size of sample. However, these results reinforce the heterogeneity of MFDH and justified the indication of array-CGH in routine care and genetic investigation of these patients.

763T

Analysis of Candidate Gene Sequences in Suicide Severity. *C.C. Zai, V. Goncalves, A.K. Tiwari, V. de Luca, J.B. Vincent, J.L. Kennedy.* Dept Neurogenetics, CAMH, Toronto, ON, Canada.

Background: A number of novel candidate genes have been identified in recent genome-wide association studies of suicide attempt in bipolar disorder (Willour et al, 2011; Perlis et al, 2010). However, the historical candidate genes, including NTRK2, HTR1A, and HTR1B, did not appear to be among the top findings in these studies. Perhaps a combination of rare and common variants may contribute to the predisposition to suicidal behaviour. Method: We analyzed variant data from high-throughput DNA sequencing of 202 genes (Nelson et al, Science 2012) in our bipolar disorder cases of self-reported and genetically ascertained European ancestry. We analyzed the phenotype of suicide attempt as well as suicide severity score (from the Schedule for Clinical Assessment in Neuropsychiatry SCAN: 0=non-suicidal; 1=suicide plan/ideation; 2=suicide attempt without serious harm; 3=suicide attempt with serious harm; 4=suicide attempt designed to end life; N=227). We conducted preliminary analysis using PLINK, with history of alcohol use disorder as well as sex and age were included as covariates. A total of 3199 DNA variants across the 202 gene regions were analyzed. Results: Among the findings, we found a number of DNA variants in NTRK2 (rs41312188, rs2289656; p<0.05) and HTR1A (rs34118353; p<0.05) to be nominally associated with suicide severity scores. These regions are also nominally significant for the analysis of lifetime history of suicide attempt. Conclusions: We conducted a high-throughput targeted sequence analysis of suicide severity in bipolar disorder and found a number of gene regions to be possibly associated with suicidality. We will be using various methods to analyze rare variants, including gene-based tests, and incorporating functional annotation. The top findings will be validated by Sanger sequencing or SNP genotyping. We will attempt to replicate the validated results in other bipolar disorder samples. We will further explore the interaction of these gene variants with history of alcohol abuse/ dependence.

764F

Targeted sequencing of 100 psoriasis candidate loci in 4,845 samples of European descent. *M. Zawistowski[1], P. Stuart[2], L. Tsoi[1], Y. Li[1], E. Ellinghaus[3], V. Chandran[4], T. Tejasvi[2], C. Helms[5], K. Callies Duffin[6], M. Weichenthal[3], J.J. Voorhees[2], A.M. Bowcock[5], P. Rahman[7], A. Franke[3], G.G. Krueger[6], D.D. Gladman[4], R.P. Nair[2], H.M. Kang[1], G.R. Abecasis[1], J.T. Elder[2].* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Dept of Dermatology, University of Michigan, Ann Arbor, MI; 3) Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; 4) Toronto Western Hospital, Toronto, ON; 5) Department of Genetics, Washington University at St. Louis, St. Louis, MO; 6) School of Medicine, Dermatology, University of Utah, Salt Lake City, UT; 7) 7Memorial University, St. John's, NF, Canada.

Psoriasis is a chronic, immune-mediated inflammatory disease of the skin with prevalence ranging from 0.2 to 2% in different populations. To date, genome-wide association studies have identified 36 psoriasis susceptibility loci in Europeans, with several showing strong evidence for multiple associated variants, indicating both genetic and allelic heterogeneity. In most cases, the precise causal variants at each locus are yet to be determined. With the goal of identifying these causal variants, we selected 100 candidate regions from among the confirmed psoriasis loci as well as additional less strongly-associated regions, totaling 5.7 Mb of sequence and containing 769 genes, for deep, targeted resequencing. In each region, we sequenced all exons of transcription units within 250 Kb of the strongest previous association signal; except in the MHC, where we targeted five specific candidate genes comprising 137 Kb of sequence. In addition, we used a whole-region sequencing strategy that included non-coding intergenic and intronic sequence for ten regions with the most significant association signals, and a full transcription unit plus flanking sequence approach for four established loci of lesser significance. The dataset consists of 4,845 samples of predominantly European ancestry from six different centers and included ~30% psoriatic arthritis cases, ~30% cutaneous psoriasis cases and ~40% controls. We predict that our sample should capture >99% of variants in the target regions with allele frequency at least 0.1%. The sequencing averaged 80x coverage across the target regions and discovered over >140,000 total variants, including ~9,000 protein-altering coding variants not previously observed by either the 1000 Genomes or the NHLBI Exome Sequencing projects. We will assess the role of rare variation in psoriasis using single variant and multi-marker gene-level burden tests. Based on a set of preliminary variant calls, we designed a custom exome array to validate the genotype calls within the sequencing dataset, particularly for singletons, as well as for replication of interesting findings in independent samples.

765W

Admixture mapping of uterine leiomyomata in the African American population. *K. Zhang, H. Wiener, B. Aissani.* Univ Alabama at Birmingham, Birmingham, AL.

Background. The genetic basis of susceptibility to uterine leiomyomata (UL or fibroids), a condition with an estimated cumulative incidence of more than 80% among African American women of age 50, remains largely unexplored. The higher prevalence of UL among African American women compared to European American women prompted us to adopt admixture mapping to evaluate candidate chromosomal regions for UL. Study design. We genotyped a total of 1,583 MALD (mapping by admixture linkage disequilibrium) SNPs in a sample of 393 premenopausal UL cases ascertained by ultrasonography in the NIEHS uterine fibroid study. We selected 1,167 MALD SNPs overlapping candidate chromosomal regions recently identified by other groups through linkage scan (2q37, 3p21, 5p13, 10p11, 11p15, 12q14 and 17q25) and admixture mapping (2q37, 4p16.1, 10q26). We also genotyped an additional set of 252 MALD SNPs on chromosome 1 to re-assess suspected UL loci on this chromosome as well as 164 randomly selected MALD SNPs on chromosomes 15, 16, 18, 19 and 21. Analysis. The ADMIXMAP program was used in case-only models to compare ancestry at test SNPs with genome-wide ancestry and statistical significance was assessed using standard normal Z statistics, with a threshold of Z >4.0 considered statistically significant. Results. In models with no stratification for body mass index (BMI), a phenotype associated with UL albeit in a complex relationship, SNP loci at several chromosomes (1, 2, 3, 4, 5, 11, 12 and 15) reached statistical significance (Z >4.0; p=5.5E-05-3.9E-15), with those on chromosomes 1p11, 2q37 and 12q24 being the most significant (Z >5.0). After stratification by BMI, only chromosomes 1q42, 2q37 and 15q24 remained significant at p=0.05 across all the BMI categories (BMI<25; 25-30 and >30), with the association peaking at rs756784 on 1q42 (Z=4.0-4.7; p=7.6E-05-1.4E-08), at rs256552, rs256550 and rs12479375 on 2q37 (Z=4.8; p=2.2E-02-1.7E-06), and at rs1562250 and rs8030499 on 15q24 (Z=2.4-3.2; p=1.8E-02-3.6E-03). Our data show that only rs756784 on 1q42 remains significant across the BMI categories after correction for multiple testing. Conclusion. While this study replicated previous findings, only the new candidate 1q42 remained statistically significant after correction for multiple testing. Co-localization of UL-associated phenotypes such as the age at menopause and obesity-related traits on 1q42 further supports a candidate status for this region.

766T

Characterizing genetic risk factors of obesity in a coronary artery disease cohort. *S.G. Gregory[1], L.K. Kwee[1], E. Grass[1], C. Haynes[1], N. Stitziel[2], S. Kathiresan[2], E.R. Hauser[1], W.E. Kraus[1], S.H. Shah[1].* 1) The Duke Institute of Molecular Physiology, Duke University Medical Center, Durham, NC; 2) Massachusetts General Hospital, Harvard Medical School, Boston, MA.

Purpose: Obesity is a chronic metabolic disorder that is contributing to morbidity and mortality at epidemic proportions in western and, increasingly, developing nations. Although lifestyle contributes to the development of obesity, there is an established genetic role in the disease state, which has only been partly elucidated by the common variant common disease hypothesis. While obesity leads to the development of several chronic metabolic disorders, here we report the identification of obesity genetic risk factors in a coronary artery disease (CAD) cohort using a single nucleotide variant (SNV) 'exome' microarray that contains both common and rare coding variants. Methods: DNA was obtained from subjects selected from the CATHGEN biorepository of patients referred for cardiac catheterization at Duke University. Samples were genotyped on the Illumina Infinium HumanExome BeadChip, which includes >240K coding SNVs from >20,000 coding and genomic sequences with a combination of rare and common non-synonymous variants. After quality control measures and filtering of monomorphic probes, 65,853 polymorphic autosomal SNVs were available for analysis in European Americans (EA, n=1192), and 97,118 SNVs in African Americans (AA, n=1528). We used race-stratified linear regression models to test each genetic variant for association with BMI, adjusting for age, sex and ethnicity (via principal components obtained from Eigenstrat analysis). We combined the EA and AA results using meta-analysis. Results: Three and six SNVs were significantly associated with BMI in the AA and EA samples, respectively, after Bonferroni correction for multiple tests. No SNVs were significantly associated with BMI in the meta-analysis. Several other SNVs in genes that may play a role in the development of CAD and obesity were suggestively associated with BMI (p<10-5) (AA: ACACB, IFT88 and GRB7; EA: DYNC1H1, MTRR, MUC5B and PRSS2; meta-analysis: ITGB2). Conclusion: We have identified coding and putatively causal variants within several genes that are associated with BMI (and implicated in the development and CAD). These variants may provide the functional link between previous association studies that have identified the gene but not the coding variant in the candidate gene. We are currently carrying out interaction analysis and validation of our initial results.

767F

Discovery and fine-mapping of BMI loci using Metatochip: a trans-ethnic meta-analysis from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *J. Gong[1], F. Schumacher[2], L. Hindorff[4], U. Lim[3], J. Haessler[1], C. Carlson[1], S. Rosse[1], P. Bůžková[6], R. Cooper[10], G. Ehret[11], C. Gu[12], M. Irvin[9], M. Graff[14], L. Fernandez-Rhodes[14], E. Boerwinkle[7], T. Matise[5], L. Marchand[3], C. Kooperberg[1], D. Crawford[8], C. Haiman[2], K. North[13,14], U. Peters[1], the Population Architecture using Genomics and Epidemiology (PAGE) Study.* 1) Public Health Science, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Preventive Medicine, Keck School of Medicine / Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 3) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 4) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Department of Genetics, Rutgers University, Piscataway, NJ; 6) Department of Biostatistics, University of Washington, Seattle, WA; 7) The Human Genetics Center and Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX; 8) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Biostatistics, University of Alabama, Birmingham, AL; 10) Preventive Medicine and Epidemiology, Loyola University, Chicago, IL; 11) Division of Cardiology, Geneva University Hospital, Geneva, OH, Switzerland; 12) Department of Biostatistics, Washington University, St. Louis, MO; 13) Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 14) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Genome-wide association studies (GWAS) primarily performed in European-ancestry (EA) populations have identified numerous loci associated with body mass index (BMI). However, it is still unclear whether EA associations can be generalized to other ancestry groups. In addition, for the most part the putative functional variant(s) in these BMI loci remain under investigation. Different local Linkage Disequilibrium (LD) structure between ethnic groups can help narrow in or fine-map these BMI-related loci. Therefore, we used the Metabochip to densely genotype and evaluate 21 BMI GWAS loci identified in EA studies in 29,151 African Americans (AA), 15,000 Asians, and 6,211 Hispanics from the Population Architecture using Genomics and Epidemiology (PAGE) consortium. Our preliminary analysis showed that among the 21 loci, FTO was associated with BMI in all of the three ethnic groups; MC4R in AA and Hispanics; SEC16B, TMEM18, ETV5, GNPDA2, TFAP2B, and BDNF in AA (at 5.8×10-5 significant level; 0.05/average number of SNPs across 21 loci). Moreover, Metabochip-wide trans-ethnic meta-analyses (fixed-effects models) revealed two novel BMI loci, COBLL1-GRB14 (rs10184004, p=2.2×10-7) and TCF7L2 (rs7903146, p=5.4×10-8) loci, that were significant when adjusting for the total number of SNPs tested across the chip (2.5×10-7). The analysis in AA showed that fine-mapping in AA is a powerful approach to narrow in on the underlying causal variants in BMI GWAS loci identified. We will expand upon this work by including additional Asian, Hispanic and Native American samples into this study (totaling ~20,000 Asian, 20,000 Hispanics and 500 Native Americans) providing improved power to study known loci and identify novel loci in a trans-ethnic meta-analysis.

768W

Fine-mapping type 2 diabetes susceptibility loci with the Metabochip. *A.P. Morris[1], T.M. Teslovich[2], T. Ferreira[1], A. Mahajan[1], Y. Lee[2], N.W. Rayner[1], H. Chheda[3], L. Eisele[4], M. Franberg[5], H. Grallert[6], T. Green[7], S. Gustafsson[8], H.A. Kestler[9], B. Sennblad[5], T. Sparso[10], R.J. Strawbridge[11], D. Thuillier[12], L. Yengo[12], R. Mägi[13], I. Prokopenko[14], C.N.A. Palmer[15], D. Altshuler[7], M. Boehnke[2], M.I. McCarthy[1] on behalf of the DIAGRAM Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 3) Institute of Molecular Medicine, University of Helsinki, Finland; 4) Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, Germany; 5) Atherosclerosis Research Unit, Department of Medicine Solna, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden; 6) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, Neuherberg, Germany; 7) Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA; 8) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University Hospital, Uppsala, Sweden; 9) Research Group Bioinformatics and Systems Biology, Institute of Neural Information Processing, University of Ulm, Ulm, Germany; 10) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 11) Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; 12) Lille Institute of Biology, European Genomics Institute of Diabetes, Lille, France; 13) Estonian Genome Center, University of Tartu, Tartu, Estonia; 14) Genomics of Common Disease, Imperial College London, London, UK; 15) Medical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK.

Genome-wide association studies (GWAS) have been successful in identifying loci for type 2 diabetes (T2D). These loci are usually characterised by common lead SNPs with association signals that map to large genomic intervals. Consequently, there has been limited progress in localising causal variants for T2D and establishing the functional impact of these loci in the pathogenesis of the disease. To address these shortcomings, we combined Metabochip data from 25,154 type 2 diabetes (T2D) cases and 50,269 controls of European ancestry, supplemented by imputation up to the 1000 Genomes Project reference panel (March 2012 release), to facilitate fine-mapping of 39 established loci for the disease that are covered on the array. We aimed to: (i) delineate association signals in established loci arising from multiple independent causal variants; (ii) assess the evidence for association with low-frequency (LF) variants, minor allele frequency (MAF) <5%, in established loci; and (iii) define 'credible sets' of SNPs that account for 99% of the probability of including the causal variant at each association signal.

We undertook approximate conditional analyses implemented in GCTA, using 3,298 cases and 3,708 controls from GoDARTS as reference, to identify loci with multiple SNPs at nominal significance ($p<10^{-5}$), reflecting increased prior odds of association in Metabochip fine-mapping regions. These data suggest that association signals mapping to *TCF7L2* and *KCNQ1* can each be explained by three causal variants, whilst those at *DGKB*, *CDKN2A/B*, *KCNJ11*, *HMGA2*, *HNF1A* and *GIPR* can be delineated by two. The conditional analysis suggests that LF associated variants mapping to *TCF7L2* (rs140820620, MAF=1.9%, $p=6.5\times10^{-13}$; rs180726800, MAF=2.0%, $p=1.9\times10^{-6}$), *KCNJ11* (rs61763083, MAF=0.3%, $p=7.5\times10^{-7}$), *HMGA2* (rs116521220, MAF=0.6%, $p=6.3\times10^{-6}$), and *HNF1A* (rs1800574, MAF=2.6%, $p=1.5\times10^{-6}$), are independent ($r^2<0.05$) of the common lead SNPs at these loci. The 99% credible sets map to <10kb at *MTNR1B* (rs10830963 only), *CDKN2A/B* (both signals map to the same <2kb interval), *TCF7L2* (primary signal includes 3 SNPs, 4.3kb), *HNF1B* (7 SNPs, 5.8kb), and *GCK* (3 SNPs, 9.8kb).

We are currently following up these association signals in an additional 12,413 T2D cases and 32,091 controls to confirm fine-mapping intervals for functional investigation.

769T

GTF2IRD1 is an epigenetic regulator involved in facial skin patterning that may underpin facial abnormalities of William-Beuren Syndrome patients. *C.P. Canales[1], S. Corley[2], P. Kaur[3], I. Smyth[4], M. Wilkins[2], E.C. Hardeman[1], S.J. Palmer[1].* 1) Cellular and Genetic Medicine Unit, University of New South Wales, Kensington, Sydney, New South Wales, Australia; 2) School of Biotechnology and Biomolecular Sciences, The New South Wales Systems Biology Initiative, University of New South Wales, Sydney, Australia; 3) Epithelial Stem Cell Biology Laboratory, Peter MacCallum Cancer Centre, Melbourne, Australia; 4) Dept of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia.

Williams-Beuren Syndrome (WBS) is a genetic disorder associated with multisystemic abnormalities, including craniofacial dysmorphology. It is caused by a deletion of 1.55 Mb including 28 genes on chromosome 7q11.23. Analysis of genotype-phenotype correlations in atypical smaller deletions has implicated two evolutionary-related transcription factors GTF2I and GTF2IRD1 as prime candidates for the cause of the facial dysmorphology. We have investigated the involvement of GTF2IRD1, a transcriptional repressor discovered in our laboratory, in the control of epidermal proliferation and differentiation and its role in the causation of the typical WBS craniofacial features. Protein interaction studies indicate that GTF2IRD1 gene silencing occurs via chromatin-modifying epigenetic mechanisms. We have generated a targeted Gtf2ird1 knockout (KO) mouse that shows striking similarities to aspects of the human disease including enlarged lips caused by an extreme thickening of the epidermal layer, associated with an expansion of the basal proliferative zone. The expression of Gtf2ird1 in facial skin shows a dynamic pattern during development that matches with the affected areas of the KO mice, revealing how the expression of this protein correlates with patterning of the future lips. To assess the basis of the facial abnormalities, we have evaluated expression of the cell proliferation marker Ki67 and markers of epidermal differentiation, K10 and K14, and tested skin barrier function by toluidine blue whole mount embryo staining. Our results show increased Ki67, K10 and K14 expression in lips of null animals confirming a hyperproliferative phenotype that begins very early in life. CT-scan analysis of skulls showed no major differences in hard tissue morphology suggesting that the role of Gtf2ird1 is restricted to soft tissue. In order to elucidate gene expression alterations due to the lack of this transcriptional repressor in the skin, we have performed comparative RNA-seq analysis from lip skin samples, which has revealed important clues concerning the molecular mechanisms involved. We conclude that GTF2IRD1 plays an important role in how facial skin is patterned during development. It is a crucial component of the transcriptional machinery for proper cell proliferation and differentiation in specific regions of the skin, including playing an important role in regulation of components that define proper distribution of the epidermal keratin network.

770F

Probabilistic method with correction for imputation error improves fine-mapping resolution. *B. Pasaniuc[1], G. Kichaev[2], N. Zaitlen[3], G. Bhatia[4], E. Kostem[5], F. Hormozdiari[5], E. Eskin[5], A.L. Price[4].* 1) Pathology and Laboratory Medicine, Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Dept of Biostatistics, UCLA; 3) School of Medicine, UCSF; 4) Dept of Epidemiology and Biostatistics, Harvard School of Public Health; 5) Dept of Computer Science, UCLA.

Standard statistical approaches for fine-mapping causal variants use marginal association statistics to prioritize SNPs for functional assays. Here, we propose a probabilistic approach that uses information across all SNPs in a locus to approximate the posterior probability that a given SNP is causal. This approach provides appropriate weights to SNPs both within and between loci. We use simulations based on 1000 Genomes data to show that our approach drastically reduces the total number of SNPs that need to be tested in functional assays before identifying the causal SNPs. For example, an average of 69 SNPs per locus need to be tested to identify 90% of causal SNPs, as compared to 110 SNPs per locus for locus-specific ranking based on marginal association statistics or 304 SNPs per locus for genome-wide ranking based on marginal association statistics. Although our simulations use a flat prior for a SNP being causal, our framework naturally allows for priors that incorporate external information such as functional annotations. Differential imputation error across SNPs leads to biases in the prioritization of SNPs, because imputation error deflates the estimated effect size and reduces the priority of a SNP (e.g. an average of 131 SNPs to be tested per locus to identify 90% of causal variants if the causal is imputed vs. 69 if causal is typed). We use the fact that observed effect sizes at nearby SNPs follow a multivariate normal distribution (determined by effect sizes at causal SNPs and population LD) to incorporate imputation error in approximating the posterior probability of being causal. We use 1000 Genomes simulations to show that our approach leads to superior performance vs. the standard approach of ignoring imputation error (e.g. an average of 96 vs. 131 SNPs per locus need to be tested to identify 90% of causal variants). Differential imputation error across ethnicities can be incorporated into the model, further improving fine-mapping resolution in multi-ethnic fine mapping studies. Finally, we extend our approach to model admixture association signals in recently admixed populations and show that this provides a further boost in fine-mapping resolution.

771W

The association of common SNPs with cholesterol and lipid levels in Latvian population. *J. Klovins, I. Radovica, D. Fridmanis.* Latvian Genome Ctr, Latvian Biomedical Ctr, Riga, Latvia.

Low high density lipoprotein (HDL) cholesterol level, high low density lipoprotein (LDL) cholesterol level and high triglyceride (TG) level are well known risk factors for coronary artery disease (CAD) and mortality in type 2 diabetes (T2D) patients. Recent genome-wide association studies (GWASs) and other human genetic studies have localized many common SNPs in different loci that influence blood lipid levels including previously known loci that are potentially involved in lipid metabolism. There are at least 40 loci associated with one or more blood lipid levels and far more SNPs, but not all of them are successfully replicated. In this study we report an association of 139 SNPs selected from more than 30 genes with quantitative measures of blood lipids in 1345 individuals selected from Genome Database of Latvian Population. 26 SNPs were nominally ($p<0.05$) associated with TH, 58 SNPs were associated with HDL-C, 22 SNPs with LDL-C and 40 SNPs were associated with TG. We have performed an imputation of SNPs for the densest genotyped loci. Allele dosage test was also performed and indicated a strong correlation between the number of risk alleles and quantitative measures of all available lipids. SNPs from CEPT gene were most strongly associated with HDL-C levels (rs3764261; p= 2.2 × 10-10). In summary we have shown that genetic variation strongly affects blood lipid levels in Latvian population.

772T

Genetic Variation and Lumbar Spine Bone Mineral Density in National Health and Nutrition Examination Survey 1999-2002. *J. McLean[1], A. Yesupriya[2], T. Fakhouri[2], Y.-H Hsu[3, 4], D. Kiel[3, 4], A. Looker[2].* 1) Division of Health and Nutrition Examination Surveys, NOVA Research/National Center for Health Statistics, Hyattsville, MD; 2) National Center for Health Statistics,Centers for Disease Control and Prevention, Hyattsville, MD; 3) Institute for Aging Research, Hebrew SeniorLife, Boston, MA; 4) Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA.

Bone mineral density (BMD) is a heritable and multifactorial trait used to diagnose fracture risk and osteoporosis. Genome wide association studies (GWAS) of BMD have identified significant genetic variants in European and Asian populations. Since BMD varies across populations, we aimed to replicate GWAS results in a multiethnic population. The relationship between 23 polymorphisms within 14 candidate genes and lumbar spine BMD (g/cm^2) (LSBMD) was examined by race/ethnicity (non-Hispanic white, non-Hispanic black, and Mexican American) in a nationally representative sample of 7226 participants' age 20+ years from the National Health and Nutrition Examination Survey 1999-2002. We used linear regression that assumed an additive genetic model and controlled for age. False discovery rate (FDR) adjusted p-values were used to correct for multiple comparisons and analyses accounted for the NHANES survey design. Ten polymorphisms within ESR1 (rs1038304 and rs4870044), TNFRSF11B (rs11995824, rs4355801, rs6469804, and rs6469804), MEPE (rs1471403), LRP5 (rs3736228), and CCDC170 (rs6929137) were significantly related to LSBMD in non-Hispanic whites. The two polymorphisms within ZBTB40 (rs6426749 and rs7524102) that have been associated with BMD in East Asian and Northern European populations were significantly associated with LSBMD in non-Hispanic blacks but not in non-Hispanic whites. There were no statistically significant relationships between polymorphisms and LSBMD in Mexican Americans. Significant results from GWAS in European and Asian populations are not consistent with race/ethnic specific results for non-Hispanic whites, non-Hispanic blacks and Mexican Americans in the United States. This indicates a need for further research focusing on these groups.

773F

Polymorphisms of DNA leading to increased susceptibility for Type 2 Diabetes in South Asians: A systematic review and meta-analysis. *Z.N. Sohani[1,2,3], W.Q. Deng[1,3], S.S. Anand[1,2,3].* 1) Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; 2) Population Genomics Program, McMaster University, Hamilton, Ontario, Canada; 3) Population Health Research Institute, Hamilton General Hospital, Hamilton, Ontario, Canada.

Background: Type 2 diabetes (T2D) is a metabolic disorder resulting from interplay between pancreatic β cell function and insulin resistance. People of South Asian ancestry are up to four times more likely to develop T2D than white Europeans and are forecasted to carry 25% of the world's burden by the year 2030. Single Nucleotide Polymorphisms (SNPs), which explain the most genetic variation in humans, may contribute to the increased susceptibility for T2D in South Asians. However, risk from genetic variants has not been fully quantified for this ethnic group. The purpose of this study was to i) ascertain risk estimates from T2D associated polymorphisms in South Asians, and ii) compare effect sizes, risk alleles, and risk allele frequencies (RAFs) from this meta-analysis with genome wide estimates in white Europeans. Methods: A MEDLINE, EMBASE, and CINHAL search identified 3499 studies. An additional 17 studies were identified by hand-search and consultation with experts. We included 88 case control studies, which met our inclusion criteria. Data were extracted independently by two reviewers. SNPs were meta-analyzed using a random effects model and weighted by inverse variance. Odds ratios were determined for the risk allele. Results: After considering the linkage disequilibrium structure, 264 SNPs were considered from all included reports comprised of 65317 cases and 88511 controls. From these, 71 SNPs were conducive to a meta-analysis. Significant association with T2D was found in 22 SNPs from 17 nuclear genes (ADCY5, CDKAL1, CDKN2A/2B, ENPP1, FTO, GLIS3, HHEX, IDE, IGF2BP2, ITLN1, KCNJ11, PGC1α, PPARγ, SLAMF1, SLC30A8, TCF7L2, and UCP2) and 2 SNPs from mitochondrial DNA (mtDNA). The majority of SNPs from nuclear genes increased odds of T2D by between 15-35% per risk allele. Risk alleles from mtDNA SNPs increased odds of T2D by 90%. Interestingly, effect sizes in South Asians were similar to those reported in white Europeans. However, we observed variation in RAFs and alleles. Specifically, risk allele for HHEX was the minor allele in South Asians and common allele in Europeans. Conclusions: This is the first meta-analysis to investigate polymorphisms associated with T2D in South Asians. We demonstrate that similar effect sizes for T2D SNPs are apparent between South Asians and Europeans, but there is variation in RAFs. Differences in risk allele for HHEX may be indicative of selection or a result of genetic drift. Further investigation is warranted.

774W

Assessment of Cytochrome P450 Genetic Variability on Methadone Dose and Tolerance. *H. Tsai[1,2], S. Wang[3], S. Liu[3], I. Ho[3,4,5], Y. Chang[3], C. Chen[3], K. Lin[3], A. Chen[6], Y. Liu[3,7,8].* 1) Dept Pediatrics, Northwestern Univ Sch Med, Chicago, IL; 2) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County, Taiwan; 3) Center for Neuropsychiatric Research, National Health Research Institutes, Zhunan, Miaoli County, Taiwan; 4) Center for Drug Abuse and Addiction, China Medical University Hospital, Taiwan; 5) Graduate Institute of Clinical Medical Science, China Medical University, Taiwan; 6) Department of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, New York, USA; 7) Graduate Institute of Drug Safety, China Medical University, Taiwan; 8) Department of Psychiatry, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan.

Objectives: Methadone is well known as a maintenance drug to prevent heroin-addicted patients from experiencing withdrawal symptoms. Importantly, previous studies have suggested that methadone dose is one of the key factors in relation to success in individuals with methadone maintenance treatment (MMT). In this study, we constructed gene matrix using previously identified gene polymorphisms from CYP2C19, CYP2B6 and CYP3A4, individually, and simultaneously determined whether there was genetic influence on methadone dose and methadone tolerance. Methods: A total of 366 heroin addicts undergoing MMT were recruited in this study. Data were collected using interviewer-administered assessments. The SNPs genotyped and used to construct gene matrix included two from CYP2C19; four from CYP2B6; and five from CYP3A4, separately. After adjusting covariates, regression analyses were performed to examine associations of CYP2C19, CYP2B6 and CYP3A4, separately, with methadone dose and methadone tolerance. Results: Our results indicated that methadone dose and methadone tolerance were dominantly affected by gene dose of CYP2C19, but not CYP2B6 and CYP3A4. In addition, the results also suggested that dominant genetic influence of the CYP2C19 gene dose on methadone dose was only found among the study patients with negative unrine morphine, not with positive urine morphine. Conclusions: The findings in this study have suggested that accounting for the CYP2C19 gene dose may serve as a potential indicator when assessing methadone dose and methadone tolerance, respectively. Ultimately, taking into account genetic information of CYP2C19 and other CYP isoenzymes together in methadone treatment will facilitate understanding interindividual variability of the clinical pharmacokinetics of methadone.

775T

Mapping expression quantitative traits in mesenchymal stem cell cultures derived from nonsyndromic cleft lip and palate patients' *orbicularis oris* muscle. *C. Masotti[1], A.C. Nica[2], S. Ferreira[1], L. Bomilcar[1], D. Meyer[1], D.Y. Sunaga[1], D.F. Bueno[1], N. Alonso[3], D. Franco[4], E. Dermitzakis[2], M.R. Passos-Bueno[1].* 1) Department of Genetics, University of São Paulo, Univ de Sao Paulo, Sao Paulo, Brazil; 2) Department of Human Genetics and Development, University of Geneva, Switzerland; 3) Department of Plastic Surgery, Faculty of Medicine, University of São Paulo, Brazil; 4) Department of Plastic Surgery, Hospital Clementino Braga Filho, Faculty of Medicine, Federal University of Rio de Janeiro, Brazil.

Mesenchymal stem cells derived from dental pulp of nonsyndromic cleft lip and palate (NS CL/P) patients exhibit a distinct expression signature at loci related to extracellular matrix modeling and epithelial-mesenchymal transition processes. *Orbicularis oris* muscle is compromised in NS CL/P patients and is an accessible source of mesenchymal stem cells. In this work we established mesenchymal stem cell cultures derived from *orbicularis oris* muscle (OOMMSC) of 43 NS CL/P patients and four controls in order to identify genetic variants associated with gene expression variation and disorder susceptibility. By correlating genome-wide expression and genotype microarrays, we mapped 119 expression quantitative trait loci (eQTLs) related to 18 genes ($p<0.0001$; FDR=14%). We did not observe significant enrichment of any GO term among genes for which we identified eQTLs. Thirty out of the 119 eQTLs are shared with other tissues or cell types, such as lymphoblastic cell lines, liver, brain cortex, T-cells, and fibroblasts. Twelve previously reported NS CL/P candidate genes have at least one OOMMSC eQTL located within their putative regulatory regions, but none of these genes is directly regulated by these eQTLs. Four eQTLs for *SENP5* and *CEP19* genes ($p<0.05$, FDR=57%) are in linkage disequilibrium with NS CL/P genome-wide associated SNPs in 3q29 region ($r^2>0.52$; $D'>0.9$). The haplotype containing the at-risk variant (rs10489880) is associated with higher expression of SENP5 and lower expression of *CEP9*. *SENP5* is a SUMO-specific protease that plays a critical role in the regulation of sumoylated protein levels in the cell, and sumoylation was demonstrated as an essential mechanism for palate development. Further case-control association study of OOMMSC eQTLs will elucidate the potential role of these variants in NS CL/P susceptibility. Funding: FAPESP/CNPq.

776F

Variations in *ORAI1* gene associated with Kawasaki disease. *Y. Onouchi[1,22,24], R. Fukazawa[2,24], K. Ozaki[1], M. Terai[3], H. Hamada[3], T. Honda[3], H. Suzuki[4], T. Suenaga[4], T. Takeuchi[4], K. Yasukawa[3], R. Ebata[5], K. Higashi[5, 6], T. Saji[7, 24], Y. Kemmotsu[7, 24], S. Takatsuki[7, 24], K. Ouchi[8], F. Kishi[9], T. Yoshikawa[10], T. Nagai[11], K. Hamamoto[12], Y. Sato[13], J. Abe[14, 24], M. Seki[15, 17, 24], T. Kobayashi[16, 17, 24], A. Takahashi[18], T. Tsunoda[19], M. Kubo[20], Y. Nakamura[21], A. Hata[22], T. Tanaka[1, 23].* 1) Lab. for Cardiovascular diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 2) Dept. of Pediatrics, Nippon Medical School, Tokyo, Japan; 3) Dept. of Pediatrics, Tokyo Women's Medical University, Yachiyo Medical Center, Yachiyo, Chiba, Japan; 4) Dept. of Pediatrics, Wakayama Medical University, Wakayama, Wakayama, Japan; 5) Dept. of Pediatrics, Chiba University Graduate School of Medicine, Chiba, Chiba, Japan; 6) Dept. of Cardiology, Chiba Childrens' Hospital, Chiba, Chiba, Japan; 7) Dept. of Pediatrics, Toho University School of Medicine, Tokyo, Japan; 8) Dept. of Pediatrics, Kawasaki Medical School, Kurashiki, Okayama, Japan; 9) Dept. of Molecular Genetics, Kawasaki Medical School, Kurashiki, Okayama, Japan; 10) Dept. of Pediatrics, Fujita Health University, Toyoake, Aichi, Japan; 11) Dept. of Pediatrics, Dokkyo Medical University Koshigaya Hospital, Koshigaya, Saitama, Japan; 12) Dept. of Occupational Therapy, International University of health and welfare, Fukuoka, Fukuoka, Japan; 13) Dept. of Pediatrics, Fuji Heavy Industry LTD. Health Insurance Society General Ohta Hospital, Ohta, Gunma, Japan; 14) Dept. of Allergy and Immunology, Division of Immunology, National Research Institute of Child Health and Development, Tokyo, Japan; 15) Dept. of Cardiology, Gunma Children's Medical Center, Shibukawa, Gunma, Japan; 16) Division of Clinical Pharmacology and Toxicology, The Hospital for Sick Children, Toronto, Ontario, Canada; 17) Department of Pediatrics, Gunma University Graduate School of Medicine, Maebashi, Gunma, Japan; 18) Lab. for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 19) Lab. for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 20) Lab. for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 21) Lab. for Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan; 22) Dept. of Public Health, Chiba University Graduate School of Medicine, Chiba, Chiba, Japan; 23) Dept. of Human Genetics and Disease Diversity, Tokyo Medical and Dental University, Tokyo, Japan; 24) Japan Kawasaki Disease Genome Consortium.

Kawasaki disease (KD; MIM611775) is a systemic vasculitis syndrome with unknown etiology which predominantly affects infants and children. Recent findings of susceptibility genes for KD have suggested up-regulation of Ca^{2+}/NFAT pathway as one of the main pathophysiological processes in KD. In this study, we focused on *ORAI1*, a gene for a channel involved in store operated Ca^{2+} entry located on 12q24 where positive linkage signal was seen in our previous sib pair study of KD, and conducted a genetic association study. By re-sequencing 23kb of *ORAI1* region for 24 subjects including 12 KD cases and 12 controls, we identified 68 variants and then performed linkage disequilibrium analysis with 94 Japanese healthy individuals. After selecting 9 tagging SNPs which represent 37 variants with minor allele frequencies larger than 5%, we performed an association study using 730 KD cases and 1315 controls. Among the 9 SNPs examined, a non-synonymous SNP (rs3741596; p. S218G) showed a nominal association with KD (OR = 1.19, 95%CI 1.02~1.40, P = 0.028). The same trend of association was observed in an independent case control panel (1586 KD cases and 1097 controls) and a significant P value was observed in a meta-analysis (OR = 1.20, 95%CI 1.08~1.33, P = 0.00098). Furthermore, we also found a rare 6 base-pair insertion polymorphism which cause elongation of proline repeat within N-terminal cytoplasmic domain of the ORAI1 protein was overrepresented in KD cases (rs78448924; OR = 3.91, 95%CI 1.30~11.80, P = 0.010). These data indicate altered ORAI1 function confers susceptibility of KD and further highlight importance of the Ca^{2+}/NFAT pathway in the disease pathogenesis.

777W

Elucidation of auto-antibody profiles and genetic risk alleles associated with antinuclear antibody (ANA) positive fraction of healthy population. *P. RAJ[1], L. WANG[1], Quan-Z. LI[1], D. KARP[1], I. DOZMOROV[1], N. OLSEN[2], KM. SIVILS[3], JA. JAMES[3], JA. KELLY[3], B. LAUWERYS[4], PK. GREGERSEN[5], EK. WAKELAND[1].* 1) IMMUNOLOGY, UT SOUTH WESTERN MEDICAL CENTER, UNIVERSITY OF TEXAS, DALLAS, TX; 2) Division of Rheumatology, Penn State Hershey Medical Center, 500 University Drive, Hershey PA 17033; 3) Oklahoma Medical Research Foundation, 825 North East 13th Street, Oklahoma City, Oklahoma 73104, USA; 4) Service de Rhumatologie, SSS/IREC/RUMA, Cliniques Universitaires Saint-Luc & Université catholique de Louvain Avenue Hippocrate 10, bte B2.5390, 1200 BRUXELLES, Belgium; 5) Feinstein Institute for Medical Research in Manhasset, New York, USA.

Antinuclear antibodies (ANA) are the serologic hallmarks of autoimmunity and are frequently seen in patients with systemic autoimmune diseases such as SLE. Many studies have shown that a good number of healthy individuals also display ANA, raising questions concerning the clinical significance of being ANA positive and the relevance of ANA to diagnosis of autoimmune disease. We screened 2223 healthy Caucasians and 143 SLEs for ANA by ELISA and found that ANA is a continuous trait with > one normal distribution in the general population, leading to the delineation of subsets of individuals who manifest low, moderate, and high ANA levels without any apparent disease. Our data show that about 26% of the normal population are ANA positive, with 17% (382 of 2223) having moderate (20-40 ELISA units) and 9% (197 of 2223) having high (>40 ELISA units) ANA levels. Further, to elucidate the disease associated auto-antibody profiles and to distinguish ANA positive normal subjects from SLE patients, we performed protein array screening to detect antibodies against over 90 different auto-antigens implicated in various autoimmune diseases. We found that ANA positive fraction of healthy population makes significant amount of IgG and IgM antibodies against many antigens with more significant and frequent antibodies to extracellular matrix proteins such as collagens and antiphospholipids such as cardiolipin. We noticed stronger and diverse autoantibody signatures in females than males. Next, we genotyped all the ANA negative and ANA positive individuals with Immunochip and performed a quantitative association test on 1583 PCA corrected Caucasians with ANA as dependent variable, and we found strongest association of HLA alleles with ANA. LD analysis of data suggested multiple independent association signals in HLA class I, II and III region. As expected, HLA-DR and DQ alleles were also found associated with ANA as peak SNP rs9268832 in HLA-DRA gene and rs2395252 in HLA-DQA2 gene reached suggestive genome wide p value of 7.63606E-07 and 7.50887E-05, respectively. The minor allele of rs3135388 SNP, a tag of DRB1*1501 allele was found significantly (p=5.23E-03) associated with high ANA level. Further, we studied the expression of HLA-DRB1 in human macrophages by RNAseq and found that ANA associated alleles of rs3135388 and rs9268832 SNPs are also associated with significant (>2 fold) increase in HLA-DRB1 expression. More results and conclusion from study will be discussed.

778T

Meta-analysis followed by replication identifies novel genetic variants in X chromosome associated with systemic lupus erythematosus in Asians and analysis of genetic differences between male and female cases. *J. Yang, Y. Zhang, W. Yang, Y.L. Lau.* Paediatrics and Adolescent Medicine, University of Hong Kong, Pokfulam, Hong Kong.

Systemic lupus erythematosus (SLE) is an autoimmune disease with a strong genetic component and dramatic gender difference in disease prevalence. Involvement of genetic variants on the X-chromosome in SLE susceptibility, especially in non-European populations, has not been adequately explored. In this study, we aimed at identifying genetic variants on the X chromosome associated with disease susceptibility in Asian populations. Meta-analysis was performed on two genome-wide association studies (GWAS) on SLE on Chinese Han populations, focusing on X chromosome SNPs and female samples at first. The prominent signals were further replicated in three independent cohorts, with a total of 5021 female cases and 5700 matched controls. Furthermore, 355 male cases and 4388 corresponding male controls were also used for replication to explore possible gender differences. Variants in or near PRPS2, NAA10 and TMEM187 were shown to be associated with SLE in Asian populations, surpassing genome-wide significance. In addition, two previously reported loci in European populations, IRAK1 (rs1059702), and MECP2 (rs2734647) were also replicated in our study. We also identified a variant upstream of L1CAM with suggestive association, which showed clear interaction with SNPs in NAA10. Independent contributions of these variants were tested using various methods and clearly independent effect of IRAK1 and MECP2 was demonstrated. This study delineates multiple independent signals towards SLE susceptibility from the X-chromosome in Asian populations, expanding our understanding of the role of X-chromosome variants in SLE susceptibility. A comparison was conducted genome-wide on effect sizes of susceptibility variants between male and female cases, and their implication on the sex differences on SLE prevalence was discussed.

779F

Variable set enrichment analysis of exome sequencing data reveals interesting candidate genes for left ventricular hypertrophy. *W. Yang[1], U. Broeckel[3], A.J. Stoddard[3], D. Zhi[4], M.R. Irvin[5], H.K. Tiwari[4], S.C. Hunt[6], D.C. Rao[1], D.K. Arnett[5], C.C. Gu[1,2].* 1) Division of Biostatistics, Washington Univ in St Louis, St Louis, MO; 2) Department of Genetics, Washington Univ in St Louis, St Louis, MO; 3) Department of Medicine, Medical College of Wisconsin, Milwaukee, WI; 4) Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 5) Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 6) Department of Internal Medicine, University of Utah, Salk Lake City, Utah.

Backgrounds: Analysis of variants from next-generation sequencing (NGS) studies of complex diseases poses a great challenge because of the large number of rare variants, effects of which may vary in size and direction. Many existing methods 'collapse' clusters of rare variants to detect aggregate effects, and their power can be seriously compromised when the clusters contain only a few informative variants or when the effects are inconsistent. On the other hand, set enrichment analysis, which has been widely used for microarray gene expression and pathway-based genome-wide association studies, remains powerful even when a large number of irrelevant variants exist in the set. Methods: A recently developed method called Variable Set Enrichment Analysis (VSEA) was applied to the exome sequencing data of 21 African American subjects from the HyperGEN (Hypertension Genetic Epidemiology Network) study, to prioritize genes associated with a left ventricular mass trait (LVMHT27). It evaluates the enrichment of association signals indiscriminate of rare or common variants within genes, based on SNP level statistics. We tested 3 SNP level statistics including the t-score, log likelihood ratio, and the regression coefficient in a linear model. Their performance was compared with that of a step-up collapsing method. Results for chromosomes 17 and 18 are completed and reported below. Results: Among the 1192 genes that have data on chromosomes 17 and 18, VSEA test returns 52~68 genes with false discovery rates (FDR) q-value<0.05, depending on the type of SNP level statistics in test. Despite the small sample size, VSEA detected many known candidate genes for heart disease and related traits, including ACE (min q-value=0.0021), ITGA3(0.0046), IGF2BP1(0.0051), ALOX15(0.021), and LPO(0.041). Genes from several highly related families, like calcium channel, potassium channel, integrins, the solute-carrier family, and the Wnt family, were also identified, though the relationship of these genes with heart diseases is less clear. In contrast, the test using step-up collapsing only discovered two solute-carrier family genes (p=0.042~0.046). Conclusion: The analyses demonstrate power of the variable set enrichment analysis empirically in presence of large number of rare and common variants. While more research is needed to evaluate the influence of SNP-level test statistics, the VSEA approach seems to provide a simple and effective way to prioritize genes in NGS studies.

780W

Novel IRF6 Mutations in Families with Van Der Woude Syndrome and Popliteal Pterygium Syndrome from Sub-Saharan Africa. *A. Butali[1], P.A Mossey[2], W.L Adeyemo[3], M.A Eshete[4], L.A Gaines[1], R.O Braimoh[5], S.B Aregbesola[5], R. Rigdon[1], C.I Emeka[3], O. James[3], M.O Ogunlewe[3], A.L Ladeinde[3], A. Fikre[4], M. Ibrahim[4], E.G Paul[4], A.A Adeyemo[6], J.C Murray[1].* 1) Pediatrics, Univ Iowa, Iowa, IA; 2) University of Dundee, UK; 3) University of Lagos, Nigeria; 4) Addis Ababa University, Ethiopia; 5) Obafemi Awolowo University, Nigeria; 6) National Institute for Health, MD.

Orofacial clefts (OFC) are complex genetic traits that are often classified as syndromic or non-syndromic clefts. To date, over 500 syndromic clefts have been identified and included in the Online Mendelian in Man (OMIM) database and for many of these; the underlying genetic factors have been identified. Van der Woude syndrome (VWS) is an autosomal dominant disorder affecting 1/40,000 people worldwide and is one of the most common syndromic clefts, accounting for 2% of all OFC. Popliteal pterygium syndrome (PPS) is a rare dominant disorder affecting 1/300,000 people worldwide and in most cases is an allelic, but more severe form of VWS. Variants in the IRF6 gene have been reported world-wide to cause VWS and PPS. We conducted Sanger sequencing of the high risk exons for IRF6 (exons 3, 4, 7 and 9) in four families with VWS and one family with PPS from Nigeria and Ethiopia. For the VWS families, we found a novel nonsense mutation in exon 4 (p.Lys66X), a novel splice site mutation in exon 4 (p.Pro126Pro) and a previously reported splice mutation in exon 7 that changes the acceptor splice site. A previously known missense mutation was found in exon 4 (p.Arg84His) in the PPS family. All the mutations segregate in the families. Our data confirms the presence of IRF6-related VWS and PPS in sub-Saharan Africa and highlights the importance of screening for novel variants in known genes when studying diverse global populations. This is important for counseling and prenatal diagnosis for high risk families, especially in Africa where there are limited genetic counselors. Funding: This study was support by the NIH/ NIDCR Grants K99/R00 5K99-DE022378-02, R37-DE08559 and Face Base grant U01 DE-20057.

781T

A replication study for four keloid loci at 1q41, 3q22.3-23 and 15q21.3 in the Japanese population. *A. Watanabe[1, 2], R. OGAWA[3], M. SASAKI[2], A. FUJITA[2], B.T. Naing[1], S. AKAISHI[3], H. HYAKUSOKU[3], T. SHIMADA[1, 2].* 1) Dept Biochem & Molec Biol, Nippon Med Sch, Tokyo, Japan; 2) Div Pesonalized Genetic Med, Nippon Med Sch Hosp, Tokyo, Japan; 3) Dept Plastic and Reconstructive Surgery, Nippon Med Sch Hosp, Tokyo, Japan.

Keloid is a dermal fibroproliferative growth that results from dysfunction of the wound healing processes. A previous genome-wide association study (GWAS) reported four novel keloid-susceptibility loci at 1q41, 3q22.3-23 and 15q21.3 (Nakashima M, et al. Nature Genet. 2010;42,768-772). Here, we investigated the association of these loci with keloid by using an independent Japanese sample set. We performed case-control association analysis using 204 patients with keloid subjects in Nippon Medical School Hospital. The four SNPs at 1q41, 3q22.3-23 and 15q21.3 were genotyped using Small Amplicon Genotyping method. In this study, the two from four loci identified by previous GWAS, rs873549 on chromosome 1 (odds ratio (OR) =4.8, 95% CI: 1.941-14.192) in recessive manner and rs8032158 located in NEDD4 on chromosome 15 (OR= 3.019, 95% CI: 1.712-5.308) in dominant manner were significantly associated in keloid compared with control. Especially, in rs8032158, the risk for severe type of in the scar severity scale (OR= 2.563, 95% CI: 1.264-5.238) increased in comparison with mild type. Individuals with two risk alleles of the two loci showed higher risk of developing keloid than those with one risk allele alone. Our findings elucidated the significance of genetic variation at these 2 loci in keloid in the Japanese population.

782F

Identification of potentially causative variants underlying triglyceride levels. *S.M. Raj[1], A. Coventry[1], C. Ballantyne[2], C.F. Sing[3], R. Gibbs[2], E. Boerwinkle[2, 4], A.G. Clark[1].* 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI 48105; 4) Human Genetics Center, Health Sciences Center, University of Texas, Houston, TX 77030.

There is enormous incentive to identify the causative variants that underlie associations identified in GWA studies. Here we present an approach to optimize the allocation of sequencing resources for GWAS follow-up studies. The method specifies both informative individuals and the specific genomic regions to resequence, and it specifies an analytical approach to identify phenotypic effects. This approach was applied to a known GWAS signal for serum triglyceride levels in European-Americans from the Atherosclerosis Risk in Communities (ARIC) study. We also explored methods for querying GWAS genotype data to identify rare variants responsible for differences in triglyceride levels. We first carried out single-marker association tests to identify markers that show the strongest association with triglyceride levels. The shared haplotype structure in the region of these associations specified a 110 kb region on chromosome 7 that we sequenced (to 30-fold depth) in 502 individuals from the tails of the distribution of triglyceride levels. Analysis of these data revealed strong associations with TRIG levels at candidate sites that also showed strong prior biological relevance. We identified a 2 kb region in *BCL7B* which is the target of many transcription-factor binding sites as a candidate for influencing triglyceride biology. The method used here lends stronger support for candidate regions to motivate biochemical or molecular-based validation of their role in disease etiology. Identification of potentially causative variants may therefore be useful in developing therapeutic applications.

783W

The contribution of regulatory variation to facial masculinity in humans. *A.A. Zaidi[1], P. Claes[2], W. Yao[1], K. Daniels[2], M.D. Shriver[3].* 1) Genetics, Pennsylvania State University, University Park, PA., United States; 2) University of Leuven, Leuven, Belgium; 3) Anthropology, Pennsylvania State University, University Park, PA., United States.

The genetic basis of sexual dimorphism in the human face is an understudied phenomenon. Here we present an objective measure of facial masculinity and that it has a bimodal distribution with a great deal of overlap between the two sexes (t=43.32, N = 999, p-value < 2.2 × 10-16). This dimorphism is likely to be under the influence of sex-specific hormones such as testosterone and estrogen. According to our hypothesis, sexual dimorphism in the face is a signature of sexual selection in humans and might be different across populations due to differences in mate preference. In order to test this, we compiled a list of putative androgen response elements (AREs), identified in a previous study using chromatin immunoprecipitation (ChIP) for the androgen receptor (AR) protein. This list was filtered for AR binding sites that also overlap with other potentially functional non coding regions such as DNAseI hypersensitivity clusters, using ENCODE data. A subset of these with high Fst were then selected to represent regions that have recently diverged across human populations. We then tested for association between genetic variants in these regions with overall facial masculinity and with specific regions of the face that are observed to be different between males and females in a sample of 416 individuals with varying levels of African and European ancestry. With this approach, we have found a number of putative androgen response elements that are significantly associated with facial masculinity.

784T

Genetic associations with high-grade dyaplasia and colorectal cancer in patients with colonic inflammatory bowel disease: preliminary results from ImmunoChip using a targeted analytic approach. *J. Knight[1,2], M. Waterman[3,4], J.M. Stempak[3], K. Krishnaprasad[5], I. Cleynen[6], L.P. Schumm[7], S. Vermeire[8], D.P. McGovern[9], S.R. Brant[10], G. Radford-Smith[5], M.S. Silverberg[2,3], International IBD Genetic Consortium.* 1) Neuroscience Research, Centre of Addiction and Mental Health, Toronto, Ontario, Canada; 2) University of Toronto, Toronto, ON, Canada;; 3) Division of Gastroenterology, Mount Sinai Hospital, Toronto, Toronto, ON, Canada; 4) Department of Gastroenterology, Rambam Health Care Campus, B. Rappaport Faculty of Medicine, The Technion - Israel Institute of Technology, Haifa, Israel; 5) Inflammatory Bowel Diseases Research Group, Queensland Medical Research Institute, Brisbane, QLD, Australia; 6) Translational Research in GastroIntestinal Disorders, K.U.Leuven, Leuven, Belgium; 7) University of Chicago, Chicago, IL; 8) Department of Gastroenterology, University hospitals, Leuven, Belgium; 9) Inflammatory Bowel and Immunobiology Research Institute, Cedars Sinai Hospital, Los Angeles, CA; 10) The Meyerhoff Inflammatory Bowel Disease Center, Johns Hopkins Hospital, Baltimore, MD.

Colonic IBD is a significant risk factor for colorectal cancer (CRC) and high-grade dysplasia (HGD). Extensive involvement of the colon, the co-existence of primary sclerosing cholangitis (PSC), disease activity, family history of CRC and disease duration have been shown to increase the risk for CRC. However, no specific genetic association has been repeatedly shown to be associated with CRC in patients with colonic IBD. We aimed to identify genetic associations with CRC/HGD in patients with colonic IBD using data from the ImmunoChip in a large multi-national cohort of patient with colonic IBD.

Members of the International IBD Genetic Consortium (IIBDGC) were asked to identify patients with colonic IBD who developed CRC or HGD, verified by pathology. Demographic and clinical data were also collected. For each HGD/CRC case 1-2 controls were matched (by IBD subtype, disease duration, endoscopic extent (Montreal Classification), ethnicity, co-existence of PSC and gender). Phenotypic variables were compared between CRC/HGD cases and controls using $\psi 2$ for categorical variables and student's t-test for continuous variables. Preliminary analysis of genotypes generated using the iChip data was undertaken using logistic regression with appropriate covariates. Specifically investigating recently-reported IBD associations, PSC associations and SNPs in CRC-associated genes.

Overall, 585 colonic IBD cases (390 UC, 171 CD, 19 IBDU) were reported, 274 had either HGD or CRC and 311 were matched controls. There were 58% males, 95% Caucasian, and 12% current smokers. The mean age at diagnosis was 33. Disease duration to CRC/HGD diagnosis was significantly longer than length of F/U in controls (230 months vs. 180 months, p=0.005). There was no significant difference in proportions of patients with extensive disease between the groups. However, PSC was more common in cases vs. controls (11.3% vs. 2.9%, p=0.0004). Genetic data from iChip were available on 120 cases, 227 controls (59.3% of the total cohort; 50% males). Only Caucasians were selected for analysis. There were no significant genetic associations.

In this initial analysis, no genetic associations could be found related to risk of CRC/HGD in IBD. The IIBDGC will continue to gather cases to increase the sample size.

785F

Association of genetic variations in anti-inflammatory cytokine pathway genes in the outcome of tuberculosis. *S. Aggarwal[1, 2], S. Ali[2], R. Chopra[2], A. Srivastava[2], R.N.K. Bamezai[2].* 1) Department of Biochemistry, All India Institute of Medical Sciences, Raipur, 492099 Chattisgarh, India; 2) National Centre of Applied Human Genetics, School of life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India.

Tuberculosis (TB) constitutes the major cause of death due to infectious diseases. Cytokines play a major role in defense against Mycobacterium tuberculosis infection. Polymorphisms in the genes encoding various cytokines have been associated with tuberculosis susceptibility or resistance. This study aimed to investigate potential associations of forty SNPs in anti-inflammatory cytokines (IL10, TGFB1, IL6, IL4 and IL13) and receptors (IL10RA, IL10RB, TGFBR1, TGFBR2, IL6R, IL4R, IL5RA, IL5RB and IL13RA1) with tuberculosis. Our study included 327 clinically categorized tuberculosis patients representing North Indian population and 1294 unrelated healthy individuals. Analysis of rs3024498 located in the 3'UTR region of IL10, was observed to be significantly associated with the risk towards tuberculosis (AA/AG+GG, OR=1.52, 95%CI=1.11-2.10, P=0.01). Other polymorphisms, rs7281762 of IL10RB, (GG/GA+AA, OR=0.7, 95%CI=0.55-0.89, P=0.004) provided protection and rs2228048 of TGFBR2 (TT/CT+CC, OR=2.7, 95%CI=1.2-6.02, P=0.02) risk towards the disease, respectively. Our results demonstrate that the polymorphisms in cytokine genes may be valuable markers to predict the protection or risk towards tuberculosis. In conclusion, this study provides an interesting cue to cumulative polygenic host component which regulates tuberculosis pathogenesis.

786W

Assessing genetic association and gene-gene interaction between PTPN22 and CSK for SLE susceptibility in Asians. *X.R. Kim-Howard[1], C. Sun[1], A. Adler[1], S.C. Bae[2], S.K. Swapan[1].* 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea.

Regulation of effector T cell responses are essential for maintaining immune homeostasis and tolerance. Protein tyrosine phosphatase non-receptor type 22 (PTPN22) negatively regulates T cell activation associated with multiple autoimmune diseases, including systemic lupus erythematosus (SLE). It encodes intracellular lymphoid-specific phosphatase (Lyp), which has a prominent role in lymphocyte activation and differentiation. While functional coding variant (R620W, rs2476601) is robustly replicated in European derived populations the risk variant is absent Asian populations. Regulatory variant (rs34933034) of CSK, encodes protein c-Src tyrosine kinase (Csk), is associated with SLE in European-Americans. Lyp physically interacts with Csk during activation of downstream Src kinases, which may increase susceptibility to SLE. Our goals are to (a) assess these variants in an Asian population (Korean) using imputation-based association, (b) explore gene-gene interactions, and (c) identify association with SLE clinical manifestations (i.e, lupus nephritis) and autoantibodies (i.e, anti-Ro, anti-La, anti-dsDNA). We genotyped SNPs from PTPN22 and CSK in 1710 SLE cases and 3164 controls from Korea. Next, a comprehensive imputation-based analysis was used to add out-of-study GWAS controls and increase SNP density using 286 Asians available from 1000Genomes Project. Case-only analysis used specific SLE clinical manifestations as the phenotype. Since we observed significant association for PTPN22 and CSK SNPs we tested for potential genetic interaction between SNPs. This data analysis is in progress. We analyzed 138 SNPs from PTPN22 and 57 SNPs from CSK. As expected, known PTPN22 SNP rs2476601 was monomorphic in our Korean cohort. The most significant PTPN22 SNPs were in complete LD (r2=1), intronic rs12746551 and rs33965092 (P=9.2×10-6, OR=2.09). In CSK published SNP (rs34933034) was not significant (P=0.30) in Koreans. The most significant SNP was intronic rs4886627 (P=4.0×10-4, OR=1.15). We observed suggestive interaction (P=0.07) between intronic SNPs rs33965092 (PTPN22) and rs11635664 (CSK). Clinical phenotype analysis is ongoing. Previously reported functional variants rs2476601 (PTPN22) and rs34933034 (CSK) were monomorphic and non significant, respectively. However, other potentially SNPs from both genes could be associated with SLE. Additionally, we observed suggestive SNP-SNP interaction.

787T

Correlation between SNPs within the MHC region and immune responsiveness to childhood vaccinations. *Y.A. Talzhanov[1], B. Yucesoy[2], V.J. Johnson[3], N.W. Wilson[4], R.E. Biagini[6], W. Wang[2], B. Frye[2], D.N. Weissman[5], D.R. Germolec[7], M.I. Luster[2], M.M. Barmada[1].* 1) Center for Computational Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Toxicology and Molecular Biology Branch, CDC/NIOSH, Morgantown, WV; 3) BRT-Burleson Research Technologies, Morrisville, NC; 4) Department of Pediatrics, School of Medicine, University of Nevada, Reno, NV; 5) Division of Respiratory Diseases Studies, CDC/NIOSH, Morgantown, WV; 6) Biomonitoring and Health Assessment Branch, CDC/NIOSH, Cincinnati, OH; 7) Toxicology Branch, DNTP/NIEHS, Research Triangle Park, NC.

The present study investigated association between genetic variability in major histocompatibility complex (MHC) and immune response to childhood vaccinations. 141 healthy infants who have been immunized, according to standard vaccination scheme, with hepatitis B (HBV), 7-valent pneumococcal conjugate (PCV7), and diphtheria, tetanus, acellular pertussis (DTaP) vaccines were participated in the study. Using multiplex immunoassays we have assessed vaccine specific antibody responses and total serum immunoglobulin levels (IgM, IgA, IgG and IgG subclasses). We collected genotype information using Illumina Goldengate MHC panels. Our analysis showed association between single nucleotide polymorphisms (SNP) within MHC region and vaccine specific antibody responses, serum levels of immunoglobulins (IgG, IgA, IgM) and IgG isotypes (IgG1-IgG3). Functional annotation of SNPs using RegulomeDB showed that both significant and correlated SNPs regulate the expression of a group of genes involved in antigen processing and presentation, including HLA-G, HLA-A, HLA-DRB5, HLA-DQA1, HLA-DQB1 and TAP-2. The results suggest that immune response to common childhood immunization is affected by genetic variations within MHC region, which in turn may influence vaccine efficacy. These novel regions within MHC may serve as candidates for future genetic and mechanistic studies.

788F

The CTLA4 gene and severe bacterial infections. *H. Chi[1,2,3], NC. Chiu[1,2], FY. Huang[1], SC. Chang[4], WF. Chen[4], WS. Lin[4], CL. Lin[4], YJ. Lee[1,4,5,6].* 1) Pediatrics, Mackay Memorial Hosp, Taipei, Taiwan, 92, Sec. 2, Zhongshan N. Rd., Zhongshan District Taipei; 2) Department of Nursing, Mackay Medicine, Nursing and Management College; 3) Department of Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine; 4) Department of Medical Research, Mackay Memorial Hospital Tamsui District; 5) Department of Biomedical Sciences, Mackay Medical College; 6) Department of School of Medicine, Taipei Medical University.

The CTLA4 gene encodes a T-cell receptor, namely, cytotoxic T-lymphocyte antigen 4 (CTLA4), which is involved in controlling the proliferation and apoptosis of T lymphocytes. Among them, rs3087243 G>A and rs231775 A>G are the most associated markers with autoimmune diseases. In this study, we investigated whether the CTLA4 gene was associated with severe bacterial infections in Han Chinese children. Subjects and Methods Patients The patients were 142 unrelated children (54 boys, 88 girls) with severe bacterial infections (SBI) (bacterial septicemia, meningitis, pneumonia, or cellulitis). Their age at diagnosis was 4.9 ± 3.6 years (range 0.0 - 17.7 years). Among them, 55 were infected by pneumococcus. The control subjects were 920 subjects. They included hospital personnel and individuals who underwent routine health examinations or minor surgery. All patients and control subjects were ethnic Han Chinese in Taiwan. Our institutional review board approved this study, and all subjects or their guardians gave informed consent. CTLA4 genotyping We genotyped -318 C/T, +49 A/G, and CT60 polymorphisms of the CTLA 4 gene by PCR-restriction fragment length polymorphism (PCR-RFLP) using MseI, BstEII, and NcoI enzymes, respectively as we have reported. The -318 C/T polymorphism was amplified with modified primers 5'-GTT AGG GAT GCC CAG AAG AT-3' and 5'-CTC AAC TGA ACA AAA CAA GC-3'resulting in a 172 bp product. The +49 A/G polymorphism was amplified with primers 5'-AAG GCT CAG CTG AAC CTG GT-3' and 5'-CTG CTG AAA CAA ATG AAA CCC-3' resulting in a 153 bp product. The CT60 polymorphism was amplified with primers 5'-CAC CAC TAT TTG GGA TAT ACC-3' and 5'-AGG TCT ATA TTT CAG GAA GGC-3' resulting in a 216 bp product. Statistical analysis Genotype, allele and carrier frequencies of the -318 C/T, +49 A/G, and CT60 polymorphisms of the CTLA4 genes were determined by direct counting. Agreement with Hardy-Weinberg equilibrium (HWE) was tested for genotype frequencies of the controls using Haploview 4.2 Results The -318, +49, and CT60 polymorphisms of the CTLA4 gene were in Hardy-Weinberg equilibrium in controls and in patients. Statistical analysis revealed that no significant difference in the frequencies of genotypes, alleles, carriers, or haplotypes of the 3 SNPs of the CTLA4 gene between patients and controls. The results were similar when patients with pneumococcal infections were analysed separately. Discussion More patients must be collected.

789W

Association of Toll-like receptor 4 with otitis media in the Finnish population. *E. Einarsdottir[1], L. Hafrén[2], E. Kentala[2], E. Leinonen[3], J. Kere[1,3], P.S. Mattila[2].* 1) Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden; 2) Dept of Otorhinolaryngology, Helsinki University Central Hospital, University of Helsinki, Finland; 3) Folkhälsan Institute of Genetics, Helsinki, Finland.

Background & aims: Otitis media (OM) is one of the most common infectious diseases in childhood, leading to doctor's visits and antibiotic prescriptions. Despite being an infectious disease, the risk of developing recurrent (RAOM) and chronic (COME) otitis media with effusion has a strong genetic component. Toll-like receptor 4 (TLR4), a receptor for LPS and an important component of the innate immune system, has been suggested previously as a plausible candidate for variation in OM susceptibility. Genetic variation within this locus has, however, not been studied comprehensibly previously. The aim of the current study was to get a clear picture of TLR4 involvement in otitis media in Finnish patients.

Study design: We studied 21 SNPs within and upstream of the TLR4 coding region. Markers were chosen to tag the common variation in the LD block within which TLR4 lies. Two rare coding variants (Asp299Gly and Thr399Ile) were also considered. 592 cases and 746 controls from Finland were analyzed. All participants gave their informed consent to participate in the study. Cases were recruited when coming for surgery due to RAOM or COME; controls were healthy blood donors. In addition to all OM, we also looked at patient subgroups with COME, RAOM, early onset of OM, and patients requiring multiple rounds of draining tubes. Plink was used to look at association of each marker & haplotypes to OM, Haploview was used to visualize LD patterns.

Results: We identified association of variants at the TLR4 locus to OM in all patient groups (all OM and all sub-phenotypes). The strongest association (p-uncorr = 0,004-0,000038) was found to two markers upstream of the TLR4 gene, with ORs approx. 1,3 in RAOM, COME and all OM. The two patient groups with the more severe phenotypes (early onset & multiple draining tubes) yielded association to the same markers, with ORs of 2,4 and 1,6, respectively. The previously studied rare coding variants in TLR4 were excluded as major risk variants for OM in Finland.

Conclusions: Our results suggest genetic variation within the TLR4 locus as an important modifier of otitis media susceptibility in the Finnish population.

790T

Multiple eQTLs of *TNFSF8* are associated with pathological immune responses in leprosy. *V.M. Fava[1,2], A. Cobat[1,2], V.T. Nguyen[3], N.N. Ba[3], M. Orlova[1], J. Manry[1,2], A.C.P. Latini[4], M.M.A. Stefani[5], M. Mira[6], V.H. Thai[3], L. Abel[7,8,9], A. Alcais[7,8,9], E. Schurr[1,2].* 1) The McGill International TB Centre, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Departments of Human Genetics and Medicine, McGill University, Montreal, Quebec, Canada; 3) Hospital for Dermato-Venereology, Ho Chi Minh City, Vietnam; 4) Lauro de Souza Lima Institute, Bauru, São Paulo, Brazil; 5) Tropical Pathology and Public Health Institute, Federal University of Goiás, Goiânia, Brazil; 6) Core for Advanced Molecular Investigation, Pontifical Catholic University of Paraná, Curitiba, Paraná, Brazil; 7) Laboratoire de Génétique des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale, U980, 75015 Paris, France; 8) Université Paris René Descartes, Faculté Médicine Necker, 75015 Paris, France; 9) Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 11065, USA.

A genome-wide association study (GWAS) in Chinese leprosy patients identified six genes associated with leprosy. While five of these genes have been validated, one gene, *TNFSF15*, has resisted all attempts of replication by independent groups. While evaluating the clinical characteristics of the leprosy GWAS sample, we noted that more than 85% of the patients presented with disabilities. Leprosy type 1 reversal reactions (T1R) are the major cause of nerve damage in leprosy patients due to excessive immune responses directed against Schwann cells. We hypothesized that the lack of validation of *TNFSF15* was in truth due to an association with the endophenotype T1R. To test our hypothesis, we studied SNPs in the *TNFSF15* genomic region as risk factors for T1R in three independent populations: a family based sample from Vietnam and two case-control samples from Brazil, comprising a total of 1768 subjects. For the family-based design, we used a set of leprosy families without T1R as controls. In Brazil, the cases were T1R patients while the controls were leprosy patients without T1R. In total, we genotyped 56 SNPs across all three samples. In the Vietnamese sample, eight SNPs were strongly associated with T1R ($p<0.001$). Unexpectedly, the majority of SNPs were in four bins that extended to the neighboring *TNFSF8* gene and not to the physically closer *TNFSF15* gene. Indeed, of the 19 cis-eQTL for *TNFSF8* described in public data bases, 17 had $p < 0.05$ in support of association with T1R. Multivariate analyses revealed independent association of two bins that carried numerous eQTLs for *TNFSF8*. The bins associated with T1R in the Vietnamese sample were fragmented in the Brazilian samples. Yet, two bins were validated. In Goiania, four SNPs of an eQTL bin displayed $p<0.05$ while in Rondonópolis eleven distinct SNPs presented $p<0.05$ in favour of association with T1R. Taken together, our results identify a set of genetic risk factors for T1R in Vietnam and Brazil. These risk SNPs are eQTLs for *TNFSF8* that despite of being in the same SNP bin may independently act on *TNFSF8* expression levels. Our results illustrate the importance of taking in account endophenotypes in genetic association studies.

791F

Immuno-genetic role of mannan-binding lectin in Indian visceral leishmaniasis. *A. Mishra[1], A. Nath Jha[1], H. V. Tong[2], L. Singh[1], T. P. Velavan[2], K. Thangaraj[1].* 1) Centre for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India; 2) ITM, University of Tubingen, Wilhelmstr, Germany.

Introduction Visceral leishmaniasis (VL) is endemic in middle east region of India and accounts for more than 90% of Indian VL and always fatal if untreated. It is multifactorial disease where host immunity and genetic factor play their role in outcome of VL. The mannose-binding lectin (MBL) play important role in innate immunity. In human mutation in MBL2 gene (chromosome 10), affects level of MBL in serum. Higher MBL level is associated with disease progression while lower level protect against development of VL. The aim of this study is to define the role of MBL2 immuno-genetic factor in Indian VL. Method The study was designed as matched case-control study (by caste, age group, sex and geographical location) and conducted in endemic region of Indian VL. Total 322 unrelated subjects (197 case and 125 Controls) were selected for study, Mean age of cases were (29.62+/- 16.95) while controls have (32.98+/- 15.04); the male:female ratio in cases were (113:84) while in controls were (69:56). Asertainment of cases was based on presence of typical clinical features of VL while control subjects had never suffered from or been treated for VL or other infectious disease. We have analyzed MBL serum level (ELISA) and several regulatory polymorphisms at promoter and exon 1 (genotyping) in MBL2 gene via case-control approaches. Result Serum level of MBL were higher ($p = 0.027$) in individuals with VL than in controls. We have not found any association in genotype and allele frequency distribution along with haplotype carriage between case-control groups of VL, although haplotype LYQC is significantly associated (OR 0.4 (CI 0.15-1), p = 0.48) but present in 7.14% subjects. Conclusion This is first case-control genetic study which indicates level of MBL in serum and distinct polymorphism role of MBL2 gene in Indian VL (no association). This study is totally different from Brazil study (another major foci of VL), overall it reflect the clinical complication and unique genetic architecture of Indian population.

792W

Association of single nucleotide polymorphisms (SNPs) in the gene coding for the lens epithelium-derived growth factor (LEDGF/p75) with the outcome of HIV infection in Brazilian HIV+ patients. *C.PB. Passaes[1,2,3], C.C. Cardoso[4], D.G. Caetano[3], S.L.M. Teixeira[3], M.L. Guimarães[3], D.P. Campos[5], V.G. Veloso[5], M.O. Moraes[6], M.G. Morgado[3].* 1) Service d'Immuno-Virologie Commissariat à l'énergie atomique et aux énergies alternatives, CEA, Fontenay aux Roses, France; 2) Unité de Régulation des Infections Rétrovirales, Institute Pasteur, Paris, France; 3) Laboratório de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz - Fiocruz - Rio de Janeiro, Brasil; 4) Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil; 5) Instituto de Pesquisa Clinica Evandro Chagas, Fiocruz, Rio de Janeiro, Brazil; 6) Laboratório de Hanseníase, Instituto Oswaldo Cruz - Fiocruz - Rio de Janeiro, Brasil.

The lens epithelium-derived growth factor p75 (LEDGF/p75), coded by the *PSIP1* gene, is an important host co-factor which interacts with HIV integrase (IN) to target integration of viral cDNA into active genes. The aim of this study was to investigate the association of SNPs in PSIP1 with disease outcome in Brazilian HIV infected patients. We performed a case-control study in a cohort of 171 HIV+ patients followed up at Instituto de Pesquisa Clínica Evandro Chagas and classified as rapid progressors (RP, n=69), typical progressors (TP, n=79) and long-term non progressors (LTNP, n=23) according to the time of HIV infection, CD4+ T cells counts, date of the last negative serology and presence of an AIDS-defining events. A group of 96 healthy volunteers was used as population controls. The *PSIP1* exons 8-14 (coding for the integrase binding domain) were sequenced and the SNP rs61744944 was characterized. To increase gene coverage, 9 tag SNPs were selected from HapMap database and genotyped using TaqMan assays. The control population was analyzed separately to determine the frequency of each SNP, as well as deviations from Hardy-Weinberg Equilibrium (HWE). Pairwise linkage disequilibrium (LD) patterns were also estimated for this population using the r^2 coefficient. The association between PSIP1 SNPs/haplotypes and disease progression was assessed by logistic regression models. After HWE analyses, a total of 8 SNPs were characterized. No additional SNPs were identified in PSIP1 exons 8-14. There were no clear LD patterns among the SNPs analyzed ($r^2<0.8$). The logistic regression models suggested that patients carrying the allele T at rs61744944 were more likely to develop the LTNP outcome (OR = 4.98; $_{95\%}$CI = 1.03 - 24.19; p = 0.05) using TP as control group. The same trend was observed when LTNPs were compared to the RP group (OR = 3.26). These results were reinforced when all SNPs were combined, since the haplotype rs61744944.T/rs17337140.G/rs2737829.C/rs10119931.C/rs10283923.G/rs10962048.G/rs7470146.G/rs2277191.G was associated to the LTNP status (OR = 6.05; p = 0.08 f and OR = 3.44; p = 0.12 for comparisons against TP and RP, respectively). In spite of the limited number of LTNPs, these data suggest that PSIP1 gene could be associated to the outcome of HIV infection. Further analyses of this gene may guide the identification of causative variants to help predicting disease progression.

793T

Identification of genetic susceptibility loci for dengue shock syndrome at the PLCE1 gene using a targeted re-sequencing approach. *E. Png[1], T.N. Chau[2], N.M. Nguyet[2], D.T. Kien[2], N.T. Quyen[2], D.T. Trung[2], J. Pang[1,3], B. Wills[2,4], N. Van Vinh Chau[5], C.C. Khor[1,3,6], C.P. Simmons[2,4], M.L. Hibberd[1,3].* 1) Infectious Diseases, Genome Institute of Singapore, Singapore, Singapore; 2) Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam; 3) School of Public Health, National University of Singapore, Singapore; 4) Centre for Tropical Medicine, University of Oxford, Oxford, United Kingdom; 5) Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam; 6) Department of Paediatrics and Department of Ophthalmology, School of Medicine, National University of Singapore, Singapore.

In our recent genome wide association study (GWAS) of dengue shock syndrome (DSS), we identified common SNPs within the PLCE1 gene to be significantly associated. These SNPs reside within a 50Kb region with high linkage disequilibrium and hence, could be merely tagging causal variant(s). To fine map and identify functionally important variants, we performed targeted re-sequencing of this region of interest (ROI) in 240 DSS cases and 240 controls from Vietnam . We used the Fluidigm Access Array to perform long range PCR to generate 1Kb amplicons to cover the ROI. An aliquot of the pooled amplicons was used to generate library using the NexteraXT kit from Illumina. The libraries were barcoded and sequenced at 96 plex per lane using 100bp paired end reads on Illumina HiSeq 2000 platform. The sequenced data was aligned to the human reference (Hg 19) using BWA software, and genetic variants were identified using GATK Unified Genotyper. On average each library yielded 191Mb of sequences that allowed each targeted base of the ROI to be covered ≤4000x. After performing quality control filters, we identified 414 variants in 429 samples, of which 70% were novel. We performed association analysis on this dataset and found 23 variants exceeding a nominal $P \leq 0.05$. The smallest P value was observed at rs147206950 (P 0.005, OR= 0.36, 95% C.I. 0.17- 0.76), which had a significantly larger effect size than the previously reported GWAS SNP, and is predicted to be a regulatory variant. The eight sequencing SNPs that overlapped with the previous Illumina Human 660W BeadChip data had >99% concordance, emphasizing the confidence of the sequencing SNP calls. We further used this dataset as reference to impute the genotype of all the variants found through sequencing into the entire GWAS dataset of 2,008 cases and 2,018 controls. We found we could accurately recapitulate the previous GWAS associations and also identified an additional 22 variants with predicted P values $\leq 1 \times 10\text{-}7$. We next plan to perform de-novo genotyping of these variants in an additional 1,800 DSS cases and 2,900 controls from Vietnam, looking to validate these new associations and identify the causative variant(s).

794F

Nha Trang Birth Cohort Study: a population-based study for the identification of phenotype expression of genetic risk factors for childhood infectious diseases in Vietnam. *M. Yasunami[1], R. Miyahara[1], L.-M. Yoshida[1], H. Nakamura[1], K. Takahashi[1], D.T. Vu[2], X.M. Trinh[3], T.C. Bui[3], D.A. Dang[2], H.T. Le[4], K. Ariyoshi[1].* 1) Clinical Medicine, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; 2) National Instituteof Hygiene and Epidemiology (NIHE), Hanoi, Vietnam; 3) Pasteur Institute Nha Trang, Nha Trang, Vietnam; 4) Khanh Hoa Health Service Department, Nha Trang, Vietnam.

The differential susceptibility to certain infectious disease is attributable to genetically-determined responsiveness to microbial pathogens in part. The genetic polymorphisms associated with functional variations in Toll-like receptors (TLRs) are candidates for such risk factors of infection/infectious diseases. We conducted a population-based study of innate immune phenotype analysis for the better understanding of genetic predisposition to infectious diseases. 1,999 newborn babies, who were born in Khanh Hoa Provincial Hospital, a principal medical facility in Nha Trang City, central Vietnam from May 2009 to May 2010, were enrolled to a longitudinal study. Genomic DNA samples were prepared from umbilical cord blood collected at the entry. The events of severe diseases were passively but comprehensively identified by admissions to the pediatric ward of the same hospital in which they were born. The cohort children were invited to medical check at community health stations at 24 month of age and venous blood samples were collected upon their guardians' informed consent. A simple procedure of small-scale whole blood culture was established; fresh whole blood was mixed with cell culture medium at 1:9 ratio and incubated at 37 degree in the presence of LPS, Pam3CSK4 or other chemical stimulants of immune-competent cells. Twenty hours later, the cells were collected for RNA preparation. The levels of induced mRNA expression of cytokine genes such as IL8 and those of internal standards such as RPLP0 were quantified by real time RT-PCR assay using Fluidigm Biomark platform. Among 1,999 cohort children, 1,494 were responded to the invitation to 24-month medical check and 1,348 blood samples were successfully collected. The first 368 blood samples were examined for gene expression phenotype by the small-scale whole blood culture. The incidence of severe acute lower respiratory tract infection (ALRI) was 77.3/person/year (231 events in 1,494 persons in two years). A major haplotype of eight SNPs in the TLR10-TLR1-TLR6 region (4p14) was associated to ALRI as well as basal and inducible expression of several immune-related genes including chemokines CCL3 and IL8, and TLR1 itself. We observed a significant non-additive interaction between genetic polymorphism (a dose-effect of SNP haplotype) and gene expression phenotype (classification of multiple chemokine mRNA levels by cluster analysis) to the establishment of the risk for ALRI (hazard ratio of approx. 3).

795W

Systematic characterization of allelic architecture at the non-coding type 2 diabetes locus 9p21 using haplotype analysis in complete sequence data, imputation, and functional testing. *V. Agarwala[1,2,3], J. Wright[2], J. Flannick[2,4], A. Morris[5], M. Rasmussen[6], A. Mahajan[5], H. Kang[7], C. Fuchsberger[7], M. Kellis[6], M. Boehnke[7], M. McCarthy[5], D. Altshuler[2,4,8], The Go-T2D Consortium.* 1) Program in Medical & Population Genetics, Broad Institute, Cambridge, MA; 2) Harvard-MIT Division of Health Sciences and Technology, MIT, Cambridge, MA, USA; 3) Program in Biophysics, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA; 4) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Computer Science and Electrical Engineering Department, MIT, Cambridge, MA USA; 7) Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 8) Department of Genetics, Harvard Medical School, Boston, MA, USA.

The causal variant(s) underlying signals discovered in genome-wide association studies for complex traits are, in most cases, unknown. The allelic architecture at these loci - the number, frequencies, and effect sizes of causal mutations - is not yet understood, in part because GWAS directly tested only common SNPs. The non-coding chr9p21 GWAS locus (near CDKN2A/B) harbors one of the strongest known signals for type 2 diabetes (T2D). Interestingly, multiple independent signals at this locus create a three-tiered (risk, neutral, and protective) haplotypic association. The availability of whole-genome sequencing data enabled us to test, for the first time, three current hypotheses about allelic architecture at 9p21: (1) that common variant(s) causally modulate risk of T2D, (2) that rare variants create 'synthetic' common variant associations, and (3) that rare variants (individually or in aggregate) have effects on T2D, independent of the common signals. To systematically test these hypotheses, we identified all variants, common and rare, across a 5Mb region around the 9p21 locus in ~2800 whole-genome sequenced Europeans. We confirmed the known haplotypic association, and enumerated all variants present on the risk/protective haplotypes by integrating estimated phase data with information from ancestral recombination graphs. We tested variants, individually and in combination, for (a) association to T2D, and (b) ability to explain the common signals. We performed analysis in the sequenced sample, and also after imputation into ~35K samples densely genotyped on Metabochip. Observed genetic data place an upper bound on the contribution of rare variants at 9p21. Moreover, we reject the synthetic association hypothesis: we find no set of rare variants (excluding singletons) that can explain the haplotypic association, suggesting that causal variants underlying this GWAS signal are indeed common. After imputation, the credible set of variants with >99% probability of explaining the signal included only 1 common variant on the risk and 4 common variants on the protective haplotypes. Lastly, we conducted functional screens for regulatory activity across the 9kb region in multiple cell types. These experiments identified a sub-fragment which has beta cell-specific activity, and which includes the sites of the top candidate causal variants. The approach described here represents a general method for elucidating the causal architecture at disease-associated loci.

796T

A reliable algorithm for identifying modifier genes and the mutants who carry them in family-based studies of complex disorders. *P. Belleau[1], S. Dubois[1], S. Desjardins[1], R. Arsenault[1], E. Shink[1], J.L. Anctil[2], G. Côté[2], M.A. Walter[3], M.A. Amyot[4], V. Raymond[1].* 1) Neurosensory Systems Laboratory, CHUL Research Center (CHU de Québec), Quebec City, QC, Canada; 2) Ophthalmology, Universite Laval, Quebec City, QC, Canada; 3) Medical Genetics, University of Alberta, Edmonton, AB, Canada; 4) Ophthalmology, Universite de Montreal, Montreal, QC, Canada.

Our goal is to characterize gene-gene interactions implicated in open-angle glaucoma, a genetically complex disorder of the eye. Using a pedigree-based genome-wide linkage strategy and 154 heterozygotes for the autosomal dominant glaucoma myocilin K423E mutation, we recently mapped at 20p13, a modifier locus for variability of ages-at-onset (AAO). To characterize this modifier, we designed a pedigree-based algorithm that optimizes the identification of reliable double-mutants for genome-targeted sequencing. Our three-stage algorithm exploits a dataset obtained after genotyping 184 pedigree members with microsatellite and SNP markers covering the locus. The 1st stage is the identification of alleles associated with the modifier. The 2nd stage is the identification of double-mutants which simultaneously carry $MYOC^{K423E}$ and 1 marker allele associated with extreme values for AAO within the distribution of all AAO. The 3rd stage is haplotypes building starting with the alleles defined in 2. More formally, we first define Φ as the kinship coefficient ($\Phi(X,Y) \geq 0.0625$ meaning that X and Y are people closer or equal to first degree cousins). Second, let D be a double-mutant carrying $MYOC^{K423E}$ and the allele A under study. The neighborhood of D is then defined as X, the set of MYOC carriers **not** carrying A where $\Phi(D,X) \geq 0.0625$. We compute for each $MYOC^{K423E}$+A double-mutant the difference between this double-mutant AAO and the median of the AAO of his/her neighborhood. For all double-mutants, we then evaluate the median of the preceding difference and call this median I. We next compare I with the value Ip (p: permutation) calculated from permutations of AAO within the respective neighborhood of each double-mutant of the pedigree. The alleles for which the proportion of Ip > I is max. or min. are considered associated with the modifier. From the pool of $MYOC^{K423E}$ heterozygotes which carry the **associated alleles (As)**, we only keep for sequencing the double-mutants who are at the extremes of the distribution of the AAO. Conversely, we keep as controls, $MYOC^{K423E}$ carriers who **do not harbor** the As alleles at the opposite extremes. Finally, we test for **associated haplotypes** starting at the As alleles. In conclusion, we designed a reliable pedigree-based algorithm that selects those unequivocal double-mutants that participate in gene-gene interactions. This algorithm can be applied to other quantitative traits and to other types of allelic markers for complex traits.

797F

Fine-mapping of Type 2 Diabetes susceptibility loci via trans-ethnic meta-analysis. *M. Horikoshi[1, 2], J.R. Huyghe[3], F. Takeuchi[4], A. Mahajan[1], J. Asimit[5], E. Zeggini[5], N. Kato[4], YY. Teo[6], A.P. Morris[1], M.I. McCarthy[1, 2], T2D-GENES.* 1) The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 3) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Michigan, US; 4) Research Institute, National Center for Global Health and Medicine, Tokyo, Japan; 5) Wellcome Trust Sanger Institute, Cambridge, UK; 6) Department of Statistics & Applied Probability and Department of Epidemiology & Public Health, National University of Singapore, Singapore, Singapore.

Nearly 70 type 2 diabetes (T2D) susceptibility loci have been identified to date, but the underlying causal variant(s) are not well characterized. One approach to fine-map these loci is to combine genome-wide association (GWA) data from diverse populations to increase sample size and leverage differences in linkage disequilibrium (LD) between ethnicities. We previously reported trans-ethnic fine-mapping analyses undertaken in five T2D loci: *CDKAL1*, *CDKN2A/B*, *FTO*, *IGF2BP2* and *KCNQ1*, which demonstrate strong signals of association and differential patterns of LD across ancestry groups. In the current analysis, we increased the sample size to 64,727 (21,997 T2D cases and 42,730 controls), representing five ancestry groups (European, East Asian, South Asian, Hispanic and African American), and combined GWA data imputed up to the 1000 Genomes March 2012 reference panel of 1092 individuals from diverse ancestry groups. Using trans-ethnic meta-analysis, we defined 'credible sets' of SNPs at each locus that have 99% probability of containing the causal variant, and evaluated the evidence for multiple independent association signals via conditional analyses. None of the lead SNPs at any of the loci showed substantial evidence for heterogeneity in allelic effects between ethnicities, suggesting that the underlying causal variants are shared across ancestry groups and are amenable to trans-ethnic fine-mapping. By comparing credible sets of SNPs identified through ancestry specific and trans-ethnic meta-analyses, we saw improved resolution in fine-mapping at all five loci by combining data from diverse populations. The resolution was most improved at *CDKAL1*, where the 99% credible set included 5 SNPs mapping to 12.3kb, compared with 15 SNPs in 34.4kb in East Asians and 8 SNPs in 40.5kb in Europeans, South Asians and Hispanics. We observed evidence for multiple independent association signals at genome-wide significance at *KCNQ1* and *CDKN2A/B* after conditional analysis. At *KCNQ1*, we observed three association signals defined by rs2237896 ($P_{cond}=9.5\times10^{-12}$), rs234864 ($P_{cond}=8.4\times10^{-9}$), and rs231353 ($P_{cond}=1.7\times10^{-11}$). At *CDKN2A/B*, there are two distinct signals defined by rs10965248 ($P_{cond}=7.2\times10^{-40}$), and rs10757282 ($P_{cond}=2.0\times10^{-12}$), which capture the previously reported haplotype effect at this locus. Our study demonstrates the benefits of trans-ethnic analysis for dissecting association signals at GWA loci and prioritizing regions for functional follow-up.

798W

Fine mapping of inflammatory bowel disease risk loci using immunochip. *H. Huang[1,2], L. Jostins[3], KK. Farh[1,2], B. Bulik-Sullivan[1,2], The International IBD Genetics Consortium.* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK.

Inflammatory bowel disease (IBD) is an autoimmune disorder that affects 2.5 million people of European ancestry. There are two etiologically related forms of IBD, Crohn's disease and ulcerative colitis. The latest study in IBD genetics reported 163 risk loci. Comprehensively assessing variants in these loci gives insights into the genetic mechanisms of IBD. Here we propose a Bayesian based fine mapping approach that can define the number of independent association signals in each locus, the disease form of each signal and a minimal set of variants that we are 95% confident to contain the causal variants.

We constructed a posterior probability function to characterize the relation between genomic variants and the observed diseases. The most probable causal variants were identified by maximizing this function. Due to the LD, multiple variants may be tied for the causal signal. We therefore reported a credible set of variants such that the sum of their posterior probabilities ≥ 95%;. We apply this approach to 66,849 QC'ed European samples from the International IBD Genetics Consortium's immunochip project (18,603 CD, 14,307 UC and 33,938 controls). Out of 186 high-density regions in the immunochip, 76,862 variants in 110 regions have overlaps with the reported risk loci. These variants were imputed to the 1000 genome reference. 237,736 variants passed the imputation quality threshold and were used in the fine mapping.

We found that known causal variants were usually mapped as the only or the top variant in the credible set. For example, 3 NOD2 causal variants, fs1007insC, R702W and G908R were the only credible variants for the first 3 independent associations respectively. Out of the 110 fine-mapped regions, 50 have multiple independent associations and 78 have genome-wide significant (GWS) primary associations. Here we report fine-mapping results for these 78 associations. We reduced 11 associations to a single variant and 33 associations to ≤5 variants. Credible sets for 34 associations were localized to a region < 10Kb (28 of them have ≤ 5 variants). 13 associations have coding variants in the credible sets and 2 of them are the only variant. 9 out of the 11 single-variant credible sets are non-coding, suggesting high confident non-coding causal variants. We observed that in general, sizes of the credible set decrease with sample size, or the significance of the association signal.

799T

Exome array analysis of rare and low-frequency variation in multiple sclerosis. *M. Mitrovic[1], C. Cotsapas[1,2,3], International Multiple Sclerosis Genetics Consortium.* 1) Department of Neurology, Yale University School of Medicine, New Haven, CT; 2) Department of Genetics, Yale University School of Medicine, New Haven, CT; 3) Medical and Population Genetics, the Broad Institute of Harvard and MIT, Boston, MA.

Multiple sclerosis (MS) is a common complex disorder in which genetically susceptible individuals encounter environmental triggers and initiate an inflammatory reaction against self-antigens in the central nervous system. As a consortium, we have performed several genome-wide association studies (GWAS) and identified almost 100 loci harboring common variants contributing to MS risk. However, these only explain a fraction of the total heritability of disease risk. Low-frequency and rare variants may explain a further portion of this heritability and define new loci modulating pathobiology. To assess the contribution of this fraction of genetic variation to MS susceptibility, we are currently genotyping 42,000 cases and 36,000 unaffected controls drawn from our diverse collection using the Illumina HumanExome beadchip. We will test the following hypotheses: i) rare and/or low-frequency coding variants collectively contribute significantly to the disease risk; ii) genes perturbed by those variants form clusters with group enrichment for association; iii) cumulative rare/low-frequency scores are conditional on common variants; and iv) whether genes with evidence of conservation are more burdened in cases than controls. In the context of this experiment we aim to establish the role of rare variation in the architecture of common disease. MS is a prototypical common, complex disease with a polygenic architecture that has been tractable to GWAS. It thus serves as a model for other such diseases and the conclusions we draw here are likely applicable across the spectrum of phenotypes currently under genetic investigation.

800F

Variations in *Shank3* in Colombian individuals with Autism Spectrum Disorders. *L. Munera, P. Katherine, M. Oscar, M. Lattig, Liga Colombiana de Autismo (LICA).* Universidad de los Andes, Bogotá, Colombia.

Autism spectrum disorders (ASDs) are a group of common and very highly hereditable neurodevelopmental impairments, characterized by the presence of three main behaviours: alteration of social interactions, abnormal language development and the presence of stereotype behaviours. These disorders are highly heterogeneous and in some cases its cause can be explain by highly penetrant de novo mutations; or by the impact of various mutations in a particular individual, each with a moderate effect, that act in synergy producing the autistic behaviours. It has been shown that a variety of genes are associated with ASDs and interact with the neuroligins and neurexins in the glutamatergic synapsis, being SHANK3 one of these genes. The relationship between SHANK3 and ASDs has been demonstrated with diverse genetic and functional studies that suggest a role of this gene in the pathogenesis of autism and in the formation of the dendritic spines. To recognize the importance of variations on this gene on our population, we made a screening of variations in a group of ten Colombian individuals with a diagnostic of ASDs and we determined the possibly pathogenic effect of each of these changes. We found 9 variations, of which just one was found in the exonic region. As far as the intronic variations two changes occurred (rs113644328, MAF=0.005 and rs148315568, MAF=0.004) and stood out because of their low frequency, regardless, the bioinformatic analysis of these two changes did not show any significant importance. A missense mutation in exon 19 (rs61729471; p.Ala721Thr) with low frequency (MAF= 0.012) was found in heterozygosity in one individual. Based in the structure and conservation of this position, the bioinformatic analysis suggest that this change produces a possible reduction in the stability of the protein; and the evaluation of different scenarios indicate that the effect might possibly be due to changes in the protein phosphorylation. Given the heritability of this change from an apparently healthy mother and the presence of the variation in a control population of other study, it can be presumed that even though this change is not responsible of ASDs in this individual, it might have a small effect which in synergy with other changes produces the autistic phenotype.

801W

The Role of Mitochondrial Variants in the Risk of Age-related Macular Degeneration (AMD) in the National Health and Nutrition Examination Surveys (NHANES). *N. Restrepo, R. Goodloe, S. Mitchell, D. Murdock, J. Haines, D. Crawford.* Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Substantial progress has been made in identifying susceptibility variants for age-related macular degeneration (AMD), one of the most common causes of adult onset blindness. Major autosomal risk factors include Complement Factor H (CFH) Y402H (rs1061170) and Age-related Maculopathy Susceptibility-2 (ARMS2) rs10490924. Non-autosomal variants such as those in the mitochondria also contribute risk towards AMD. To further characterize the association between AMD risk and mitochondrial variants, we as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed the National Health and Nutrition Examination Surveys (NHANES) III and 2007-2008 for targeted genotyping of 60 mitochondrial SNPs. Fundus photography was available for participants >40 years. AMD cases were defined among older adults >60 years that presented with early/late AMD. Controls were defined among older adults >60 years without evidence of AMD. A total of 332, 37, 47 cases and 1500, 430, 270 controls from non-Hispanic whites, non-Hispanics blacks, and Mexican Americans, respectively, were available for study. We performed tests of association between AMD and mitochondrial variants using logistic regression assuming a dominant genetic model unadjusted and adjusted for age, sex, body mass index, and smoking status (current vs. none). Tests were performed for individual SNPs and by haplogroups J, T, and U. In non-Hispanic whites at a liberal significance threshold $p<0.05$, two variants were associated with AMD: rs28357682 (MT-CYTB) (OR= 2.17, 95% CI= 1.05-4.52; padjusted=0.04) and mt16362 (OR= 1.81, 95% CI=1.10-3.00; padjusted=0.02). In our Mexican Americans, mt16111 was associated with AMD (OR=2.90, 95% CI=1.17-7.18; padjusted=0.02) and mt16362 (OR=2.48, 95% CI=1.01-6.15; padjusted=0.05). After adjustment for covariates, no mitochondrial variant or haplogroup was associated in non-Hispanic blacks, and no association generalized across all three race/ethnicities. The lack of associations in non-Hispanic blacks reflects the relatively low sample sizes and subsequent lower power to detect associations for this disease in African-descent populations, highlighting the need for further genetic association studies in diverse populations.

802T

Fine mapping of the Glucose-6-Phosphate Dehydrogenase (G6PD) gene in a multicentre study of severe malaria provides evidence for balancing selection. *G.M. Clarke, K.A. Rockett on behalf of The MalariaGEN Consortium.* Wellcome Trust Centre for Human Genetics, University of Oxford, UK.

G6PD deficiency is the most common human enzyme defect affecting more than 400 million people worldwide. Approximately 140 different molecular variants have been identified in the G6PD gene making it one of the most polymorphic loci in the human genome. The specific variants causing G6PD deficiency are localized to specific geographic regions. The most common deficiency haplotype in sub-Saharan Africa is composed of a derived allele at two non-synonymous SNPs at coding positions 202 and 376. Individuals carrying this haplotype exhibit around 12% of normal enzyme; individuals carrying a derived allele only at 376 exhibit around 85% of normal enzyme activity. G6PD deficiency is associated with protection from malaria infection by the plasmodium falciparum parasite, although reports are conflicting as to whether the effect is in males, females, both or neither. Derived alleles at 202 and 376 are not present in Vietnam and Papua New Guinea and are at a lower frequency in Gambia. At these sites and others there are additional variants that play a role in G6PD enzymopathy. To further investigate the association between G6PD deficiency and severe malaria, we genotyped 17,900 controls and 12,131 severe malaria cases from ten African and two Asian sites at 143 SNP variants spanning the G6PD locus. Variants typed included 202 and 376 as well as another 13 with mutations classed as causing G6PD deficiency by the World Health Organization (WHO). The strongest association with severe malaria was at 202 where we found significant evidence of an increased risk of severe malarial anaemia (SMA) in males with a similar trend in females and a significant reduction in the risk of cerebral malaria (CM) in males with a similar trend in females. To investigate the net effect of the WHO classified G6PD variants, many of which are extremely rare and/or exert only modest or weak effects of G6PD deficiency, we developed a surrogate score of G6PD deficiency for each individual. Increases in the G6PD deficiency score were also associated with an increased risk of SMA and a decreased risk of CM suggesting that the effects on severe malaria seen at 202 are indicative of the effect of G6PD deficiency per se. We present detailed results of our analyses and an investigation of the evidence that the contrasting effects of G6PD deficiency on CM And SMA may plausibly lead to balancing selection through a complex interaction of environmental heterogeneity and genetic factors.

803F

Bayesian conditional analysis of 180,000 individuals for migraine reveals local genetic substructure and nine additional genes.

V. Anttila[1,2,3], H. Huang[1,2], N. Eriksson[4], B.M. Neale[1,2], M.J. Daly[1,2], A. Palotie[2,3,5]. 1) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) 23andMe, Mountain View, CA, USA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK.

Migraine is a common, episodic neurovascular disorder with significant heritability. Recently, our Consortium conducted a GWAS meta-analysis of 23,000 migraineurs and 95,000 controls, uncovering genome-wide significant association at 12 loci (Anttila et al, Nature Genetics, in press). In addition, we identified 134 additional loci with genome-wide suggestive p-values (less than 1 × 10-5). We set out to apply a new Bayesian analysis framework to the dataset, allowing for a rapid testing of SNPs conditional on local genotypes from summary-level data. Additionally, we wanted to identify new genetic loci for migraine, using existing data and new samples from 23andme, for a total of 34,000 migraineurs and 147,000 controls. SNP data from the 146 implicated loci were analyzed in a meta-analysis of the previous cohorts and the new samples. In the conditional analysis, six new significant genes (TMEM51, MBOAT4, SUV39H2, PRKG1, HPSE2, TMEM91) were identified. All genes are from loci with p-values <1 × 10-6 in the single-marker analyses. In addition, multiple signals were identified at four previously reported loci (PRDM16, MEF2D, TRPM8, LRP1), where accounting for the multiple signals increases the association p-values by several orders of magnitude (e.g. LRP1 p-value decreases from 3.94 × 10-19 to 1.41 × 10-26). In the expanded meta-analysis including the additional samples, four new loci showed significant association to migraine (FGF6, ASTN2 [distinct locus from previously reported], PRKG1 [significant also in the conditional analysis], MRVI1). At the MRVI1 locus, significant association is observed to a missense variant for the first time in migraine. In addition, MRVI1 interacts directly with PRKG1 in regulating many cellular functions; PRKG1 is a master regulator of many pathways involving in nociception and learning as well as modulating intracellular calcium in neuronal cells, an emerging mechanism in neuropsychiatric diseases. We report the analysis of common variants influencing the genetic susceptibility to migraine, including nine new loci, as well as multiple local independent associations underlying previously reported loci. The newly identified loci include the first common missense variant in common forms of migraine. The implication of the genes at these loci, especially the PRKG1/MRVI1 pair and their biological pathways and mechanisms, make these findings a potentially interesting addition to our knowledge of the genetics of migraine.

804W

Evidence for shared genetic loci between body mass index and menarche timing among 5,357 Hispanic/Latina women in the Population Architecture using Genomics and Epidemiology Study.

L. Fernández-Rhodes[1], C. Carty[2], C. Kooperberg[2], U. Peters[2], T. Matise[3], J.L. Ambite[4], E. Demerath[5], J. Dreyfus[5], M. Gross[5], N. Pankratz[5], E. Ramos[6], L. Hindorff[6], M. Daviglus[7], R. Kaplan[8], S. Castañeda[9], G. Heiss[1], U. Lim[10], C. Haiman[11], L. Le Marchand[10], K.E. North[1], N. Franceschini[1]. 1) University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Rutgers University, Piscataway, NJ; 4) Information Sciences Institute, Marina del Rey, CA; 5) University of Minnesota, Minneapolis, MN; 6) National Human Genome Research Institute, Bethesda, MA; 7) Northwestern University, Chicago, IL; 8) Albert Einstein College of Medicine, Bronx, NY; 9) San Diego State University, San Diego, CA; 10) University of Hawaii, Honolulu, HI; 11) University of Southern California, Los Angeles, CA.

Prior studies have identified several genetic loci for both body mass index (BMI), a measure of obesity, and age of menarche, a marker of puberty timing in population studies. The epidemiologic literature supports an inverse association among these traits, wherein childhood obesity associates with early menarche, which in turn associates with increased BMI and obesity in adulthood. We recently reported the association of several BMI-related single nucleotide polymorphisms (SNPs) with menarche timing among women of European ancestry. We found that 58 of 64 BMI SNPs were associated with menarche in the expected inverse direction and several of the associations had not been previously reported with menarche timing. However, the association of BMI loci with age at menarche has not been investigated among Hispanic/Latina (H/L) women, a diverse ethnic group with disparities in both obesity and early pubertal development. We examined the associations of SNPs in 25 validated BMI loci with fine mapping on the MetaboChip array with self-reported age at menarche among 5,357 H/L women, primarily of Mexican descent, in two studies (Multiethnic Cohort Study and Women's Health Initiative, reported at an average of 69 and 60 years respectively). Due to differences in linkage disequilibrium (LD) between European and H/L populations, we interrogated all SNPs within 500kb of the published European SNP at each BMI locus. Each study performed a linear regression analysis for the association of SNPs with menarche timing (additive genetic model, adjusted for birth year and ancestry); these results were combined in a fixed-effect meta-analysis. SNPs in 9 of 25 BMI loci were significantly associated with menarche when accounting for multiple testing (*NEGR1, TNNI3K, CDKAL1, TFAP2B, MAP2K5, SH2B1, FTO, MC4R, QPCTL*; p<0.002). Three of these loci also showed inverse effects for BMI as compared to menarche (*NEGR1, MC4R, QPCTL*; prior information on BMI effects was missing for the remaining loci). At six loci the significant SNPs were in low LD (r^2<0.2 HapMap CEU) with the published SNP(s) suggesting allelic heterogeneity in these regions. Two BMI SNPs at *KCTD15* also had inverse effects on BMI and menarche (p<0.02). Our study provides evidence supporting the association of BMI loci with age at menarche in H/L women, a diverse admixed population, and shows the contribution of allelic heterogeneity at these loci.

805T

A polymorphism of *Interleukine-13 (IL13)* is associated with susceptibility to food allergy in the Japanese population. *T. Hirota[1], M. Tamari[1], M. Kubo[2], S. Sato[3], M. Ebisawa[3], T. Imai[4], M. Sakashita[5], S. Fujieda[5].* 1) Laboratory for Respiratory and Allergic Diseases, IMS, RIKEN YokohaLaboratory for Respiratory and Allergic Diseases, IMS, RIKEN, Yokohama, Japanma, Yokohama, Kanagawa, Japan; 2) Laboratory for Genotyping Development, IMS, RIKEN, Yokohama, Japan; 3) Clinical Research Center for Allergy and Rheumatology, Sagamihara National Hospital, National Hospital Organization, Sagamihara, Japan; 4) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 5) Department of Otorhinolaryngology-Head & Neck Surgery, Faculty of Medicine, University of Fukui, Fukui, Japan.

Background: The prevalence of food allergy (FA) has increased over the past two decades in Japan. FA is an adverse health effect arising from a specific immune response that occurs reproducibly on exposure to a given food. Recent GWASs have shown that *IL13* locus on chromosome 5q31 is associated with bronchial asthma and atopic dermatitis. But influences of genetic variations in the *IL13* gene on susceptibility to FA in the Japanese population are unclear. Objective: To investigate the association between polymorphisms of *IL13* gene and FA and FA related phenotypes in the Japanese population. **Methods:** We resequenced the *IL13* region by PCR-directed sequencing. Four tag SNPs were selected using the Tagger algorithm and genotyped by Taqman and Invader methods. We performed an association study of FA using a total of 603 subjects with FA and 938 controls in the Japanese population. Among the FA cases, a total of 182 subjects have bronchial asthma. **Results:** We identified a total of 17 polymorphisms including a non-synonymous substitution (Arg144Gln). We found that rs1295686 was associated with FA under an allelic model ($P = 0.000047$; OR, 1.37; 95% CI, 1.18-1.60). In further analyses of patient subgroups, we observed a strong association between rs1295686 and FA with bronchial asthma ($P = 0.0000081$; OR, 1.68; 95% CI, 1.34-2.12). **Conclusion:** rs1295686 of *IL13* gene is significantly associated with FA in the Japanese population. Although further genetic and functional analyses are needed, our findings could help elucidate common genetic factors for bronchial asthma, atopic dermatitis, and FA.

806F

No association between the neuropeptide Y gene polymorphisms and smoking habit in Japanese. *M. Isomura, C. Matsuda, T. Nabika.* Functional Pathology, Shimane University, Shimane, Shimane, Japan.

Purpose: Smoking induces various adverse effects such as chronic obstructive pulmonary diseases, lung cancer, as well as heart diseases. Quit smoking has been a large issue in public health to reduce smoking related diseases. Because nicotine in a cigarette produces activities in brain reward system, smoking habit has a strong potential for addiction. Recently, it has been shown that a NPY (neuropeptide Y) participates in brain reward system and genetic polymorphisms in NPY gene associates with smoking habits. To explore relationship between NPY and smoking habits, we have conducted genetic association study. Methods 3,020 individuals attending health check examinations carried out at rural area in Shimane prefecture in Japan. Smoking history and status were self-reported. Mental status of participants was assessed from response to questionnaire. Three SNPs in NPY gene were genotyped by TaqMan method. Concentrations of NPY in serum were measured by using Neuropeptide Y EIA Kit (Phoenix pharmaceuticals). Results: Genotypes of three SNPs in NPY gene were successfully determined. Association between genotypes and smoking habits were examined by comparing allele frequencies of these SNPs according to smoking status (never, or ever smoker). However, no statistical differences were observed in allele or genotype frequencies between these groups ($p=0.095$). Although several reports showed that NPY concentrations in serum vary among genotypes of a SNP, rs16147, no difference was observed between CC genotype (1.52ng/ml) and TT genotype (1.57ng/ml) in our study. Discussion: In Japanese, polymorphisms in NPY gene showed no association with smoking habit, even though several confounding factors were adjusted. Discrepancy of published result and our result may be due to difference in ethnic group. Our results indicated that polymorphisms in NPY gene had little contribution to smoking habits in Japanese.

807W

Use of eQTLs to Unravel the Etiology of Nonsyndromic Oral Clefting. *C. Malcher[1], C. Masotti[1], L.A. Brito[1], C.F.S. Bassi[1], A.G. Morales[1], S.G. Ferreira[1], D.F. Bueno[1], G.S. Kobayashi[1], N. Alonso[2], M.R. Passos-Bueno[1].* 1) Human Genome Research Center, Institute of Biosciences, University of São Paulo, São Paulo, Brazil; 2) Division of Plastic Surgery, Department of Surgery, School of Medicine, University of São Paulo, São Paulo, Brazil.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a complex disorder determined by the interaction between genetic and environmental susceptibility factors. Among the genetic factors, several susceptibility loci have been found, but the underlying molecular mechanisms remains elusive. Considering that variation of gene expression is an important mechanism in complex diseases, we investigated if expression quantitative trait loci (eQTLs) mapped by our group in mesenchymal stem cells derived from orbicular oris muscle (OOMMSC), a tissue affected in the NSCL/P phenotype), are associated with the disease itself. We conducted a structured association test in a sample of 454 NSCL/P patients and 379 controls. Informed consent was obtained from each participant included in this study, as approved by the Research Ethics Committee of the Institute of Biosciences, University of São Paulo, Brazil. We tested 5 putative OOMMSC eQTLs: rs5011163, rs1505443, rs4793213, rs4793229 and rs1242500. These eQTLs regulate or are located within genes related to RNA processing, *FGF* pathway, DNA repair and cell growth and differentiation, and these processes are known to be relevant for NSCL/P susceptibility or palate development. All individuals were genotyped for these five variants with the Taqman method. To correct for population structure, the individuals were also genotyped for a panel of 40 ancestry-informative markers (AIMs). Chi-square analysis showed no evidence of deviation from HWE for all of the variants in our control sample. The mean ancestry contribution of European, African and Amerindian in the case group was 0.6/0.22/0.18, while it was 0.7/0.19/0.11 in the control group. We did not find association for any of the OOMMSC eQTLs tested: rs5011163 ($p=0.057$), rs1505443 ($p=0.664$), rs4793213 ($p=0.6$), rs4793229 ($p=0.9$), rs1242500 ($p=0.4$). Despite the lack of association, our results exclude some potential molecular mechanisms through which these variants could be involved in the etiology of NSCL/P. A genomic approach will be necessary in order to evaluate the role of the eQTLs in NSCL/P susceptibility. Support: FAPESP/CNPq.

808T

ECM Remodelling genes in intracranial aneurysm: a south Indian perspective. *S. Sathyan[1], L. Koshy[1], H. Easwer[2], S. Premkumar[3], J.P Alapatt[3], M. Banerjee[1].* 1) Human Molecular Genetics, Rajiv Gandhi centre for biotechnology, Thiruvananthapuram, KERALA, India; 2) Department of Neurosurgery, Sree Chitra Tirunal Institute for Medical Science and Technology,-Thiruvananthapuram, KERALA, India; 3) Department of Neurosurgery, Calicut Medical College, Calicut,KERALA, India.

Subarachnoid Hemorrhage accounts for 15% of strokes, and occurs at a fairly young age. Intracranial aneurysm (IA) accounts for 85% of Subarachnoid Hemorrhage. In general population 2-3% individuals harbor intracranial aneurysm and the aggregate worldwide incidence of SAH is 10.5 cases per 100,000 person-years. Extracellular matrix remodeling plays an important role in maintaining the structure of intracranial arteries. Elastin (ELN) and collagen(COL3A1 and COL1A2) fibers form the structural framework for ECM providing elasticity and strength to the artery,while proteoglycan versican (VCAN) is known for its adhesive property. Lysyl oxidase's (LOX and LOXL2) is an important class of copper containing enzyme involved in covalently cross linking of these fibrous proteins by formation of aldehydes between lysine residues, which insolubilize these extracellular proteins, thus giving strength to intracranial arteries. Matrix metalloproteinase's (MMP2 and MMP9) is involved in maintaining structural protein concentration by their degrading ability. Therefore, to understand the functional significance of SNP variants of these genes involved in ECM remodeling in intracranial aneurysm patients, we screened SNPs from these genes based on functionality and tagging status. The study populations consist of 200 radiologically confirmed aneurismal cases and 235 ethnically and age matched controls from Dravidian Malayalam speaking population of Kerala. We do observe a significant allelic and genotypic association with the variants of COL1A2 MMP2 and VCAN in South Indian population. However, many of the extracellular matrix remodeling genes were not found to be associated in our population. These observations are in sharp contrast to reports in Caucasians and European ancestry. These conflicting results may be due to sampling strategies resulting in genetic heterogeneity, phenotypic heterogeneity, epigenetic variability, pertaining to complex disorder like intracranial aneurysm.

809F

Common Variants Associated with Normal Tension Glaucoma Are Also Associated with Glaucoma in Exfoliation Syndrome. *D. Wang[1], B. Fan[1], H. Levkovitch-Verbin[2], I. Pasquale[1], J. Haines[3], J. Wiggs[1].* 1) Ophthalmology, Massachusetts Eye & Ear Infirmary, Boston, MA; 2) Goldschleger Eye Institute, Sheba Medical Center, Tel-Hashomer, Israel; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Genome-wide association studies have identified several genomic regions associated with primary open angle glaucoma (POAG), including TMCO1 at 1q24, CAV1/CAV2 at 7q31, CDKN2BAS at 9p21, SIX1/SIX6 at 14q22 and GAS7 at 17p13. Recently we showed that SNPs in the CDKN2BAS region and a novel regulatory region on 8q22 are associated with the normal-tension POAG subgroup with increased susceptibility to optic nerve degeneration suggesting that the CDKN2BAS and 8q22 SNPs primarily influence the development of optic nerve disease in glaucoma. To further investigate this hypothesis we evaluated the association of SNPs located in these genomic regions with exfoliation syndrome-related glaucoma (EG) where elevated intraocular pressure is due to the accumulation of fibrillar aggregates in the ocular fluid drainage pathways. Glaucomatous optic nerve degeneration develops in approximately 50% of exfoliation syndrome (ES) patients. The lead SNP from each genomic region associated with POAG was analyzed in a discovery dataset of 222 unrelated patients with ES (108 with EG) and 344 control subjects recruited from the United States and in a replication dataset of 92 unrelated patients with ES (67 with EG) and 102 control subjects recruited from Israel. The lead SNPs rs284489 at 8q22 and rs1412829 at CDKN2BAS were significantly associated with EG in the US dataset (P = 0.02 and 0.01 respectively) and in the meta-analysis of both datasets (P = 0.008 and 0.02 respectively). No evidence of association was found between the lead SNPs from the other POAG associated regions and EG in either population. There was no evidence of association between these lead SNPs and ES without glaucoma. These results suggest that genetic variants in the 8q22 and 9p21 regions may contribute to optic nerve disease in exfoliation glaucoma, and thus may be risk factors for optic nerve disease related to glaucoma generally.

810W

Alternatively Spliced Isoform Expression of PTPN2 in Rheumatoid Arthritis. *M. Houtman, K. Shchetynsky, L. Padyukov.* Rheumatology Unit, Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden.

Background: Rheumatoid arthritis (RA) is a prototype autoimmune disease with unknown etiology. Despite extensive genetic association studies, the mechanism of disease in relation to putative function of found loci remains unclear. Expression analysis is one of the tools to link findings from genetic epidemiology studies with biological mechanisms of the disease. Variations within the gene encoding the intracellular protein tyrosine phosphatase non-receptor type 2 (PTPN2), which has been implicated in T cell receptor signaling, have recently been associated with a number of chronic inflammatory diseases, including RA. The *PTPN2* gene codes for at least two alternatively spliced isoforms - TC45 and TC48, which differ in their C-termini and localize differentially to the nucleus and cytoplasm. In this study, we investigate mRNA expression of *PTPN2* isoforms in peripheral blood cells of RA patients and healthy controls. Methods: We studied the relative mRNA expression of all *PTPN2* splice forms combined and the specific *PTPN2* splice forms by TaqMan real-time PCR in relation to house-keeping gene *ZNF592* in whole blood samples from 75 RA patients and 77 healthy controls of Caucasian ancestry. The Mann-Whitney U-test was used to evaluate differences and $P < 0.05$ was considered to be significant. Results: There was no significant difference in total *PTPN2* expression in whole blood samples from RA patients and healthy controls. However, the *PTPN2* splice forms were differentially expressed in these samples. We saw an increase in expression for *TC45* (median relative quantity (RQ) = 1.119 (interquartile range (IQR) 0.664); $P = 0.0259$) and *TC48* (median RQ = 1.047 (IQR 0.529); non-significant) in RA patients. This difference was more pronounced when comparing the ratio of splice forms ($P = <0.0001$). Conclusions: Our study shows that *PTPN2* expression of the canonical and the alternatively spliced isoform in whole blood samples of RA patients is different compared to healthy controls, although expression of total *PTPN2* did not show this clearly. It is an example of expression analysis where detection of overall gene expression without considering transcription/translation diversity may generate false negative results. Our results suggest that the structural differences between the two isoforms might affect events in the pathogenesis of RA.

811T

Comparative analysis of eQTL structures in chronic obstructive pulmonary disease (COPD) and Genotype-Tissue Expression (GTEx). *T. Huang, Q. Long, B. Zhang, S. Yoo, J. Zhu, Z. Tu, the GTEx Consortium.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Expression quantitative trait loci (eQTLs) can help people to understand the mechanisms of disease associated SNPs as majority of them are in non-transcribed regions and presumably function through gene expression regulation. GTEx, like many existing eQTL data collections (e.g., seeQTL and Genevar), is based on tissues from relatively healthy individuals. It is unclear that to what extent eQTLs identified in these normal tissues could replicate in disease populations from which GWA studies were performed. Answers to this question are important and will provide guidance on the appropriate usage of GTEx eQTLs for GWAS SNPs interpretation and may shape the future direction for projects like GTEx. To help answer this question, we calculated eQTLs in all the nine tissues from GTEx with sufficient samples, and eQTLs from COPD lung tissues profiled from Lung Genomics Research Consortium (LGRC) project. Raw data from both data sets was processed in strict parallel to ensure fair comparison. The eQTLs from LGRC COPD lung, LGRC control lung, GTEx adipose, artery, heart, lung, muscle, nerve, skin, thyroid and whole blood tissues were compared from multiple angels. Several interesting results are found: first, eQTLs from GTEx lung share a great similarity with eQTLs from LGRC control lung compared to LGRC COPD lung. Second, among the nine GTEx tissues, lung is the one showing highest similarity with LGRC COPD/control lung with respect to eQTL structures. Third, 35.6% of eQTLs linked to COPD GWAS SNPs (or SNPs in high LD, $r^2 > 0.8$) in LGRC COPD tissues are shared in GTEx. This indicates that genetic regulatory structures are partially conserved across disease and normal individuals. Interestingly, the GTEx tissue with most similar eQTL pattern for GWAS COPD associated SNPs with LGRC eQTLs is not lung but muscle. This may reflect certain pathological changes related to respiratory muscular systems in COPD lung tissues. In conclusion, we performed a comprehensive comparison of eQTL structures in disease and normal tissues and report our findings on the conservativeness of eQTL structures across disease/normal tissues, tissue specificities of these eQTLs, and their relevance to GWAS SNPs. These findings shed some new insights to the complex disease systems and may help to better utilize GTEx for disease studies.

812F

Phenotypic analysis of Peptidylarginine deiminase type 4 knock-out mice. *A. Suzuki[1], Y. Kochi[1], H. Shoda[2], K. Fujio[2], R. Yamada[1, 3], K. Yamamoto[1, 2].* 1) IMS, RIKEN, Yokohama, Japan; 2) Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Center for Genomic Medicine, Kyoto University, Kyoto, Japan.

Rheumatoid arthritis (RA) is well-known as an autoimmune disease and is a chronic inflammatory disorder characterized by the destruction of multiple joints. Many genome wide association studies were performed and multiple RA-susceptibility loci and autoimmune-susceptibility loci have been identified. These studies suggested that multiple genes and its functions were related with disease causing and development. Previouly, peptidylariginine deiminase type 4 (PADI4) was identified as a susceptibility gene for RA in a Japanese population by case-control association study (Ref 1). PADI4 is a member of the PADI gene family and converts arginine residue (peptidylarginine) to citrulline residue (peptidylcitrulline). PADI4 is highly expressed in bone marrow, macrophages, neutrophils and monocytes. Peptidylcitrulline is an interesting molecule in RA, because it is an antigen of ACPA and only PADs (translated protein from PADI genes) can provide peptidylcitrullines, via modification of protein substrates. To evaluate the importance of PADI4 gene in the progression of RA, we generated Padi4 -/- DBA1J mice. We used Padi4-/- mice to show that PAD4 is significantly affected to progress of collagen induced arthritis (CIA), well known as an RA model animal. In Padi4-/- CIA mice sera, the concentrations of serum anti-CII IgM, IgG, and levels of inflammatory cytokines decreased significantly rather than in WT CIA mice. Futhermore, the transcription levels of inflammatory cytokines in CD11b positive splenocytes from Padi4 -/- CIA mice are also significantly lower than those from WT CIA mice. As the results, we suggested that Padi4 enhanced collagen-initiated inflammatory responses. 1) Suzuki, A. et al Nat. Genet.34, 395-402 (2003).

813W

Defective PGRN gene expression in two patients with frontotemporal dementia (FTLD) with potential decreased mRNA stability due to exon 6 PGRN deletion G.101349_101355delCTGCTGT associated with familial FTLD. *E. Vitale[1,2], A. Iuliano[2], A. Polverino[2], G. Milan[3], S. Pappata'[4], P. Sorrentino[3], A. Postiglione[5], V. Alesi[8], A. Novelli[7], G. Sorrentino[2,6].* 1) Cybernetics, CNR-National Research Council, Pozzuoli (NA), Napoli, Italy; 2) School of Movement Sciences (DiSIST), University of Naples 'Parthenope', Napoli, Italy; 3) Geriatric Center "Frullone" ASL Napoli 1, Naples, Italy; 4) CNR-National Research Council, Bioimaging and Biostructures, Naples, Italy; 5) Department of Clinical and Experimental Medicine, University of Naples Federico II, Naples, Italy; 6) Neurological Institute for Diagnosis and Treatment Hermitage Capodimonte, Naples, Italy; 7) Mendel Laboratory, IRCCS Casa Sollievo della Sofferenza Hospital, Roma, Italy; 8) Medical Genetics Unit, San Pietro Hospital Fatebenefratelli, Rome, Italy.

Frontotemporal dementia (FTLD) is a neurodegenerative disease with hallmark deficits in social and emotional functions, characterized by genetic and clinic heterogeneity and progressive declines in behavior or language associated with frontal and temporal lobar degeneration. The phenotypic heterogeneity consistently reveals three different clinical manifestations such as a behavioral variant of FTD (bvFTD), primary progressive aphasia (PPA) and FTD overlapping with motor neuron disease (FTD-MND or FTD-ALS). It can also present with parkinsonian syndromes, progressive supranuclear palsy (PSP) and corticobasal syndrome (CBS). The progranulin gene (PGRN) was recently reported to cause tau-negative frontotemporal dementia linked to chromosome 17. Here we report a four-generation Southern Italian family segregating FTLD in four affected family members, two of them still alive. We collected these two cases (one male and one female) along with the unaffected brother. The two phenotypes are clinically heterogeneous for the behavioral variant (bvFTLD) in the males and for a Primary Progressive Aphasia (PPA) in the female. These individuals were analyzed by sequencing PGRN and other genes in search for a mutation. We found a g.101349_101355delCTGCTGT deletion in PGRN exon 6 CDS in the two affected individuals. The g.101349_101355delCTGCTGT deletion has been already described in two sporadic cases as causing a premature stop codon with a frameshift introduction and could be causing mRNA non-sense mediated decay resulting in PGRN protein haploinsufficiency. We ascertained this mutation in about 50 healthy controls matched by age, sex and geographic regions using allele-specific PCR mutation-primers (ARMS) and we found no deletions carried by these individuals. Moreover, a CGH assay performed on the DNA of the two patients showed no additional genomic alterations. Quantitative RT-PCR of the PGRN gene using exon 7-8 forward and reverse primers was performed on mRNA obtained from WBC from the two patients and from controls. We found a defective PGRN gene expression in the two patients leading to a possible decreased mRNA stability. These results suggest that PGRN may be a possible modifier gene that could lead the FTLD phenotype. Further investigations will focus on understanding how PGRN mutations lead to neurodegeneration and, specifically, on development of FTLD in this family.

814T

Biomarkers and Perinatal Risk Factors in Autism Spectrum Disorder. *H. Xu[1,2], H. Wang[1], Q. Dai[1], A-Q. Zhou[1], M-R. Wu[1], X-Y. Wang[1].* 1) Children's Development and Health, Hubei Provincial Maternal and Children's Health Hospital, Wuhan, Hubei, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, China.

Autism spectrum disorder (ASD) has been given more and more attention with the increasing incidence rate worldwide. Despite the complicated behavior observation scales that have been created for the diagnosis of ASD, few studies have focused on the use of biomakers for early diagnosis of ASD, which may bring the identification and prevention of ASD to an earlier age. Here we have investigated the possible use of plasma secreted amyloid precursor protein alpha (sAPP) and brain-derived neurotrophic factor (BDNF) as the peripheral biomarkers in the diagnosis of ASD from more than 150 children with ASD. A sensitive ELISA method was used to detect plasma APP and BDNF levels. There were significantly increased levels of total sAPP in 67.5% ,sAPP- α in 57.4%, sAPP- β in 62.8% , BDNF in 51.22% of the clinically diagnosed autistic children. The association was statistically significant among the total sAPP (F=59.4965 P =0.0083), sAPP- α (F= 7.5036, P <0.0001), sAPP- β (F=11.7575, P =0.0083) or BDNF (F=10.0685, P = 0.0180) between case groups and control. Correlation analysis of sAPP and BDNF with age results show: the total sAPP, sAPP- α, sAPP- β significantly positively associated with age (β were: -1.90849, -0.80611, -0.75111 respectively; P<0.05 respectively). We have also used a self-designed questionnaire to screen risk factors for ASD and found 11 perinatal factors. We have therefore concluded that prenatal high-risk factors may have certain relationship with ASD; expressions of sAPP, sAPP- α, sAPP- β and BDNF in plasma are of different severity in children with different levels of ASD. It is a potential way for early diagnosis of autism by peripheral biomarkers both in plasma and cord blood, which will promote to build a brand new ASD diagnostic system.

815F

Analysis of DNA-protein interactions using nuclear extracts around the -224 A/G single nucleotide polymorphism in the neuropeptide receptor Y2 (*NPY2R*) gene in predisposition to hypertension. *E. Albino, A. Nuñez, J. Dutil.* Biochemistry, Ponce School of Medicine, Ponce, PR.

Previous work in animal models and human populations identified Neuropeptide Y receptor 2 (*NPY2R*) as a candidate gene for HTN. In two independent Japanese populations, the GG genotype single nucleotide polymorphism (SNP) located 224 bp upstream of the transcription start site was associated with an increased risk of HTN. In addition, the luciferase activity induced by the NPY2R promoter with the G allele in position -224 was reduced by 34% compared to the promoter containing an A in that position. The aim of this study was to assess the DNA-nuclear protein interactions surrounding the *NPY2R*-224 A/G SNP. Two versions of double stranded oligonucleotide probes (50bp) corresponding to the sequence flanking *NPY2R*-224 A/G were synthesized and incubated with different amounts nuclear protein extracts for an electrophoretic shift assay (EMSA). The ENCODE project proposes CTCF as a functional element that binds to the -224 SNP in the *NPY2R* gene. We used CTCF antibody for a gel super shift assay. The DNA/nuclear protein reactions were loaded on a 4% non-denaturing polyacrylamide gel. Interestingly, the EMSA showed an allele-specific binding with the oligonucleotide containing an A nucleotide in -224 position. However, the oligonucleotide containing the G nucleotide in -224 position shows no DNA/nuclear protein interaction. Also, we were able to detect a super shift using the CTCF antibody. The data provides strong evidence for a functional role of *NPY2R* in genetic predisposition to HTN.

816W

Experimental Assessment of the Function of Disease-Associated Genetic Variants. *P. An, A. McLenithan, S. Johnson, S. Limou, CA. Winkler.* Basic Science Program, SAIC-Frederick, Inc., Basic Research Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD.

In the genomics era, one major challenge is the identification of causal functional SNPs and genes. Indeed, thousands of SNPs have been associated with various diseases through candidate gene and genome-wide association studies, but, only a few have experimental functional evidence to support their disease causality. This hinders the process of translating genetic association signals to targeted therapy. Our group has identified multiple genetic variants associated with several complex diseases, including HIV/AIDS, hepatitis B and C, and kidney diseases. Here, we employ multifaceted experimental approaches aided by bioinformatics tools to evaluate the functional effects of some of these individual genetic variants. Disease-associated SNPs functionality was assessed by real-time PCR to quantify gene expression in the Epstein-Barr virus-transformed lymphoblastoid cell lines (LCLs) from patients, by disease target cell-based gene reporter assay to test the regulatory potential of SNPs, and by electrophoretic mobility shift assays (EMSA) to estimate the impact of SNPs on DNA transcription factor binding. Several SNPs in *APOBEC3F*, *ZNRD1*, *CD4* and *SERPINA1* of AIDS candidate genes were found to affect transcription factor binding, with patterns often differing in cell types. A *ZNRD* SNP in the 5'-upstream region showed differential transcription factor binding pattern and altered promoter activity. Integration of these experimental data with in silico bioinformatics data provided a new insight into SNPs function relevant to the disease susceptibility. The biological function beyond genetic association may help understand the operating mechanisms of genetic variants in causing various diseases. (Funded by the National Cancer Institute Contract HHSN261200800001E).

817T

Generation of CLEC16A inducible knockout mouse as a novel model to study the pivotal role of NK cells in the pathogenesis of Type 1 Diabetes.
M. Bakay[1], M. Rankin[1], S. Yoeun[1], J. Kushner[2], H. Hakonarson[1]. 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Texas Children's Diabetes and Endocrinology Center, Texas Children's Hospital, Houston, TX.

CLEC16A was recently found to be associated with several autoimmune disorders including Type 1 Diabetes (T1D). The C-type lectin domain family 16, member A (CLEC16A) gene encodes protein with C-type lectin domain structure, which makes it potentially related to the immune response. Although CLEC16A has no known function so far, the discovery of the C-type lectin gene as a diabetes gene, could eventually translate preventive therapy for the disease. We hypothesize that CLEC16A plays an important role in the immune system and that knockout (KO) mouse for CLEC16A will give valuable clues in this respect, particularly in exploring effects of CLEC16A knockdown that cannot be studied ex vivo. We generated CLEC16A mutant mouse by introducing mutation in exon 3 that leads to a frameshift and a premature STOP codon. Homozygotes for the mutation had a phenotype: small body size, weak tail, they may live to adulthood but die prematurely and do not breed. Heterozygotes have a normal phenotype and lifespan and are fertile. Homozygotes were approximately 50% the weight of their unaffected littermates at one month of age. They may live until adulthood, but die by 9 months of age, none of females has reproduced. Intercrossing heterozygotes produced smaller litters and less than the expected number of homozygotes. To overcome this problem and determine the role of CLEC16A in adult mice, we have generated inducible CLEC16A KO mice (Clec16a loxP/loxP Ins-CreERT). By our design, Ozgene (Australia) has generated a conditional KO mouse for CLEC16A, by floxing exon 3. This introduces a frameshift inactivating all downstream exons. We have crossed Ozgene mice with UBC-CreERT mice (inducible cre recombinase driven by the human ubiquitin C promoter). Tamoxifen treatment will inducibly activate Cre transgene which will lead to removal exon 3, flunked by loxP sites, in Clec16a gene. This event is predicted to result in a frameshift mutation, leading to the introduction of novel amino acids and early termination codon. Deletion will be confirmed in islets and other tissues by quantitative RT-PCR and Western Blot.

Since a whole body CLEC16A KO is a new mouse model, we are in a process of characterizing this model and studying if CLEC16A affects glucose metabolism, insulin secretion and tolerance. We also will determine the role of CLEC16A and its functions in the progression of T1D by studying an inducible whole body knockout model with focus on NK cells.

818F

Genetic Polymorphism and Pathophysiology in Patients with Vitiligo.
S. Chettiar[1], K. Mistry[1], P. Agarwal[1], A. Patel[4], D. Jhala[3], D. Umarigar[2], R. Uppala[5]. 1) Department of Biotechnology, SRKI,, surat, India; 2) Department of Dermatalogy, Civil Hospital, Surat; 3) Department of Zoology, Gujarat Unviersity, Ahmedabad; 4) Genexplore Diagnostic Research Laboratory, Navrangpura, Ahmedabad; 5) Department of surgery-Transplant, University of Nebraska Medical center, Omaha, NE, USA.

Vitiligo is an acquired, idiopathic, hypomelanotic, depigmentary disorder characterized by the appearance of white patches resulting from the loss of functional melanocytes and melanin from the skin and mucous. It affects 1-4% of the world population, where as 0.5-2.5% in India with a high prevalence of 8.8% in Gujarat and Rajasthan states. Oxidative stress plays a vital role in etiology of depigmentation in skin by cellular loss, while catalase (CAT) is proven enzymatic defense system, catalyzing break down of hydrogen peroxide. Altered activity of the enzyme and increased stress markers have been reported in vitiligo patients. To investigate CAT gene polymorphism association to vitiligo suseptiblity, we investigated two CAT gene SNP including promoter region rs7943316 (T/C) and exon 9 rs769217 (A/T) with recognition site of BstXI and Hinf I respectively, in 54 vitiligo patients and 45 healthy volunteers. Catalase activity from the serum of affected and normal individual showed a significant difference, thus supporting that oxidative stress might be involved in the pathophygiology of the disorder. The genotype distribution and allele frequency of promoter region were not significantly different between vitiligo patients and healthy controls. But, the exon 9 showed significant correlation between affected and healthy individuals. Although the haplotype of two polymorphisms also showed association with vitiligo. This study suggests possible association between the CAT gene and the vitiligo susceptibility.

819W

Allele-specific regulation of DISC1 expression by miR-135b-5p.
I. Hovatta[1,2], M. Rossi[1,2], H. Kilpinen[1,3], M. Muona[1,2], I. Surakka[1,2], C. Ingle[3], W. Hennah[1,2], S. Ripatti[1,2,3]. 1) University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Wellcome Trust Sanger Institute, Hinxton, UK.

The *Disrupted-in-schizophrenia-1* (*DISC1*) gene is a risk factor for neuropsychiatric phenotypes. MicroRNAs (miRNAs) are important regulators of protein coding genes. The miRNA-mRNA target recognition mechanism is vulnerable to disruption by DNA polymorphisms. We therefore investigated whether polymorphisms in the DISC1 3'UTR affect binding of miRNAs and lead to allele-specific regulation of *DISC1*. We identified four predicted polymorphic miRNA binding sites in the 3'UTR region of *DISC1*. We tested the effect of nine miRNAs on endogenous *DISC1* expression in vitro by over-expressing these miRNAs in 293FT cells and measuring the *DISC1* expression level by qPCR. To determine whether these miRNAs regulate DISC1 levels by targeting the predicted sites, we used a luciferase reporter gene assay. We cloned either the full length *DISC1* 3'UTR or the miRNA binding sites into a luciferase expression vector and co-transfected them into 293FT cells with either the miR-135b-5p or the miR-559 precursor. We also investigated the putative allele-specificity of miR-135b-5p binding. The 293FT cell line is homozygous for the A allele of rs11122396, predicted to create a novel binding site for miR-135b-5p. Therefore, we created constructs with the G allele at rs11122396. Two of the nine miRNAs, miR-135b-5p and miR-559, significantly reduced endogenous *DISC1* mRNA expression: miR-559 by 23.7 % (p=0.009) and miR-135b-5p by 16.2 % (p=0.039). In the luciferase assays expression from the *DISC1* full length 3' UTR construct was reduced 32.1% (p=0.003) by miR-135b-5p, and by 10.3 %, (p=0.03) from the construct expressing the 60 nt miRNA binding site and flanking sequences. In contrast, miR-599 over-expression did not affect the expression of *DISC1*. When investigating the putative allele-specificity of miR-135b-5p binding we observed that miR-135b-5p had no effect on the luciferase activity of either the full length 3' UTR with G allele at rs11122396 (p= 0.49), or on the construct with the ~60 nt miRNA binding site insert (p= 0.18), indicating that miR-135b-5p binding is specific to the derived (A) allele at rs11122396. Thus, the G allele may be functionally related to the DISC1-associated phenotypes by abolishing regulation by miR-135b-5p, leading to elevated DISC1 levels.

820T

CELSR1 mutations are associated with human spina bifida.
Y. Lei[1], H. Zhu[1], W. Yang[3], M. Ross[4], G. Shaw[3], R. Finnell[1,2]. 1) Dell Pediatric Research Institute, The University of Texas at Austin. 1400 Barbara Jordan Blvd, Austin, TX, 78723; 2) Department of Chemistry and Biochemistry, College of Natural Sciences, The University of Texas at Austin, Austin, Texas; 3) Department of Pediatrics, Division of Neonatology, Stanford University School of Medicine, Stanford, CA; 4) Center for Neurogenetics, Brain and Mind Research Institute, Weill Cornell Medical College, New York, NY.

Spina bifida is one of the most common neural tube defects (NTDs), yet its complex etiology remains poorly understood. Recently, mutations in planar cell polarity (PCP) genes have been associated with NTDs, including spina bifida, in both animal models and human cohorts. We studied CELSR1, one of the PCP core genes in a California cohort to determine if CELSR1 mutations increase the risk of spina bifida. We sequenced the coding region of CELSR1 in 192 spina bifida infants. Novel and rare variants were genotyped in a control group of 190 ethnically-matched infants without malformations. Six missense mutations, absent in controls, were predicted to be deleterious by both SIFT and PolyPhen computational algorithms. Two TG dinucleotide repeat variations were detected in two spina bifida infants separately, but were not observed in controls. In functional analyses, the two TG dinucleotide repeat variants altered the subcellular localization of GFP-tagged CELSR1 protein in transfected cells. Both variants impaired interactions with VANGL2, and reduced the ability to recruit VANGL2 to the cell membrane. In total, 4.1% of spina bifida cases possessed deleterious or predicted-to-be-deleterious CELSR1 mutations. Our findings indicated that CELSR1 mutation is a risk factor for spina bifida.

821F

Functional impact of polymorphisms in the *MMP3* and *TIMP2* gene promoters in human nonsyndromic cleft lip/palate. *A. Letra[1,2,3], M. Zhao[1], R.M. Silva[1,2,3], A.R. Vieira[4], J.T. Hecht[2,3].* 1) Department of Endodontics, University of Texas Health Science Center School of Dentistry at Houston, Houston, TXHouston, Houston, TX; 2) Department of Pediatrics, Pediatric Research Center, University of Texas Health Science Center Medical School at Houston, Houston, TX; 3) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 4) Departments of Oral Biology and Pediatric Dentistry, University of Pittsburgh School of Dental Medicine, Pittsburgh, PA.

Matrix metalloproteinases (MMPs) and their inhibitors (TIMPs) are responsible for tissue remodeling during craniofacial development. Evidence from biological and human studies strongly support a role for MMP and TIMP genes as candidate genes for cleft lip/palate. We have previously shown the association of promoter polymorphisms in *MMP3* and *TIMP2* genes with cleft lip/palate. In this study, we assessed whether the previously associated promoter variants in *MMP3* (-709 A/G) and in *TIMP2* (-180 C/T) genes have functional implications and modulate gene transcription by changing the affinity of their gene promoters for transcription factors. Electrophoretic mobility shift assays, mass spectrometry analysis and luciferase reporter gene assays were performed for each gene variant. For the *MMP3* -709 A/G variant, EMSA showed DNA-protein binding complexes specific for the A allele, in which 34/204 proteins identified by mass spectrometry analyses were specific for this allele. Luciferase assays showed that allele A has lower promoter activity than the G allele in the presence of a known functional polymorphism in the *MMP3* promoter (-1171 5A6A). For the *TIMP2* -180 C/T variant, EMSA showed that both C and T alleles present potential binding sites for NF-Kappa B transcription factor, while super shift assays confirmed binding to NF-Kappa B with a higher binding affinity for allele C. Luciferase assays showed a 2.5 fold increase in promoter activity in the presence of allele T when compared to allele C. Taken together, these results show that variations in *MMP3* and *TIMP2* gene promoters may perturb gene transcription and/or function with effects on craniofacial development. Our study provides new evidence implicating *MMP3* and *TIMP2* variants in the occurrence of cleft lip/palate.

822W

Indian Hedgehog (IHH) and LRBA: New genes for Hirschsprung disease. *R.M.W Hofstra[1], Y. Sribudiani[1], R. Chauhan[1], M. Alves[1], C. Kockx[2], T. van Essen[2], R. Brouwer[2], M. van den Hout[2], W. van Ijcken[2], I.T. Shepherd[3], A.S. Brooks[1].* 1) Clinical Genetics, ErasmusMC, Rottedam, Netherlands; 2) Center for Biomics, Erasmus Medical Center, Rotterdam, Netherlands; 3) Department of Biology, Emory University, Atlanta, USA.

Hirschsprung is characterized by the absence of enteric ganglia in a variable length of intestinal tract. A large, multi-generational Dutch family with five affected family members with HSCR, revealed linkage to 4q31.3-q32.3. As the family shows an autosomal dominant mode of inheritance with incomplete penetrance, we assume that the mutation in the linkage region is necessary but not sufficient to cause disease development. To identify the mutation in the linkage region, but also mutations elsewhere in the genome, we exome sequenced two patients. One variant rs140666848 (T > C) in exon 20 of the LRBA gene was found in the linkage region. In addition, missense mutations in RET (P398L) and IHH (Q50K) were identified in each patient, respectively. Functional analysis of the RET en IHH variants showed that both mutations give rise to a non functional protein. The LRBA variant is located downstream of MAB21L2, a gene which plays a role in the proliferation of enteric neural crest cells (ENCCs) during ENS development in Zebrafish. We hypothesize that this variant might regulate MAB21L2 expression in ENS. Luciferase assays showed that the region containing the variant does show enhancer activity. Currently we are testing whether the variant, when compared to wild type sequence, show significant differences on gene expression level, using luciferase assays. Our data shows that combinations of mutations, as expected, cause autosomal dominant disease with reduces penetrance. Furthermore, we show for the first time that IHH mutations can contribute to Hirschsprung disease.

823T

Genes located in type 1 diabetes risk loci are expressed in human islet and interact in functional networks. *C.A. Brorsson, J. Stoerling, F. Pociot.* Glostrup Research Institute, Glostrup University Hospital, Glostrup, Denmark.

Background: Genome-wide association studies of type 1 diabetes (T1D) have identified 50 susceptibility loci. Although most pin-pointed candidate genes have putative functions within the immune system studies in human pancreatic islets have shown expression for 60% of these genes. Assignment of candidate genes is often based on known biological function or location in relation to the most associated SNP in a locus which could introduce bias into functional studies as all genes in a locus potentially are equally good candidates. We investigated the expression and functional interactions of all genes located within T1D loci in human islets. Methods: A list of all 857 genes located within T1D GWAS loci were downloaded from t1dbase.org. Gene expression based on RNA-seq in human islets was determined from a published study by Eizirik et al. Only genes that had an expression level above 1 in all five islet preparations in the control and/or cytokine-stimulated condition were considered expressed. Protein-protein interaction networks of expressed genes were constructed in DAPPLE. Results: Of 857 genes in T1D loci 336 were expressed in the islets representing 44 loci and 33 previously pin-pointed candidate genes. The expressed genes included 8 pseudogenes, three lincRNAs, two miRNAs and two snoRNAs. In 16 of the loci only a single gene was expressed, including nine candidate genes (*STAT4*, *CENPW*, *TNFAIP3*, *SKAP2*, *GLIS3*, *ZMIZ1*, *LMO7*, *RASGRP1* and *PTPN2*). In addition, at 7p12.1 *COBL* was the only protein-coding gene of the two genes expressed. Using the 336 islet-expressed genes as input, DAPPLE detected 123 direct interactions. A network consisting of only direct connections contained 62 genes of which nine were candidate genes (*NCOA1*, *STAT4*, *ZMIZ1*, *ERBB3*, *SH2B3*, *LMO7*, *SMARCE1*, *PTPN2* and *TYK2*). Based on the connectivity of each of the proteins 48 genes were highlighted as significant functional candidates including several HLA genes and five candidate genes (*SH2B3*, *GPR183*, *TNFAIP3*, *NCOA1* and *ERBB3*). Conclusions: These findings provide evidence that many of the genes located in T1D risk loci are expressed at the mRNA level in human islets and interact in networks. This suggests that variation that gives rise to alterations in expression levels of these genes is likely to affect network function in a way that could promote deleterious events leading to T1D. Further studies are needed to elucidate the function of the islet expressed genes in the T1D pathogenesis.

824F

Elucidating the genetics of type 2 diabetes by integrative analysis of the genome, transcriptome and epigenome of skeletal muscle and adipose tissue samples from Finnish individuals spanning the range of glucose tolerance. *J.R. Huyghe[1], S.C. Parker[2], M.R. Erdos[2], H. Koistinen[3], P.S. Chines[2], H.M. Stringham[1], L.J. Scott[1], L. Taylor[2], T. Blackwell[1], H. Jiang[1], C. Ma[1], A.U. Jackson[1], R. Welch[1], N. Narisu[2], A.J. Swift[2], L.L. Bonnycastle[2], M.L. Stitzel[2], L. Kinnunen[3], R.M. Watanabe[4,5], T. Lakka[6], M. Laakso[6], J. Tuomilehto[3], F.S. Collins[1], M. Boehnke[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) National Institute for Health and Wellfare, Helsinki, Finland; 4) Department of Preventive Medicine, University of Southern California (USC) Keck School of Medicine, Los Angeles, CA, USA; 5) Department of Physiology and Biophysics, Keck School of Medicine of USC, Los Angeles, CA, USA; 6) University of Eastern Finland, Kuopio, Finland.

Genome-wide association studies (GWAS) have identified >70 loci associated with type 2 diabetes (T2D) risk, and have significantly advanced our understanding of the pathophysiology of T2D. Yet, for most identified loci, the causal genes and functional variants remain elusive because the associated region resides in noncoding DNA. Despite the incomplete annotation of the noncoding genome, there is overwhelming evidence that GWAS loci for complex traits cluster near transcriptional regulatory elements in disease-relevant tissues. Therefore, a crucial next step in functional investigations of GWAS-identified variants is the examination of their relationship to gene expression in disease-relevant tissues, collected from subjects representing different stages of progression to T2D. To that end, as part of the Finland-United States Investigation of NIDDM Genetics (FUSION) study, we collected biopsies from two major insulin target tissues: skeletal muscle and adipose tissue. Skeletal muscle accounts for ~80% of insulin-responsive glucose uptake. Adipose tissue, while not a major site of glucose disposal, plays a key role in insulin resistance and T2D by mechanisms that remain unclear. We obtained biopsy samples from *vastus lateralis* skeletal muscle and abdominal subcutaneous adipose tissue from 324 clinically well-characterized Finnish individuals spanning the range of glucose tolerance: 125 normal glucose-tolerant, 72 impaired glucose-tolerant, 41 impaired fasting glucose, and 86 newly diagnosed T2D cases without antihyperglycemic medication. As of May 2013, sequencing of high quality (RIN 7.3-9.1) mRNA extracts to a median depth of 46.8 million aligned read pairs has been completed for 132 samples. RNA-seq quality control metrics and analyses of spiked-in RNA controls indicate excellent data quality. Genome-wide genotype and DNA methylation data are being generated using the Illumina Omni 2.5M-Quad SNP chip and Illumina HumanMethylation450 chip. A primary study aim entails examining the relation between expression quantitative trait loci (eQTL) and GWAS-associated SNPs for T2D and related traits. Integration with epigenetic marks will help to nominate putative causal variants and genes for functional follow-up. Further, this rich data resource will enable the study of the diverse molecular processes involved in insulin resistance which is a biomedical priority as most interventions proven to delay T2D onset act to reduce insulin resistance.

825W

Altered expression of *ARAP1*, at a type 2 diabetes GWAS locus, influences insulin secretion from pancreatic beta cells. *J.R. Kulzer, M.P. Fogarty, K.L. Mohlke.* Department of Genetics, UNC Chapel Hill, Chapel Hill, NC.

The biological mechanisms underlying association of the *ARAP1* locus with type 2 diabetes and proinsulin levels are unknown. The risk allele of index SNP rs11603334 increases transcriptional activity at an *ARAP1* promoter and is associated with increased levels of *ARAP1* mRNA expression in human islets. However, the functional effects of altered *ARAP1* expression on insulin processing and secretion are unknown. ARAP1 is a GTPase activating protein (GAP) that catalyzes hydrolysis of GTP bound to Arf and Rho GTPases, rendering the GTPases inactive. Arf and Rho GTPases regulate Golgi transport, membrane trafficking, and actin cytoskeleton dynamics, processes that are important to insulin processing and secretion in the beta cell. We hypothesized that ARAP1 acts through one or more Arf or Rho GTPases to exert a regulatory role on insulin processing and/or secretion. In cell culture or *in vitro*, ARAP1 has been shown to hydrolyze GTP-bound Arf1, Arf5, Arf6, RhoA, and Cdc42. The specific Arf and Rho GTPases regulated by ARAP1 in beta cells are unknown. To investigate the effects of altered ARAP1 expression on insulin processing and secretion, we knocked-down and overexpressed ARAP1 in 832/13 and MIN6 beta cell lines and measured levels of secreted insulin. Preliminary results show a 57% decrease in glucose-stimulated insulin secretion (P = .026) upon 55% siRNA knockdown of *ARAP1*. Transient overexpression of wild-type ARAP1 resulted in a 70% increase in KCl-stimulated insulin secretion (P = .028). To determine which Arf and Rho GTPases are regulated by ARAP1 in beta cells, we used GTP pull-down assays to measure the levels of GTP-bound Arf1 remaining after ARAP1 overexpression. Exogenous wild-type ARAP1 decreased Arf1-GTP levels by 47%. Exogenous ARAP1 variants with either a catalytically inactive Arf GAP domain or an inactive Rho GAP domain decreased Arf1-GTP levels by ~30%. Overexpression of an ARAP1 variant with both catalytically inactive GAP domains did not decrease Arf1-GTP levels, suggesting that both domains play a role in catalyzing Arf1-GTP hydrolysis. Arf1-GTP regulates transport between the endoplasmic reticulum and the Golgi apparatus, a critical step to the efficient packaging of secretory hormones. Ongoing work assesses the effects of ARAP1 overexpression on Arf5, Arf6, Cdc42, and RhoA. Our results suggest that altered expression of ARAP1 may affect Arf1 activity and interfere with proper regulation of insulin trafficking and secretion.

826T

The widely presumed type 2 diabetes causal variant, rs7903146, within *TCF7L2* binds a specific protein complex not seen with its closest non-causal proxy. *Q. Xia[1], S. Deliard[1], C.X. Yuan[2], M.E. Johnson[1], S.F.A. Grant[1,3,4].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Proteomics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

There have been many efforts to resolve the underlying causative mechanism for a given GWAS signal. However, in the vast majority of situations, there still remains only a list of candidate variants that could represent the causal event. The situation is somewhat more advanced for the strongest associated type 2 diabetes (T2D) locus reported to date, *TCF7L2*, where follow-up analyses in multiple ethnicities have strongly implicated rs7903146 as the causal variant within intron 3. As such, we carried out oligo pull-down combined with mass spectrophotometry (MS) to elucidate the transcriptional machinery across the SNP. Nuclear lysates from HCT116 cells, where TCF7L2 is abundantly expressed, were incubated with biotin-labeled, double-stranded 60bp oligonucleotides spanning rs7903146. The DNA-protein complexes were precipitated with streptavidin-agarose beads, and the bound proteins were isolated by denaturing SDS-PAGE. One major band was observed with the rs7903146-specific oligo that was absent with the scrambled oligo. Furthermore, an oligo coinciding with a SNP in strong linkage disequilibrium with rs7903146 in Caucasians but not in other ethnicities and widely rejected as the causal variant, namely the proxy rs12255372, also did not yield this extra band. Following digestion with trypsin, the samples were analyzed by MS. We observed that poly (ADP-ribose) polymerase 1 (PARP-1) is by far the most abundant binding factor. Among the next most abundant binding proteins, a number were shown to dimerize with the TCF7L2 protein, which resonates with our previous TCF7L2 ChIP-seq work that characterized 4 sites within the *TCF7L2* gene itself. To test for evidence of a feedback mechanism for TCF7L2 operating at this location, we cloned the genetic elements in both orientations upstream of a luciferase reporter gene. The combination of TCF7L2 and β-catenin over-expression produced a striking increase of up to 5 fold in transcription levels in 293T cells; in addition, we observed differences between alleles for rs7903146. We also found evidence for an allelic difference in the MS results for proteins with less abundant binding, namely X-ray repair cross-complementing 5 (XRCC5) and RPA/p70. Our results point to a protein complex binding across rs7903146 within *TCF7L2* and reveals a possible mechanism by which this locus confers T2D risk. Furthermore, we implicate PARP-1 as playing a role in T2D pathogenesis, a target which has classically been pursued for cancer.

827F

A role for IRF6 in cell cycle regulation and DNA damage response in mesenchymal stem cells. *G.S. Kobayashi[1], L. Alvizi[1], B.V.P. Almada[1], L.C. Andrade-Lima[2], C.F. Menck[2], M.R. Passos-Bueno[1].* 1) Human Genome Research Center, Institute for Biosciences, University of São Paulo, Brazil; 2) Institute of Biomedical Sciences, University of São Paulo, Brazil.

Purpose: The transcription factor IRF6, which regulates orofacial and epidermal development, is associated with both syndromic and non-syndromic cases of cleft lip/palate (CL/P). The most prevalent syndromic form of CL/P is van der Woude syndrome (VWS), which is thought to be caused by haploinsufficiency of IRF6. Despite the established role of IRF6 in regulating differentiation and cell cycle progression in some tissues of epithelial origin, little is known about its functions in non-epithelial cells. We have recently reported that mesenchymal stem cells from non-syndromic CL/P patients exhibit impairment of DNA repair mechanisms, which are known to be tightly related to cell cycle regulation (Kobayashi & Alvizi ***et a***l., 2013). Therefore, our objective was to verify if IRF6 participates in these cellular processes in mesenchymal stem cells, in order to further clarify its biological role in non-epithelial cells and in the aetiology of CL/P itself. **Methods:** We quantified DNA double-strand break (DSB) formation using flow cytometry for anti-γH2AX at 6 and 24 hours after exposure to H_2O_2 (100μM), using 2 VWS and 2 control *orbicularis oris* muscle-derived stem cell (OOMDSC) cultures. The proliferative profile of the cell cultures was assessed using an XTT-based assay. Real-time quantitative PCR (RT-qPCR) was carried out in 3 independent experiments to measure IRF6 mRNA expression in untreated and G_0/G_1-synchronised OOMDSCs (serum-starved for 48 hours and contact-inhibited under confluence). Synchronisation was confirmed by flow cytometry analysis using propidium iodide and RT-qPCR for proliferation marker *MKI67*. **Results/Conclusion:** VWS OOMDSCs were more sensitive to the H_2O_2 treatment, exhibiting a clear accumulation of DSBs at 6 and 24 hours of exposure as compared to controls. We did not observe consistent differences between the proliferative potential of VWS and control OOMDSCs; however, in both cases, *IRF6* was markedly up-regulated in contact-inhibited but not serum-starved quiescent cells, suggesting that this gene also plays a role in cell cycle regulation in mesenchymal stem cells. Given these results, we hypothesise that, in this cell type, IRF6 may participate in a cell cycle-integrated response to oxidatively-generated DNA damage. Further investigation will be necessary in order to confirm these results and clarify their relationship with the aetiology of CL/P. Financial support: CEPID/FAPESP, CNPq, MCT.

828W

Novel regulatory variants in family studies with adult chronic atopic dermatitis. *W.B. Jones[1], T.C Pansuriya[1], K.J Gulewicz[1], I. Oh[1], B. Borgo[1], A. Shemer[2], J.G Krueger[3], E. Guttman-Yassky[3], C. de Guzman Strong[1].* 1) Dermatology, Washington University in St.Louis, St.Louis, MO; 2) Tel-Aviv University, Israel; 3) Icahn School of Medicine at Mt. Sinia New York, USA.

Atopic Dermatitis (AD) is a complex, chronic inflammatory skin disease with a high degree of heritability. Although up to 50% of moderate to severe cases of AD is attributed to mutations in the skin barrier gene filaggrin (FLG), other genomic regions (1q, 5q, 11q, 19p, and 20q) are implicated as well. We pursued genetic studies in a tight phenotypic adult cohort for AD to better resolve the genetic architecture of AD. An adult cohort representing chronic AD demonstrated baseline skin barrier impairment owing to down-regulation of Epidermal Differentiation Complex (EDC) genes in the unaffected skin areas. Only 2 out of 31 (6%) patients in our adult cohort were heterozygous for the two most common FLG mutations, R501X and 2282del4. Although fewer FLG copy number repeats were present in AD patients vs. controls (Wilcoxon rank sum test, p=0.03), we hypothesized additional variants in regulatory elements in linkage disequilibrium (LD) with FLG (rs877776, chr 1q21) or GWAS-tagged SNP (rs7927894, chr 11q13.5) to explain the persistence of AD in this adult cohort. SNP discovery using targeted next-generation sequencing of ENCODE-annotated enhancers surrounding rs877776 and rs7927894 in our pooled AD cohort identified a total of 18 variants. Two of the variants were novel (absent in 1000 Genomes and dbSNP NCBI database) in 2 independent patients. These novel variants correlated to previously described conserved noncoding elements (184 and 180) that demonstrated enhancer and repressor activities, respectively, and are bioinformatically predicted to disrupt NFE2L2 and Sox9 transcription factor binding. The function for each of the variants was assessed using cell-based reporter assays . No appreciable difference was observed for the novel variant in CNE 180. However, the rare variant in CNE 184 exhibited a significant and surprising increase in enhancer activity. Moreover, chromosomal conformation capture (3C) assays provide compelling evidence for genomic interactions between these regulatory elements and the dysregulated EDC genes. Our results present a framework for discovering variants in regulatory elements in LD with GWAS-tagging SNPs for a complex disease. More importantly, we identify a functional novel regulatory variant nearby clinically-relevant genes that could contribute to the persistence of FLG-deficient AD into adulthood.

829T

Meta-analysis on eQTL mapping identify insertion and deletion (INDEL) specific eQTLs in LCL, PBMC and skin tissues. *J. HUANG[1], J. CHEN[2], J. Esparza[3], J. Ding[4], J. Elder[5], Y. Lee[3], M. Moffatt[6], W. Cookson[6], L. Liang[1,2].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Max-Delbrück-Center for Molecular Medicine, Berlin, German; 4) Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 5) Department of Dermatology, University of Michigan Medical School, Ann Arbor, MI 48109-0932, USA; 6) National Heart and Lung Institute, Imperial College London, London, UK.

Large scale genome-wide association studies (GWAS) for gene expression quantitative trait loci (eQTL) mapping have been primarily focused on single nucleotide polymorphisms and have helped interpret findings from GWAS for a variety of complex diseases and traits. The functional effect of structure variants, especially short insertions and deletions (INDEL) have not been systematically investigated due to limitation in the availability of high throughput techniques. In this study, we imputed 1,380,133 INDELs based on the latest 1000Genomes project panel into 3 eQTL datasets from multiple tissues including lymphoblastoid cell lines (LCLs), peripheral blood mononuclear cell (PBMC) and skin. An eQTL Meta analysis combining 741 samples from these datasets identified INDEL specific eQTLs for 325 genes (FDR<5%), which cannot be mapped to any SNP at genome-wide significant level even after imputation. Among these significant INDEL eQTL pairs, 3,232 (1.9%) show significant tissue specific effect, compared to 9,775 (1.5%) for significant SNP eQTLs pairs. We also show that INDEL specific eQTLs were enriched in Gene Ontology categories apoptosis, cell adhesion and cell cycle. Finally, imputation of 1000Genomes SNP, INDEL and cross-tissue meta-analysis together identify cis eQTLs for 6,228 genes. All eQTL results are available on our newly developed web browser. This study provides new insights into the underlying genetic architecture of gene expression, and furthermore, we illustrate how this newly developed database for SNP and INDEL specific eQTLs based on multiple tissues could help explain the association to available disease phenotypes, including asthma, atopic dermatitis and psoriasis.

830F

Integrative Genomics Approach to Unravel the Molecular Mechanisms Underlying Genome-Wide Association Results For Lung Function Measures. *M. Obeidat[1], K. Hao[2], Y. Bossé[3,4], D. Nickle[5], D. Postma[6], M. Laviolette[4], A. Sandford[1], D. Daley[1], C. Brandsma[7], M. Berge[6], R. Vessey[8], G. Opitec[8], W. Timens[7], D. Sin[1], P. Paré[1].* 1) UBC James Hogg Research Centre, Vancouver, BC, Canada; 2) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, USA; 3) Department of Molecular Medicine, Laval University, Québec, Canada; 4) Institut universitaire de cardiologie et de pneumologie de Québec, Laval University, Québec, Canada; 5) Merck Research Laboratories, Boston, MA, USA; 6) Department of Pulmonology, University of Groningen, University Medical Center, Groningen, GRIAC research institute, Groningen, The Netherlands; 7) Department of Pathology and Medical Biology, University of Groningen, Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, The Netherlands; 8) Merck & Co. Inc., Rahway, NJ, USA.

Background: SpiroMeta and CHARGE consortia published a large scale meta-analysis of lung function measures (n~49,000), which identified 26 novel loci (Artigas et al. Nat. Gen. 2011). However, the exact mechanisms underlying these associations are not fully understood. Hypothesis: A subset of SNPs which influence lung function act as eQTL in the lung to change the level of expression of their gene product. Aim: To identify SNPs that are associated with both lung function measures and mRNA levels (eQTL) in lungs. Methods: The lung eQTL were derived from genome-wide genotyping and gene expression analysis of 1,111 lung tissue samples. The study identified ~470,000 SNPs related to the level of gene expression in cis and ~17,000 SNPs in trans, at 0.1 FDR (Ke Hao et al. PLoS Gen. 2012) The SpiroMeta and CHARGE consortia have made available the top 2000 SNPs associated with forced expiratory volume in one second (FEV1), and its ratio to Forced Vital Capacity (FEV1/FVC). We undertook an integration of these 4000 SNPs with the lung specific eQTL at a 0.1 FDR. Results: From the CHARGE study associations with FEV1, 1293 SNPs were associated with cis eQTL and 976 trans eQTL, and for the FEV1/FVC associated variants, 809 were associated with cis eQTL and 236 with trans eQTL. For SpiroMeta associations' results for FEV1, 594 SNPs were associated with cis eQTL and 54 with trans eQTL, and among the FEV1/FVC associated variants, 578 SNPs were associated with cis eQTL and 91 SNPs with trans eQTL. 48 eQTL overlapped between SpiroMeta and CHARGE for FEV1 and an equal number (48 eQTLs) for FEV1/FVC, and 7 eQTL were common to both consortia associations' with FEV1 and FEV1/FVC. The mRNA levels of a number of genes regulated by the integrated SNPs also correlate with lung function measures in the lung eQTL study. Gene Ontology and pathway enrichment analyses showed lung function genes to be involved in inflammatory, tissue remodeling and developmental pathways. Future work will focus on prioritizing a number of genes for validation at protein level. Conclusion: A large number of lung function associated SNPs act as lung specific eQTL for either FEV1 or FEV1/FVC in the two consortia studied, Additionally 7 eQTL were overlapping in the two cohorts and affected both FEV1 and FEV1/FVC. The data suggest that a number of the identified variants influence lung function by modulating gene expression levels in lung.

831W

Identifying the molecular mechanisms at the vascular *PLXND1* locus associated with human waist-hip ratio. *T.S. Roman[1], J.E.N. Minchin[2], J.F. Rawls[2], K.L. Mohlke[1].* 1) Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Molecular Genetics and Microbiology, Duke University, Durham, NC.

Genome-wide association studies for waist-hip ratio (WHR), a measure of body fat distribution, have identified numerous loci including several located near genes related to angiogenesis. Stimulation of angiogenesis alters adipose metabolism, and inhibition of angiogenesis abrogates adipose tissue expansion. One locus associated with WHR by the GIANT consortium is *PLXND1* ($P = 4\times10^{-6}$, Heid 2010). *PLXND1* encodes Plexin D1, a transmembrane co-receptor expressed in endothelial cells that plays a role in physiological and pathological angiogenesis. *Plxnd1*$^{-/-}$ null mice have vascular defects and die shortly after birth, and *plxnd1*$^{-/-}$ null zebrafish are viable but also have vascular patterning defects. Human adipose tissue biopsies showed a positive correlation between *PLXND1* mRNA levels and a hypertrophic adipose morphology in visceral but not subcutaneous adipose, suggesting adipose-specific effects of Plexin D1 signaling. The DNA elements regulating *PLXND1* expression and the identity of the underlying functional GWAS variants are unknown. We hypothesize that WHR-associated DNA variants are located within *cis*-regulatory DNA elements that control tissue-specific expression of *PLXND1* and regulate fat deposition and morphology. All 37 WHR-associated DNA variants at the *PLXND1* locus are non-coding (r^2>.8, 1000 Genomes Phase 1 CEU), and all overlap with open chromatin or histone modifications that mark gene regulatory regions in human umbilical vein endothelial cells (HUVEC). We defined and tested *cis*-regulatory elements for effects on *PLXND1* expression using zebrafish transgenesis reporter assays and transcriptional reporter assays in mammalian cells. Candidate *cis*-regulatory elements recombined into a GFP expression vector were injected into one-cell stage *Tg(flk1:mCherry)* zebrafish embryos to test for regulatory activity. This zebrafish reporter assay enables identification of temporal-specific or cell-type specific regulatory elements. The same candidate *cis*-regulatory elements were tested in transcriptional reporter assays transfected into HUVEC. Preliminary data suggests that at least 3 regulatory elements show greater than 1.5-fold enhancer activity relative to an empty vector control in HUVEC. Identification of the *cis*-regulatory elements at *PLXND1* will provide greater insight into the molecular mechanisms and cell types important for regulating *PLXND1* expression, and lead to a greater understanding of how *PLXND1* influences WHR and fat deposition in humans.

832T

Genome-wide enrichments for regulatory regions across thousands of unlinked disease-associated variants. *A.K. Sarkar[1,2], L. Ward[1,2], M. Kellis[1,2].* 1) Massachusetts Institute of Technology, Cambridge, MA., USA; 2) The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Genome-wide association studies have identified thousands of non-coding genetic variants associated with disease, but many additional variants lie below the stringent p-value threshold given current cohort sizes. Here, we investigate both top-scoring and weakly-associated variants using functional genomics annotations of regulatory regions in 100 human tissues and cell lines from ENCODE and the Epigenomics Roadmap project. We find top-scoring non-coding variants are highly enriched for regulatory annotations of relevant cell lines, suggesting regulatory variation may be contributing to the molecular basis of the disease phenotype, and guiding the search for relevant cell types for QTL studies in a number of cases. We also find some surprising enrichments, suggesting new and unexpected tissue and cell types may be playing roles in previously unsuspected diseases, which can help guide directed experiments to reveal potentially novel pathways, tissues, and cell types. We also tested the enrichment of non-coding variants below the genome-wide significance threshold. Surprisingly, we found that an unprecedented number of variants are enriched for predicted enhancer regions in matching cell lines, spanning thousands of independent loci, even after excluding regions containing top-scoring variants. Our results suggest that the architecture of some human diseases is dramatically more complex than previously recognized. In particular, we find that variants associate with Type 1 Diabetes and Rheumatoid Arthritis are strongly enriched for enhancer regions active in immune cell lines and showing evidence of H3K4me1, a histone modification mark associated with active and poised enhancer regions, and H3K27ac, associated with active regulatory regions. These also show evidence of DNase Hypersensitivity in T-cells and B-cells, and are precisely enriched within high-resolution footprints and dips in the chromatin signal associated with nucleosome displacement characteristic of transcription factor binding. These results are robust to diverse permutation corrections, enrichment and association statistics, LD pruning strategies, and null distributions. We find more than 1000 distinct regions after LD pruning and excluding the MHC, and include both common and unique variants between the two traits. Overall, our results suggest a dramatically large number of additive interactions of variants with individually small effect sizes may underlie numerous human traits.

833F

Enabling phenomics with high-throughput whole-organism 3D pan-cellular imaging at cell resolution. *K. Cheng[1], P. La Riviere[2].* 1) Jake Gittlen CA Res Inst ,H059, Penn State College of Medicine, Hershey, PA; 2) Dept of Radiology, University of Chicago, Chicago, IL.

Genes and environment can impact any cell type at any developmental stage. Ideally, the comprehensive study of those influences in a multicellular organism therefore requires the ability to visualize, identify, and characterize every cell type in the whole organism across all life stages at cell resolution. The most common medical implementation of cell imaging, histology, is achieved using tissue sections of about 5 micron thickness, differential color staining of cell components, and opticical resolutions in the range of 1 micron. The destructive and tedious nature of histology precludes the study hundreds of thousands of samples, as would be required to study the functions or effects of tens of thousands of genes or chemicals. A collaborative team is working towards the ideal of high-throughput tomographic imaging and characterization of all cell types in whole small model organisms and millimeter scale tissue samples in a way that is immune to tissue opacity and pigmentation, based on the use of X-rays. Pan-cellular stains, fields of view, and phase contrast through use of monochromatic X-rays are possible at the Advanced Photon Source at Argonne National Labs and have made possible 3D imaging of whole zebrafish at larval and juvenile stages at cell resolution. We are able to recognize and characterize virtually all cell types and achieve elemental imaging of samples containing multiple stains. We propose that high-throughput adaptations of pan-cellular, whole-organism microCT can be used as a powerful foundation for a computational phenomics that will allow us to probe the function of all genes (genetic phenomics) and the effects of environmental manipulations including chemical exposures (environmental phenomics). Current barriers to automated, high-throughput, quantitative morphometric analysis and visualization can be eliminated by adoption of existing and imminent technologies. We envision the creation of one or more beamlines at the Advanced Photon Source at Argonne National Labs dedicated to high-throughput imaging of millimeter scale specimens, including whole invertebrate organisms such as flies and small vertebrate model organisms such as zebrafish. 3D images can be created, analyzed and visualized, and made available to the public in real time. Such data may profoundly enhance our understanding of the impacts of genes and environment on biological systems.

834W

Genetic variants and regulation in human complex diseases and traits. *C. Yao[1], R. Joehanes[1], A.D. Johnson[1,2], B. Chen[1], Tx. Huan[1], J.E. Freedman[3], P.J. Munson[4], D. Levy[1,2,5].* 1) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts; 2) Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland; 3) Department of Medicine, University of Massachusetts Medical School, Worcester, MA; 4) Mathematical and Statistical Computing Laboratory, National Institutes of Health, Bethesda, Maryland; 5) Center for Population Studies, National Heart, Lung, and Blood Institute, Bethesda, Maryland.

Numerous single nucleotide polymorphisms (SNPs) have been found to be associated with complex traits and diseases through genome-wide association studies (GWAS). However, for most of these disease associated SNPs, it is still unclear how genetic variation mediates the phenotype. Here, we present a systematic analysis of SNPs associated with common complex disease and traits. Using 4967 SNPs associated (at $P<5\times10^{-8}$) with 427 complex traits from dbGaP and the NHGRI GWAS Catalog, we constructed a disease network by virtue of common SNP associations. We found several 'hub' loci (HLA region, GCKR, CFH, ZNF259) associated with more than 20 traits. The network identified three large clusters: immune-associated phenotypes, metabolic-associated phenotypes, and cancer, suggesting that genetic variants and their corresponding phenotypes might be related to each other at a higher functional level. We then explored the gene expression associations of trait-associated SNPs (expression quantitative trait loci, eQTLs) in 5257 individuals from the Framingham Heart Study. At FDR <0.05, we identified 886 cis-eQTLs(local regulation) and 59 trans-eQTLs(distant regulation) for 243 complex traits. We found that genes regulated by SNPs related to immune disease and cancer are highly enriched in chr6p21 (HLA region), allograft rejection and genes regulated by metabolic related SNPs are highly associated with metabolic pathways. Furthermore, we found that the expression of transcripts regulated by trait-associated SNPs is also significantly associated with corresponding phenotypes. For example, body mass index (BMI) associated SNP rs10838738 is significantly associated with expression of PTPRJ ($P=5.8\times10^{-27}$), and expression of PTPRJ in turn is also significantly associated with BMI ($P=4.5\times10^{-14}$), suggesting that variants associated with disease phenotypes may exert their function by altering mRNA expression. In exploring eQTLs across multiple tissues (lymphoblastoid, liver and brain) from GTEX and MRCA database, we found 40% of eQTLs on average appeared in more than one tissue, suggesting many trait-associated eQTLs found in blood may have impacts in other tissues. Our findings may provide new insight into complex genetic regulatory mechanisms.

835T

READ1, a Regulatory Element within *DCDC2*, Epistatically Affects Reading and Language with both Deleterious and Protective Alleles. *N.R. Powers[1, 2], J.D. Eicher[1], Y. Kong[4, 5], L.L. Miller[6], S.M. Ring[6], J.R. Gruen[1, 2, 3].* 1) Dept of Genetics, Yale University, New Haven, CT; 2) Dept of Pediatrics, Yale University, New Haven, CT; 3) Dept of Investigative Medicine, Yale University, New Haven, CT; 4) Dept of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 5) W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University, New Haven, CT; 6) School of Social and Community Medicine, University of Bristol, Bristol, UK.

Learning disabilities are neurobehavioral disorders involving unexpected difficulty with a particular mode of learning. Two of the most common are reading disability (RD, also known as dyslexia) and language impairment (LI), which involve difficulty processing written and verbal language, respectively. Both RD and LI are substantially heritable, but like many complex genetic disorders and traits, most of this heritability is missing despite decades of study. Our lab previously identified READ1 (Regulatory Element Associated with Dyslexia 1), a highly polymorphic, purine-rich, compound short tandem repeat localized to intron 2 of the RD-associated gene *DCDC2.* In the Avon Longitudinal Study of Parents and Children (ALSPAC), a large prospective birth cohort based in the UK, we showed a six-marker risk haplotype in strong linkage disequilibrium with allele 5 of READ1 to be strongly associated with severe RD. We also showed another risk haplotype in the same haplotype block~and in strong linkage disequilibrium with allele 6 of READ1~to be strongly associated with LI. We subsequently showed READ1 to bind the potent transcriptional silencer ETV6, and to participate in a synergistic genetic interaction with a known risk haplotype in the 5' region of *KIAA0319*, another known RD-associated gene, to adversely affect performance on reading, language, and IQ measures. We have followed up on these findings by genotyping READ1 in the entire ALSPAC cohort, and examining the effect of individual alleles or groups of alleles on these phenotypic measures. As expected from previous results, READ1 alleles 5 and 6 associated strongly with severe RD and LI, respectively. Interestingly, however, some other READ1 alleles appear to have a protective effect, with odds ratios as low as 0.5 for severe RD. Whether an allele is protective or deleterious seems to depend on its structure: longer alleles relative to the most common allele tend to be deleterious, while shorter alleles tend to be protective. Strikingly, the protective alleles also show epistasis to the *KIAA0319* risk haplotype~while deleterious alleles synergize with it, protective alleles abrogate its effect on phenotype entirely. We postulate that this behavior is due to a direct regulatory interaction through ETV6 between READ1 and the *KIAA0319* promoter, and have undertaken chromatin and expression studies to test this hypothesis.

836F

A computational framework for identifying genes perturbed by MS associated variants through regulatory element disruption. *P. Shooshtari[1, 2], C. Cotsapas[1, 2].* 1) Neurology, Yale University, New Haven, CT; 2) Broad Institute of Harvard and MIT, Cambridge, MA.

Multiple sclerosis (MS) is a progressive neurological disease affecting ~1/700 women in North America, over 60% of whom will die as a direct result. Genome-wide association studies (GWAS) have uncovered tens of genomic loci harboring genetic variants predisposing to the disease. As with other complex traits, majority of these risk variants appear non-coding and we have recently shown that they are enriched in DNase I Hypersensitive Sites (DHS) active in immune cell populations. Although highlighting their regulatory effects, enrichment of these variants on DHS alone is not enough to infer their sites of action. The challenge is thus to understand how these risk variants alter gene regulation, which genes are perturbed and how they fit together into biological processes underlying the disease. For an MS causal SNP to affect the regulation of a gene, it must perturb the DHS regulating that gene's expression. Thus, we reasoned that DHS correlated to relevant genes would be enriched for GWAS p-values compared to those correlated to bystander genes in the same region. We thus devised a two-step test to identify such candidate genes: first, correlate DHS signal to the expression of nearby genes across tissues to determine which they regulate; then calculate a summary GWAS association statistic for each DHS. If these values are correlated, then DHS regulating the gene have evidence of perturbation by risk variants, and thus the gene is likely involved in disease biology. We used DHSs data and exon array data from the ENCODE project for 78 cell types. As a specific example, we studied STAT3, a gene thought to be involved in MS, and identified significantly correlated DHSs that possibly could effect regulation of this gene. Considering a window spanning 50 kbp of each side of STAT3, we found that 24% of significant MS GWAS SNPs on this region are located on correlated DHSs. Given that significantly correlated DHSs ($p < 2.5e-4$) comprise less than 4% of all DHSs surrounding STAT3, this observation implies that MS SNPs are enriched on correlated DHSs, and highlights impact of these SNPs on regulation of STAT3. Our approach thus unites functional genomics analysis and GWAS to identify genes underlying complex traits. It is a general analytical framework that can be applied to other genomic markers (e.g. methylation) and complex traits. We are currently applying this genome-wide to MS GWAS signals to robustly identify implicated genes and will discuss these findings.

837W

TNF-β Nco1 polymorphism and sepsis susceptibility following major elective surgery. *R.Nath. Srivastava[1], K. Baghel[1,2], S. Raj[1,2], A. Chandra[2], S.K. Goel[3].* 1) Orthopaedic Surgery, King George's Medical University, Lucknow, Uttar Pradesh, India; 2) Surgical Gastroenterology, King George's Medical University, Lucknow, Uttar pradesh, India; 3) Petroleum Toxicology Division, Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India.

Background: Postoperative sepsis remains a significant cause of morbidity and mortality. In injured patients, it has been shown that a polymorphism of the tumor necrosis factor-β (TNF-β) gene is related to the development of sepsis. We investigated relation of TNF-β gene polymorphism and serum level of cytokine TNF-α with the development of sepsis after elective major surgery. Methods: The study group consists of 211 patients undergoing major elective surgery. TNF-β NcoI polymorphism was studied in genomic DNA by analyzing restriction fragments of Nco1-digested DNA fragment using Polymerase Chain Reaction. All patients were followed for 1 month following surgery for any evidence of sepsis. Serum TNF-α levels were measured pre and postoperatively by Enzyme Linked Immunosorbent Assay. Genotypes and TNF-α production were related to the occurrence of sepsis if any. Results: 21.80% (n=46) of the patients developed postoperative sepsis. The overall mortality was 4.2% (n=9). The overall allele frequency of TNF-β genotype was 0.32 for TNFB1 and 0.68 for TNFB2. In TNF-β genotype, 11.84% (n=25) patients were homozygous recessive TNFB1, 41.23% (n=87) were heterozygous TNFB1/ TNFB2 and 46.91% (n=99) were homozygous dominant TNFB2. Incidence of postoperative sepsis was significantly (p=0.01) higher in patients homozygous for the allele TNFB2. When compared with patients carrying at least one TNFB1 allele (TNFB1 homozygous and heterozygous genotype), the TNFB2 homozygous genotype was associated with an Odds ratio of 2.60 (p=0.005; 95% CI 1.32 to 5.15) for the development of sepsis. Compared with the heterozygous genotype, the Odds ratio for the homozygous TNFB2 genotype was 3.00 (p=0.003; 95% CI 1.39 to 6.44). In patients with postoperative sepsis, TNF-α serum cytokine levels were significantly higher (p=0.02) in TNFB2 homozygous individual as compared to other genotypes. Conclusion: The development of sepsis was associated with higher capability to produce TNF-α after surgery. TNF-β Nco1 gene polymorphism is found to be associated with development of postoperative sepsis with increased TNF- α serum level. In patients without postoperative sepsis, the TNF-β polymorphism was not related to different levels of TNF-α production. This indicates an association between TNF-β polymorphism and postoperative sepsis, suggesting the B2/B2 genotype as a high risk factor for the development of sepsis after elective surgery.

838T

Genetic variation in the serotonin receptor gene affects immune responses in rheumatoid arthritis. *E. Hesselberg, O. Snir, P. Amoudruz, L. Klareskog, I. Zarea-Ganji, A.I. Catrina, L. Padyukov, V. Malmström, M. Seddighzadeh.* Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

The functional effects of genetic variants associated with the risk of developing rheumatoid arthritis (RA) remain elusive. To assess the impact of the RA-associated serotonin receptor 2A (HTR2A) haplotype on immune cells, we genotyped patients with established RA (n=379) for the risk haplotype via two single-nucleotide polymorphisms (SNPs) in the HTR2A locus: rs6314 and rs1328674. Low-resolution HLA-typing was also performed, and anti-citrullinated protein antibody (ACPA) levels were measured. Patients with and without the RA-associated TC haplotype were selected and T-cell and monocyte function was monitored following in vitro stimulations with staphylococcal enterotoxin B (SEB) and lipopolysaccharide (LPS) using multiparameter flow cytometry. Within the cohort, 44 patients were heterozygous for the TC haplotype (11.6%) while none were homozygous. Upon stimulation, T-cells from TC-carrier patients produced more proinflammatory cytokines (tumor necrosis factor alpha (TNF-α), interleukin-17 (IL-17), interferon gamma (IFN-γ), IL-2, and IL-10) and monocytes produced higher levels of TNF-α compared with patients carrying the non-TC haplotype (P<0.05 and 0.01, respectively). CD4+ and CD8+ T-cells reacted similarly to each other, however, this effect was more significant in CD4+ cells. Such cytokine production could be inhibited in the presence of the selective 5-HT2 receptor agonist (2,5-Dimethoxy-4-iodoamphetamine, DOI); interestingly, this effect was more pronounced in TC carriers. DOI presence also enhanced production of IL-6 and IL-10 with LPS for some patients. HLA type and ACPA-status was not found to significantly correlate to cytokine production variability in this study. Our data demonstrate that association of RA with a distinct serotonin receptor haplotype has functional impact by affecting the immunological phenotype of T-cells and monocytes.

839F

Regulation by the Asthma Susceptibility Gene DENND1B of Rab35 and TNFα signaling. *M.E. March, P.M.A. Sleiman, C. Hou, J. Bradfield, C.E. Kim, J.T. Glessner, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Asthma is a chronic inflammatory condition of the lungs, characterized by excessive responsiveness of the lungs to stimuli, in the forms of infections, allergens, and environmental irritants. Currently, 22.9 million Americans suffer from asthma, and the prevalence has increased dramatically since 1980. Asthma is the leading chronic illness in U.S. children, with 6.8 million affected in 2006. Twin studies have shown that there is a genetic element to asthma susceptibility (with heritability of the condition estimated at between 0.36 and 0.77), but only a fraction of the heritability is explained with previously identified loci. In a genome wide association screen (GWAS), we identified DENND1B as a novel asthma susceptibility locus. At least two splicing isoforms of DENND1B exist, and we have attempted to examine the relative activities of these isoforms. Overexpression of the short or long isoform of DENND1B had opposite effects on the response of cells to TNFα stimulation, with the long isoform enhancing the response and the short isoform diminishing it. The long isoform shows localization to the plasma membrane, while the short isoform shows primarily cytoplasmic staining. The conserved DENN domain functions as a guanonucleotide exchange factor for the small GTPase Rab35. We have demonstrated that TNFα stimulation activates Rab35 through promotion of the GTP-bound form of the protein. We are currently determining the role of Rab35 in TNFα signaling and the influence of both DENND1B isoforms on that signaling.

840W

Mild deficiency of Methylenetetrahydrofolate reductase (MTHFR) increases resistance to malaria in mice. *D.N. Meadows[1], M. Pyzik[1,2], Q. Wu[3], S. Torre[1,2], P. Gros[1,2,4], S. Vidal[1,2,5], R. Rozen[1,3].* 1) Department of Human Genetics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 2) Complex Traits Group, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 3) Department of Pediatrics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 4) Department of Biochemistry, McGill University, Montreal, Quebec, Canada; 5) Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada.

A common polymorphism (677C→T (A222V)) in methylenetetrahydrofolate reductase (*MTHFR*) results in a mild enzymatic deficiency and hyperhomocysteinemia. MTHFR generates 5-methyltetrahydrofolate, the primary circulatory form of folate, which is utilized in homocysteine remethylation to methionine. Homozygosity for this SNP is a risk factor for neural tube defects and may also increase risk for vascular disease, pregnancy complications and cancer. Despite these selective pressures against the T allele, homozygosity is frequent (~5-15% in many populations, with highest levels in the Mediterranean region and Hispanic populations). To determine whether mild MTHFR deficiency may have been maintained through a selective advantage against malaria, as suggested for other human mutations in the Mediterranean region, we examined Mthfr-deficient mice and MTHFR-overexpressing mice for resistance to cerebral malaria infection with *Plasmodium berghei* ANKA (*Pb*A). Male mice received 10^5 parasites and survival was monitored over 2 weeks. Compared with wild-type littermates, $Mthfr^{+/-}$ mice survived infection longer (p=0.02; log-rank test), and $MTHFR^{Tg}$ mice died more quickly (p<0.05; log-rank test). Percent parasitemia revealed a trend (p=0.067) toward lower values in $Mthfr^{+/-}$ (5.5±0.5) compared with $Mthfr^{+/+}$ (7.0±0.4) mice, with no differences in this parameter in $MTHFR^{Tg}$ mice. Splenocytes analyzed by flow cytometry with several markers for T and NK cell populations, showed increased numbers of total lymphocytes, as well as more $CD4^+$ and $CD8^+$ T cells and more $CCR4^+$ NK cells in $Mthfr^{+/-}$ animals, compared with $Mthfr^{+/+}$ mice (p<0.05 for all cell types). $MTHFR^{Tg}$ mice showed decreased numbers of NK cells (p<0.05) as well as decreased numbers of $CCR4^+$ NK cells (p<0.01). Serum interferon-γ (IFNγ) levels, measured by ELISA, were lower in $Mthfr^{+/-}$ mice (88.7±27.3 pg/mL) compared with $Mthfr^{+/+}$ mice (217.1±49.6pg/mL) (p<.05). IFNγ levels in tissues, measured by immunoblotting, were higher (by ~40%) in spleen (p<0.01) and brain (p<0.001) of $Mthfr^{+/-}$ compared with $Mthfr^{+/+}$ mice. Our results suggest that differences in the balance between pro- and anti-inflammatory immune modulators in $Mthfr^{+/-}$ mice may protect them from *Pb*A infection. This is the first time that mild MTHFR deficiency has been shown to confer an advantage, in the outcome of malarial infection, suggesting a possible mechanism that may have maintained the 677 polymorphism in human populations.

841T

***In vivo* modeling of genetic mechanisms associated with sickle cell disease nephropathy.** *B.R. Anderson[1], E.E. Davis[2], M.J. Telen[3], A.E. Ashley-Koch[1].* 1) Center for Human Genetics, Duke University Medical Center, Durham, NC; 2) Department of Pediatrics, Duke University Medical Center, Durham, NC; 3) Department of Medicine, Duke University, Durham, NC.

End-organ damage in patients with sickle cell disease (SCD) has become an emergent clinical priority over recent decades due to the increased lifespan of affected individuals. Renal failure, which occurs in 4-12% of SCD patients and is strongly associated with early mortality, has become a particular concern. The detection of sickle cell disease nephropathy (SCDN) relies on relatively late markers of the disease process, namely proteinuria and reduced glomerular filtration rate (GFR). Therefore, at-risk SCD patients cannot be identified prior to end-organ damage. A genomic region on human chromosome 22 containing two genes, *MYH9* and *APOL1*, has been associated with non-SCD nephropathy, although the primary gene responsible has remained elusive due to strong linkage disequilibrium in this region. Our group demonstrated that both *MYH9* and *APOL1* are strong, independent genetic predictors of risk for proteinuria in SCD (Ashley-Koch *et al*., 2011). Here, we use zebrafish as a model to study the contribution of each gene (*myh9* and *apol1*) to kidney function and filtration. To test independent effects of the knockdown of *myh9* or *apol1*, we injected morpholino antisense oligonucleotides in wild-type zebrafish embryos; this resulted in generalized edema and reduced glomerular filtration (as measured by quantitative dextran clearance) for both gene suppression models. These morphant phenotypes were rescued significantly by co-injection of each respective wild type human *MYH9* and *APOL1* mRNA. Importantly, co-injection of human mRNA corresponding to other *APOL* gene family members did not significantly rescue the observed phenotype, suggesting that *apol1* is indeed the functional ortholog to the human gene. Next, we investigated the possibility of a genetic interaction between *MYH9* and *APOL1* by co-suppression of each of the zebrafish orthologous genes. We observed no additive or synergistic effects. Instead, the double morphants were indistinguishable from the *myh9* morpholino alone, and neither single morpholino could be rescued by the human mRNA of the other gene. These data suggest that *MYH9* and *APOL1* may function independently, but converge on the same biological process to give rise to SCDN. By offering new insights into the contribution of genes that regulate renal function, these results further our understanding of the pathogenesis of SCDN and provide genetic markers for the identification of at-risk SCD patients prior to the onset of kidney dysfunction.

842F

Evolution and functional genomics of host-pathogen interaction in Crohn's Disease. *B.M.P. Bowen, W. Zhang, K. Hui, J.H. Cho, TAGC.* Department of Genetics, Yale University, New Haven, CT.

Crohn's disease (CD), a type of inflammatory bowel disease, affects over a million people worldwide.The prevalence of CD among individuals of Ashkenazi Jewish (AJ) ancestry is 4-7 times higher than that of non-Jewish Europeans. Genome-wide association studies have shown a remarkable overlap between the most strongly associated CD genes and mycobacterial susceptibility genes (but opposite risk alleles), which is of particular interest given the history of mycobacterial resistance documented in the AJ population. In order to determine if the role of CD genes in response to mycobacteria contributes to the AJ CD prevalence discrepancy, we combined computational and functional genomics data from AJ and European non-Jewish (EA) cases and controls. We completed whole genome sequencing of 128 AJ genomes as part of The Ashkenazi Genome Consortium, which revealed a number of CD/mycobacterial susceptibility genes containing variants with derived allele frequencies that were lower or absent in non-Jewish Europeans. This observation led us to hypothesize that these common derived alleles that currently confer susceptibility to CD conferred a past fitness advantage in terms of controlling mycobacterial infection. We then performed functional genomic studies of host response to acute and latent mycobacterial infection in monocyte-derived macrophages (MDMs) from AJ CD cases and controls, as well as EA CD cases and controls. In response to the pathogen Mycobacterium bovis BCG, a mycobacterial strain with antigenicity similar to that of tuberculosis, we report marked inter-individual variability in the efficacy of mycobacterial killing and population differences in serum cytokine levels of TNF, IFN, IL10, and IL1b throughout the course of infection. Lastly, we previously reported a NOD2 co-expression network connecting many CD susceptibility genes identified in the Immunochip, and here we identify that same co-expression network in MDM responses to BCG infection in AJ cases and controls, suggesting that CD susceptibility genes do play a role in mycobacterial clearance.

843W

Identification of a regulatory variant that binds a transcriptional activator complex including FOXA1 and FOXA2 at the *CDC123/CAMKD* type 2 diabetes GWAS locus. *M.P. Fogarty[1], S. Vadlamudi[1], M.E. Cannon[1], K.J. Gaulton[2], K.L. Mohlke[1].* 1) Dept Genetics, Univ North Carolina, Chapel Hill, NC, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

For many of the type 2 diabetes (T2D) loci identified through genome-wide association studies, signals are localized to non-protein-coding intronic and intergenic regions and likely contain variants that regulate gene transcription. The *CDC123/CAMK1D* T2D association signal on chromosome 10 spans an intergenic region between *CDC123* and *CAMK1D* and also overlaps the *CDC123* 3'UTR. To gain insight into the molecular mechanisms underlying this association signal, we examined all SNPs in LD $r^2 \geq .7$ (CEU, 1000G phase 1) with the GWAS index SNP rs12779790 overlapping a predicted regulatory region. Maps of islet and liver open chromatin, histone modifications and transcription factor ChIP-seq were used to identify potential regulatory elements. Based on these criteria, two regions containing T2D-associated variants were tested for enhancer activity using luciferase reporter assays. A 151 bp region surrounding rs11257655 located 15 kb from the 3' end of *CDC123* and 84 kb from the 5' end of *CAMK1D* displayed allele-specific transcriptional activity in rat 832/13 and mouse MIN6 insulinoma cells as well as in human HepG2 hepatocellular carcinoma cells. The rs11257655 risk allele T showed greater transcriptional activity then the non-risk allele C in all cells tested (832/13 $P = 6 \times 10^{-3}$, MIN6 $P = 2 \times 10^{-5}$, HepG2 $P = 8 \times 10^{-6}$). Site-directed mutagenesis of a second SNP included in the 151 bp region due to proximity verified rs11257655 as the driver of allelic differences in enhancer activity. Increased transcriptional activity with the rs11257655 T2D risk allele T is consistent with the direction of an eQTL for *CAMK1D* in blood (Voight, Nat Genet 42; 579). Data from the ENCODE consortium demonstrate binding of FOXA1 and FOXA2 transcription factors in HepG2 cells to a region that overlaps rs11257655. We next used DNA affinity capture and electromobility shift and supershift assays to assess whether a 20 bp region containing alleles of rs11257655 differentially binds these factors. The rs11257655 T allele showed allele specific binding of FOXA1/FOXA2 in beta cell lines and hepatocellular carcinoma cells. Taken together these results demonstrate the utility of maps of regulatory activity to guide identification of regulatory variants. The results suggest that rs11257655 is part of a *cis* regulatory complex affecting transcriptional activity through binding of a protein complex that includes FOXA1 and FOXA2 providing a potential molecular mechanism at this GWAS locus.

844T

Gene-gene interaction and RNA splicing profiles of MAP2K4 gene in rheumatoid arthritis. *K. Shchetynsky, M. Roninnger, D. Protsyuk, L. Klareskog, L. Padyukov.* Rheumatology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.

Background and objectives: MAP2K4 encodes a mitogen activated protein kinase kinase 4 (MKK4), important for optimal activation of JNK1-3 and p38 ~ the two members of the MAP kinase family. In this study we explore the interaction between MAP2K4 locus and two major known genetic risk factors for autoantibody positive rheumatoid arthritis (RA) ~ HLA-DRB1 shared epitope (SE) alleles and PTPN22 rs2476601. We also address the balance in the expression of alternative MAP2K4 splice forms and its association with known SE and PTPN22 genotypes and the autoantibody profile of the disease. **Methods:** The genotypes from 1985 patients with RA and 2252 matched healthy controls from the Swedish EIRA study population and from 863 RA cases and 1181 controls from the NARAC were used in the study. The interaction analysis was performed on 22 SNPs from the MAP2K4 locus, HLA-DRB1 shared epitope alleles and rs2476601 from PTPN22 by calculation the attributable proportion due to interaction (AP). We studied transcript diversity of MAP2K4 and investigated relative expression of MAP2K4 forms in peripheral mononuclear cells for 44 RA patients and 44 controls of Caucasian ancestry. These results were analyzed against available genotypic and phenotypic data. **Results:** We found MAP2K4 rs10468473 in statistical interaction with SE, and PTPN22 rs2476601 in autoantibody positive RA (AP 0,197 [95% CI 0,098 - 0,296] and AP 0,28 [95% CI 0,074-0,48] respectively; EIRA+NARAC) in two independent cohorts and a combined meta-cohort. In our assessment of allelic expression, RA patients heterozygous for rs10468473 demonstrated significantly elevated MAP2K4 expression in comparison to individuals homozygous for G allele. We also observed a novel 'skipped exon' type RNA splice variant of MAP2K4 in our study material, and a potential protein isoform of corresponding molecular weight. MAP2K4 splice forms were differentially expressed in peripheral blood material from 88 RA cases and controls. Within the group of RA patients, a correlation was observed between MAP2K4 variants expression and number of carried SE alleles, as well as with other phenotypic data. **Conclusion:** Our data suggest interaction between MAP2K4, PTPN22 and HLA-DRB1 in development of autoantibody positive rheumatoid arthritis. We also found an allele-specific effect of rs10468473 on MAP2K4 expression, and splicing events in transcripts from the MAP2K4 locus that could be relevant in disease pathogenesis.

845F

A polymorphism in human estrogen-related receptor beta (ESRRβ) associated with physiologic measures of noise-induced hearing loss. *I. Bhatt[1], S. Phillips[1], S. Richter[2], R. Morehouse[3], D. Tucker[1], K. Lundgren[1], V. Henrich[4].* 1) Communication Sciences and Disorders, University of North Carolina at Greensboro, Greensboro, NC. 27402, USA; 2) Department of Mathematics and Statistics, The University of North Carolina at Greensboro, Greensboro, NC 27402, USA; 3) Department of Communication Sciences & Disorders, Appalachian State University, Boone, NC, 28608, USA; 4) Center for Biotechnology, Genomics & Health Research, The University of North Carolina at Greensboro, Greensboro, NC 27402, USA.

Noise-induced hearing loss (NIHL) is a common form of hearing loss and a growing health concern despite national standards for hearing protection and public health awareness campaigns. An NIHL gene association study with college-aged musicians has associated a non-synonymous single nucleotide polymorphism (rs61742642; C→T, P386S) in the ligand-binding domain of human estrogen-related receptor beta (ESRRβ, OR = 2.8, CI = 1.4-5.9, p = 0.003) with increased susceptibility to bilateral 4 to 6 kHz hearing loss. ESRRβ is expressed in major cochlear structures except hair cells. Methods: The purpose of the study was to examine the effects of the ESRRβ polymorphism on both pre-noise exposure (baseline) cochlear physiology and on temporary changes in hearing sensitivity following a brief noise exposure. We examined baseline cochlear physiology by pure-tone audiometry, distortion product otoacoustic emissions (DPOAE growth function and high resolution DPgram), transient-evoked otoacoustic emissions (TEOAE) and TEOAE suppression in 40 and 20 participants carrying the ESRRβ rs61742642 CC vs. CT genotype respectively. Temporary changes in hearing sensitivity were induced by 10 minutes exposure to 90 dB SL narrowband noise (center frequency = 2 kHz) and the changes were evaluated by audiometry (at 2, 3 and 4 kHz). Results: Baseline DPOAE amplitudes were lower across the 2 to 4 kHz frequency range (DPgram overall mean difference (MD) = 2.66 dB, p = 0.03; DP growth function overall MD = 1298.7μPa^2, p = 0.026) in the participants with a CT genotype compared to a CC genotype. Baseline TEOAE contralateral suppression was reduced (MD = 0.35 dB, CI = -0.69 to -0.006, p = 0.046) in participants with the CT genotype. Regression analysis showed that individuals carrying the CT genotype acquired greater deterioration (p <0.001) in hearing sensitivity (MD = 3.47 dB, CI = 1.94 - 4.99) following 10 minutes of noise exposure compared to the participants with CC genotype. Conclusions: The results indicate that individuals with the ESRRβ polymorphism show reduced outer hair cell activity possibly due to poor chemical regulation. They also showed impaired activity of medial olivocochlear nerve fibers and acquired greater deterioration in hearing sensitivity following the noise exposure. These data suggest that the ESRRβ polymorphism increases susceptibility to NIHL, and also indicate the efficacy of otoacoustic emissions testing for identifying sound processing endophenotypes.

846W

Genetic risk variants for autoimmune diseases regulate gene expression in thymus. *I.S.M. Gabrielsen[1], S. Svanstrøm Amundsen[1], H. Helgeland[1], S. Tennebø Flåm[1], N. Hatinoor[1], K. Holm[2], M.K. Viken[1], B.A. Lie[1].* 1) Medical Genetics, Oslo University Hospital, Oslo, Oslo, Norway; 2) Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Oslo, Norway.

Genome-wide association studies (GWAS) have boosted our knowledge of genetic risk variants in autoimmune diseases (AIDs). Despite the fact that AIDs are extremely heterogeneous, GWAS indicate that certain loci and genes seem to predispose to multiple immune-related diseases. Most of the genetic risk variants for AIDs are located within or near genes with immunological functions, but because the majority is found to be non-coding, it's considerable to believe that these risk variants have a regulatory role. Expression quantitative trait locus (eQTL) screening is a widely used method to investigate gene regulation, and by correlating genetic association data with gene expression data, genetic risk factors acting through regulatory variants could be determined. Based on this, and the fact that thymus is an immunological important organ, we have performed an eQTL screen to investigate whether single nucleotide polymorphisms (SNPs) associated with AIDs influence gene expression in thymic tissue (i.e. whether disease associated SNPs serve cis-acting regulatory functions). The genotypes from 42 Norwegian thymic tissue samples genotyped using the Immunochip were analyzed for 420 AID-associated SNPs selected from the National Institute of Health's catalog of GWA studies (NHGRI). Gene expression was measured using Illumina Human WG-6 v3 and genotypes were correlated with gene expression of surrounding genes (± 1Mb of the SNP). By logistic regression analysis, we robustly identified 7 genes (FCRL3, LOC339804, ERAP1, RNASET2, LOC728734, SIRPG and LOC388814) that correlated with disease-associated SNPs at a study-wide level of significance ($P < 2.12\times10^{-5}$). Fine mapping of these regions led to the discovery of even stronger correlating SNPs within four regions: rs10181042 (LOC339804, $P < 3.5\times10^{-10}$), rs30377 (ERAP1, $P < 3.3\times10^{-11}$), rs429083 (RNASET2, $P < 2.2\times10^{-09}$) and rs991774 (LOC388814, $P < 7.2\times10^{-07}$). The regulatory function of the risk polymorphisms and proxy SNPs in strong linkage disequilibrium (r^2>0.8) were evaluated using the RegulomDB. Newly defined SNPs associated with AIDs from recent Immunochip projects were also included. Combining these resources we have highlighted functional variants for 7 genetic regions that potentially can represent causal autoimmune risk variants. We conclude that our study shows that several autoimmune risk variants act as eQTLs in thymus.

847T

The apical sodium-dependent bile acid transporter ASBT (*SLC10A2*) affects the progression of primary biliary cirrhosis in Japanese patients via its transcriptional activity. *K. Taira[1], T. Inamine[1], A. Kawauchi[1], S. Kondo[1], M. Nakamura[2,3], K. Tsukamoto[1].* 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch, Nagasaki, Japan; 2) Dept Hepatology, Nagasaki Univ Grad Sch, Omura, Japan; 3) Clin Res Center, Natl Hosp Organi Nagasaki Med Center, Omura, Japan.

PURPOSE Primary biliary cirrhosis (PBC) is a chronic liver disease characterized by destruction of the intrahepatic small bile ducts, leading to cholestasis, fibrosis, cirrhosis, and eventually liver failure. In order to identify the genetic determinants of PBC progression, we focused on apical sodium-dependent bile acid transporter (ASBT, encoded by *SLC10A2*), which plays a key role in reabsorption of luminal bile acids in epithelial cells of the intestines under the enterohepatic circulation, examined an association between *SLC10A2* polymorphisms and the susceptibility to PBC progression, and investigated the function of these polymorphisms. METHODS A total of 309 Japanese PBC patients were classified into the following two groups (early stage and late stage) based on liver biopsy results and/or clinical manifestations. The 12 tag single nucleotide polymorphisms (SNPs) in *SLC10A2* were genotyped by PCR-restriction fragment length polymorphism or -direct DNA sequencing. Furthermore, the 2.5-kb upstream promoter sequences of *SLC10A2* were subject to PCR-direct DNA sequencing in order to identify new SNPs within this region. The frequencies of alleles and genotypes of 15 SNPs including 3 newly identified SNPs were compared between PBC patients in early and late stages by chi-square test. To investigate the function of the SNP associated with PBC progression, the 50-bp promoter sequence containing the allele of the associated SNP, which was incorporated into luciferase reporter plasmid, was transfected into Caco-2 cells, and the luciferase activities were compared between two alleles of the associated SNP. RESULTS Five SNPs of *SLC10A2*, including 3 newly identified SNPs in the promoter region, showed the significant association with PBC progression. Among them, one of the newly identified SNPs indicated more stronger association with PBC progression and affected the transcriptional expression of *SLC10A2* in Caco-2 cells. CONCLUSION Our results suggest that PBC patients with the low-risk allele of *SLC10A2* may represent lower expression of ASBT and then diminution of reabsorption of luminal bile acids in comparison to those without the allele, resulting in a decrease in enterohepatic circulation of bile acids and eventually leading to slower PBC progression. Thus, *SLC10A2* appears to be a genetic determinant of PBC progression in Japanese patients.

848F

Comprehensive functional assays for variants of unknown significance. *J.O. Kitzman[1], L. Starita[1,2], R. Lo[1,2], R. Qiu[1], S. Fields[1,2], J. Shendure[1].* 1) Dept Genome Sciences, Univ Washington, Seattle, WA; 2) Howard Hughes Medical Institute.

With the relative ease of variation discovery by massively parallel sequencing, functional interpretation becomes an increasingly critical bottleneck in human genetics. As a case in point, clinical diagnostics frequently reveal missense variants of uncertain significance (VUS), even in genes with clear disease relevance such as the tumor suppressor BRCA1, the primary risk factor for inherited breast cancer. To address this bottleneck, we demonstrate a massively parallel mutagenesis approach called PALS (Programmed ALlelic Series). This approach uses libraries of microarray-derived mutagenesis primers to direct synthesis of every possible missense mutant of a gene of interest. As a proof of principle, we designed a library comprising all possible single-codon substitutions in the human tumor suppressor gene TP53 and a similar library for the well-characterized yeast transcription factor GAL4. Deep sequencing and tag-directed assembly (Hiatt et al., Nat Methods 2010) revealed that each library covered the majority of the designed mutational space. To demonstrate how PALS libraries could be used to comprehensively profile loss-of-function mutations, we subjected the GAL4 library to a function-dependent, in vivo screen in budding yeast. In this type of selection experiment, mutant haplotypes' fitness values are determined by deep sequencing to estimate their abundances before and after selection. This assay revealed a subset of missense variants which, together with the clear loss-of-function variants (premature stops and frameshift mutations), exhibited no activity and were depleted by selection. This approach - massively parallel synthesis and sequencing coupled to functional selection - provides a general framework for deep surveys of the functional impacts of mutations to clinically relevant genes.

849W

Expression of Two Genes from the Candidate Locus of Chromosome X in Rheumatoid Arthritis and Systemic Sclerosis. *L.M. Diaz-Gallo[1,2], K. Shchetynsky[1], A. Nordin[1], A.I. Catrina[1], J. Martín[1,2], L. Padyukov[1].* 1) Karolinska Institutet, Stockholm, Stockholm, Sweden; 2) 2.Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Granada, Spain.

There is increasing evidence of association between variants at Xq28 genomic region and autoimmune diseases (AIDs). The genes interleukin-1 receptor-associated kinase 1 (IRAK1) and methyl-CpG-binding protein 2 (MECP2) are located there. We evaluated if there was differential mRNA expression of the IRAK1 and MECP2 between rheumatoid arthritis (RA) patients, systemic sclerosis (SSc) patients and healthy controls. Additionally, we assessed the correlation of IRAK1 and MECP2 expression levels with the rs1059702 and rs17435 polymorphisms located in this locus and previous associated to AIDs. The gene expression of IRAK1 and MECP2 was measured using quantitative PCR by TaqMan assays. We studied 50 RA patients, 35 SSc patients and 52 healthy controls from a European descendent population, all females. Predesigned TaqMan 5' SNP genotyping assays were used to genotype the mentioned SNPs in the available genomic DNA from the studied subjects. The IRAK1 and MECP2 expression was significantly different between the RA patients, SSc patients and controls (p=0.0003, p= 0.0006 respectively, Kruskal-Wallis test). When we compared each group of patients against controls we observed that the IRAK1 expression was significantly decreased in the RA patients compared with controls (p= 0.0022;RA relative quantity (RQ) mean=42.37; controls RQmean=60.28, Mann-Whitney test). But there was no significant difference between the levels of IRAK1 expression between SSc patients and controls (p=0.39; SSc RQmean=46.8; controls RQmean=42.12). We found no significant difference in the MECP2 expression levels between RA patients and controls (p=0.05; RA RQmean=45.72; controls RQmean=57.06). Meanwhile, the SSc patients exhibited a significant increased expression level of MECP2 compared with controls (p=0.017; SSc RQmean=51.86; controls RQmean= 38.71). Finally, we observed that there was a moderated positive correlation between the IRAK1 and MECP2 expression in the studied individuals (correlation coefficient=0.613, p<0.0001). Although the difference was not statistically significant, the homozygous for the minor allele of the rs1059102 tend to express higher levels of both IRAK1 and MECP2 genes. Our study showed an altered expression of IRAK1 and MECP2 genes in both RA and SSc compared to the controls. This study increases the evidence that suggest an important pathogenic role of this X chromosome locus in AIDs.

850T

An enhancer element harboring variants associated with systemic lupus erythematosus engages the *TNFAIP3* promoter to influence A20 expression.
S. Wang[1], F. Wen[1], G.B. Wiley[1], M.T. Kinter[2], P.M. Gaffney[1]. 1) Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma city, OK, 73104; 2) Free Radical Biology and Aging Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, 73104.

Functional characterization of causal variants present on risk haplotypes identified through genome-wide association studies (GWAS) is a primary objective of human genetics. In this report, we evaluate the function of a pair of tandem polymorphic dinucleotides, (rs148314165, rs200820567, collectively referred to as TT>A) recently nominated as causal variants responsible for genetic association of systemic lupus erythematosus (SLE) with tumor necrosis factor alpha inducible protein 3 (*TNFAIP3*). *TNFAIP3* encodes the ubiquitin-editing enzyme, A20, a key negative regulator of NF-κB signaling. A20 expression is reduced in subjects carrying the TT>A risk alleles, however, the underlying functional mechanism by which this occurs is unclear. We used a combination of electrophoretic mobility shift assays (EMSA), mass spectrometry (MS), reporter assays, chromatin immunoprecipitation-PCR (ChIP-PCR) and chromosome conformation capture (3C) EBV transformed lymphoblastoid cell lines (LCL) from individuals carrying risk and non-risk *TNFAIP3* haplotypes to characterize the effect of TT>A on A20 expression. Our results demonstrate that the TT>A variants reside in an enhancer element that binds NF-κB and SATB1 enabling physical interaction of the enhancer with the *TNFAIP3* promoter through long-range DNA looping. Impaired binding of NF-κB to the TT>A risk alleles or knockdown of SATB1 expression by shRNA, inhibits the looping interaction resulting in reduced A20 expression. Together, these data reveal a novel mechanism of *TNFAIP3* transcriptional regulation and establish the functional basis by which the TT>A risk variants attenuate A20 expression through inefficient delivery of NF-κB to the *TNFAIP3* promoter. These results provide critical functional evidence supporting a direct causal role for TT>A in the genetic predisposition to SLE.

851F

Gene-gene interactions of EG-VEGF, PKR1 and PKR2 genes and the risk of recurrent miscarriages.
M. Su[1], P. Kuo[1], S. Lin[2]. 1) National Cheng-Kung University Hospital, Tainan City, Taiwan; 2) Insitute of clinical medicine, National Cheng-Kung Univeristy, Tainan City, Taiwan.

Background: Endocrine gland-derived vascular endothelial growth factor (EG-VEGF) and its receptor genes (PKR1 and PKR2) play an important role in human early pregnancy. Our previous study showed that PKR1 and PKR2 polymorphisms are associated with recurrent miscarriages (RM). This study was conducted to find EG-VEGF, PKR1 and PKR2 variants in the coding regions of idiopathic RPL patients and further evaluate gene-gene interactions in 3 genes. Methods: Two hundred and ninety one blood samples from 142 RPL women and 149 controls were nucleotide sequenced in the coding regions of EG-VEG, PKR1 and PKR2. Gene-gene interaction was evaluated in 3 gene variants using multifactor dimensionality reduction (MDR) method. Result(s): One each nonsynonymous variant of 3 genes were identified, and PKR1(I379V) and PKR2(V331M) were significantly associated with idiopathic RM (p=0.006 and p=0.002, respectively). Genetic interactions were founded not only between PKR1(I379V) and PKR2(V331M), but also among EG-VEGF (V67I), PKR1(I379V) and PKR2(V331M) (p=0.01 and p=0.01, respectively). Women carried low-risk genotypes reduced 77% risk of experiencing miscarriages compared with those carried high-risk genotypes. Conclusion(s): The present study corroborates the clinical relevance of the EG-VEGF system in human early pregnancy, and provides evidence for the gene-gene interactions of EG-VEGF and PKR variants.

852W

Gene- asbestos exposure interactions on lung cancer risk.
C. Liu[1,2], I. Stücker[3], C. Chen[4,5,6], G. Goodman[7], M.K. McHugh[8], A.M. D'Amelio[8,9], C.J. Etzel[8], S. Li[1], X. Lin[10], D.C. Christiani[1,11]. 1) Dept Environmental Hlth, HSPH, Boston, MA, USA; 2) Institute of Environmental Health, College of Public Health, National Taiwan University, Taipei, Taiwan; 3) INSERM U 754-IFR69, Villejuif, France; 4) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Mailstop M5-C800, Seattle, WA 98109-1024, USA; 5) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA; 6) Department of Otolaryngology: Head and Neck Surgery, School of Medicine, University of Washington, Seattle, WA, USA; 7) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Mailstop M3-A306, Seattle, WA 98109-1024, USA; 8) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, 1155 Pressler Boulevard, Unit 1340, Houston, TX 77030, USA; 9) Biomath/Biostatistics Program, The University of Texas Graduate School for Biomedical Sciences, Houston, TX 77030, USA; 10) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 11) Department of Medicine, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA.

Occupational asbestos exposure has been found to increase lung cancer risk based on epidemiologic studies. In order to clarify the respective roles of genetic factors and asbestos exposure on lung cancer risk, we conducted the asbestos exposure-gene interaction analyses among Caucasian populations who are current or ex-smokers. The stage 1 discovery stage included 833 Caucasian cases and 739 Caucasian controls using Illumina Human 610-Quad BeadChips. Several independent populations were included in the second replication stage. Cumulative lifetime asbestos exposure score (AES), which was previously developed and validated in our study population, was calculated from self-reported duration and intensity of occupational and nonoccupational exposures. The top ranked SNPs from discovery stage were replicated within International Lung and Cancer Consortium (ILCCO). A two-stage replication approach was conducted. First in silico replication was conducted in those groups that have GWAS and asbestos exposure data, including 1548 cases and 1544 controls. Followed by de novo genotyping to replicate the results from in silico replication, 1539 cases and 1761 controls were genotyped. In order to have the data comparable between studies, genotyping results were imputed by MACH using 1000 Genome dataset. Asbestos exposure measurements were recategorized as 'high' or 'low and none' asbestos exposure effects based on each group's previous findings. Logistic regression adjusted for potential confounders was used to assess the SNP-asbestos exposure interaction effect on lung cancer risk. The top associated findings been found are located at 22q13.31, 8p22, 16p11.2, which have been reported in association with carcinogenesis and inflammatory response. Further functional analysis will be conducted among the regions.

853T

***RBMS1* genotype strongly influences adiposity and liver function in rural but not urban dwellers.** *M.E.S. Bailey[1], C.A. Celis-Morales[2], M. Staunton[1], N. Ulloa[3], C. Calvo[3], F. Perez-Bravo[4], J.M.R. Gill[5].* 1) School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; 2) Human Nutrition Research Centre, Institute for Ageing and Health, Newcastle University, Newcastle, UK; 3) Department of Clinical Biochemistry and Immunology, Faculty of Pharmacy. University of Concepcion, Chile; 4) Laboratory of Nutritional Genomics, Department of Nutrition, Faculty of Medicine, University of Chile, Santiago, Chile; 5) Institute of Cardiovascular and Medical Sciences, CMVLS, University of Glasgow, Glasgow, UK.

RBMS1 encodes a set of single-stranded DNA-binding proteins that may have roles in regulation of DNA replication, gene expression, cell cycle and cell fate. GWAS meta-analysis and other studies have implicated *RBMS1* as a risk gene in obesity and type 2 diabetes. We show that *RBMS1* strongly influences levels of subcutaneous body fat and cardiometabolic risk-related phenotypes across populations but only in those living a more traditional lifestyle. We analysed an *RBMS1* SNP, rs7593730 (C/T alleles), for association with a range of quantitative trait measures in a cross-sectional population sample from Chile (n=300), including male and female adults of both native American (Mapuche) and European ethnicity living in urban and rural environments. Participants were assessed for a wide range of physical, metabolic, social and environmental variables, including body fat (aggregate of 4 skinfold measures). The T allele (f=0.14 in Europeans, 0.05 in Mapuche) was found to be strongly associated with lower levels of subcutaneous fat, insulin resistance (HOMA-IR) and ALT, and with higher levels of HDL cholesterol (all p<0.00015), with similar effect sizes in Europeans and Mapuche. The association with body fat was still observed (genotype p= 0.009) after adjustment for age, sex, ethnicity, sociodemographic factors, activity (sedentary time, MVPA) and overall food intake. This SNP was also weakly associated with several other variables, including cardiorespiratory fitness. The genotypic association with HOMA-IR was partially independent (genotype p=0.009) of body fat. Genotype was found, in fully adjusted models, to interact strongly with both living environment (rural vs urban; interaction p=0.0004) and fitness (interaction p=0.006) in its effect on adiposity. In stratified analyses, genotype was found to influence body fat strongly in rural dwellers but not urban dwellers, and more strongly in fitter individuals. In fully adjusted models, rs7593730 explained >17% of the variance in adiposity in the rural group (p<0.00001), each copy of the T allele lowering fat levels by >0.4 SDs. These findings support the idea that there may be subgroups defined by genotype in whom activity-boosting interventions to reduce adiposity may be more effective. Furthermore, they suggest that a portion of the cardiometabolic disease risk associated with a Westernised lifestyle may result from the suppression of specific genetic influences on primary risk factors in the urban environment.

854F

Gene-Lifestyle Interaction and Type 2 Diabetes. *R.A Scott on behalf of The InterAct Consortium.* MRC Epidemiology Unit, Cambridge, United Kingdom.

Rapid progress has been made in the understanding of the genetic basis of type 2 diabetes (T2D). It is uncertain whether testing for genetic susceptibility to T2D may be useful to guide decisions about lifestyle interventions. The InterAct study includes 12,403 incident T2D cases and a representative subcohort of 16,154 individuals from a total cohort of 340,234 participants from 8 European countries followed for 3.99 million person-years. We studied the combined effects of an additive genetic T2D risk score comprising 49 variants associated with T2D and modifiable and non-modifiable risk factors using Prentice-weighted Cox-regression and random effects meta-analysis methods. The effect of the genetic score differed significantly by age at study entry (p_{int}=1.20×10^{-4}), due to a larger genetic effect in cases who developed diabetes early (<55 years), compared to later. Relative genetic risk (per standard deviation (4.4 risk alleles)) was also larger in participants who were leaner, both in terms of BMI and waist circumference (WC) (HR normal weight 1.62 (1.50, 1.74), overweight 1.46 (1.37, 1.56), obese 1.27 (1.17, 1.39), p_{int}=1.50×10^{-3}; HR low WC 1.60 (1.49, 1.72), medium WC 1.53 (1.39, 1.68), large WC 1.29 (1.18, 1.40), p_{int}=7.49×10^{-9}). Only one variant (ADCY5 rs11717195 for SNP by BMI interaction) was below the Bonferroni-corrected threshold for individual significance (p_{int}=7.2×10^{-6}), again showing a larger genetic effect size in smaller individuals. Absolute risk demonstrated the strong overall preponderance of obesity for T2D risk; the cumulative T2D incidence over 10 years (per hundred) rose from 0.25, 0.44, and 0.53 to 0.89 across quartiles of the genetic score in normal weight individuals, compared to 4.22, 5.78, 5.83 and 7.99 in obese individuals. We detected no significant interactions between the genetic score and sex, diabetes family history, physical activity, or dietary habits assessed by a Mediterranean diet pattern score. The high absolute risk associated with obesity at any level of genetic risk highlights the importance of universal approaches to lifestyle intervention.

855W

Gene carbohydrate and gene fiber interactions and type 2 diabetes in diverse populations from the National Health and Examination Surveys as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study. *R. Villegas, R.J. Goodloe, B. McClellan Jr, J. Boston, D.C. Crawford.* Med, Vanderbilt Univ, Nashville, TN.

Both environmental and genetic factors impact Type 2 diabetes (T2D). Environmental modifiers of known genotype-T2D associations may account for some of the "missing heritability" of these traits. To identify such modifiers, we genotyped 15 T2D-associated variants identified through genome-wide association studies (GWAS) in 6414 non-Hispanic white, 3073 non-Hispanic black and 3633 Mexican American samples collected for the National Health and Nutrition Examination Surveys (NHANES). In this paper we evaluated interactions between these variants with carbohydrate intake and with fiber intake. We performed logistic regression analysis testing for SNP associations with T2D levels in self-identified European American, African American and Mexican American. Only 1 association generalized across all three populations, rs7903146 (TCF7L2). High intake of carbohydrates was associated with higher risk of T2D while fiber was inversely related to T2D risk in analysis conducted in all participants and in analysis stratified by race. However the trend for the association between fiber and T2D in blacks was no significant and the association between carbohydrate intake and T2D in Mexicans was of marginal significance. We identified one SNP × carbohydrates interactions at a significance threshold of p < 0.05, in blacks (with IGFBP2, rs4402960) and one in Mexican Americans (with PPARG rs1801282). We found one gene-fiber interactions with ADAMT59 rs4607103 in African Americans and another gene-fiber interaction among Mexican Americans with PPARG rs1801282, both at P<0.05 level. We also evaluated the combined effect of genetic variants by calculating a genetic risk score (GRS) and interactions of the GRS with carbohydrate intake and fiber across the racial/ethnic groups. The odds ratio for T2D with each GRS point (per risk allele) was 1.10 (95% CI: 1.06-1.14) for European Americans, 1.07 (95% CI: 1.02-1.13) from non- Hispanic blacks and 1.12 (95% CI: 1.06-1.18) for Mexican Americans. The interaction factor between the GRS and both carbohydrate intake and fiber intake was significant in the non-Hispanic Blacks group but no association was found between the GRS with these 2 nutrients in non-Hispanic Whites or in Mexican Americans. Our results suggest that carbohydrate and fiber may modify genotype-phenotype associations and that this association differs by race.

856T

Variants in the Glucagon Gene are Associated with Baseline Weight and Glycemic Response to Metformin and Intensive Lifestyle Interventions in the Diabetes Prevention Program. *A.H. Winters[1], K.A. Jablonski[2], S.E. Kahn[3], W.C. Knowler[4], E.S. Horton[5], K.J. Mather[6], R.F. Arakaki[7], J.C. Florez[8, 9, 10], T.I. Pollin[11], DPP Research Group.* 1) Human Genetics, University of Maryland, Baltimore, Baltimore, MD; 2) The George Washington University Department of Epidemiology and Biostatistics The Biostatistics Center Rockville, MD; 3) Division of Metabolism, Endocrinology and Nutrition Department of Medicine VA Puget Sound Health Care System and University of Washington Seattle, WA; 4) Diabetes Epidemiology and Clinical Research Section National Institute of Diabetes and Digestive and Kidney Diseases Phoenix, AZ; 5) Joslin Diabetes Center, Harvard Medical School, Boston, MA; 6) Division of Endocrinology and Metabolism, Indiana University School of Medicine, Indianapolis, Indiana; 7) Department of Medicine Clinical Research, University of Hawaii, Honolulu, Hawaii; 8) Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit) Massachusetts General Hospital; 9) Program in Medical and Population Genetics Broad Institute; 10) Department of Medicine Harvard Medical School; 11) Endocrinology, Diabetes and Nutrition, University of Maryland, Baltimore, Baltimore, MD.

The glucagon gene (*GCG*) codes for glucagon, glucagon-like peptides 1 and 2 (GLP1 and GLP2), glicentin, and oxyntomodulin, which have a number of different physiological effects including glucose homeostasis, weight gain and satiety. We hypothesized that variation in *GCG* affects glycemic traits and weight and modifies glycemic response to weight- and glucose-lowering interventions. The Diabetes Prevention Program (DPP) was a clinical trial testing metformin or a lifestyle intervention (systematic caloric and fat intake reduction and increase in physical activity) versus placebo on reducing type 2 diabetes development in those at high risk. We evaluated whether SNPs in *GCG* modify effects of these interventions in the DPP. Thirteen SNPs tagging common variations in *GCG* were evaluated for association with weight, fasting and two hour glucose, glycated hemoglobin and oral disposition index at baseline and 1 year. Among 2815 DPP participants, the T allele of rs2892827 (DPP minor allele frequency=0.14-0.19) was associated with greater weight at baseline (β=1.46[0.19, 2.73] kg/T allele; p=0.02). There were no significant SNP x treatment interactions on baseline-adjusted one year weight (interaction p=0.25 for lifestyle and p = 0.24 for metformin). There was a significant interaction on fasting glucose with both treatments (p=0.02 for SNP*lifestyle and p=0.04 for SNP*metformin), explained by a borderline association with increased baseline-adjusted fasting glucose at one year in the placebo group (β=1.53[-0.03, 3.08]; p=0.055), but not in the lifestyle (β=-0.93[-2.14 ,0.28]; p=0.13) or metformin (β=-0.59[-1.75, 0.57]; p=0.32) groups. Three other SNPs in *GCG* had significant interactions with lifestyle on weight change as measured by baseline-adjusted one year weight (rs11897425 p=0.03; rs13010545 p=0.006; rs10200420 p=0.02), showing evidence that lifestyle reduces the effect of the SNP on weight. No other significant interactions with metformin were seen. rs2892827 is 5' of *GCG* near its known transcription factor binding sites, and the major A allele is conserved in mammals. Thus the T allele may directly influence glycemic response to anti-diabetic interventions, or alternatively be in linkage disequilibrium with other variants with these effects. Further exploration of these genetic factors will increase our understanding of the contributions of glucagon gene products to glucose homeostasis and as modifiers of the effects of metabolic interventions.

857F

Genetic and Environmental Influences on the Age of Onset of Age-Related Macular Degeneration (AMD) in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA) Cohort. *L. Shen[1], R. Melles[1], S. Sciortino[1], D. Ranatunga[1], L. Walter[1], L. Sakoda[1], R. Whitmer[1], T. Hoffmann[2], M. Kvale[2], Y. Banda[2], N. Risch[2], C. Schaefer[1], E. Jorgenson[1].* 1) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 2) Institute for Human Genetics University of California, San Francisco.

Age-Related Macular Degeneration (AMD) is a highly heritable, common disease with a prevalence of 30% in non-Hispanic white adults aged 75 years or older. In addition to genetic factors, environmental risk factors, including smoking behavior, are known to affect the risk of developing AMD. Here we evaluate the effect of previously identified AMD risk variants and environmental risk factors on the age at onset of AMD in the Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging (GERA) cohort (n=110,266). Cohort members completed a health survey questionnaire that included information on smoking behavior and alcohol consumption. A genetic risk score (GRS) was calculated based on 10 independent risk variants that had been identified in prior studies. During the 6 year follow-up period after the GERA enrollment, 3,278 non-Hispanic white AMD incident cases were identified from the electronic medical record (EMR). These cases had an average of 13.9 years of continuous observation via EMR prior to the first AMD diagnosis, limiting the possibility that subjects might have been diagnosed with AMD at a younger age. The mean onset age was 76.16 years (SD=8.52). Subjects were divided into quintiles based on the distribution of the GRS in 78,697 non-Hispanic white GERA cohort members. Cases in the highest GRS quintile had a mean age of onset 2.44 years younger (p=0.01) than those in the lowest quintile. Current smokers had a 2.34 year earlier onset (p=0.007) than non-smokers. Subjects who exceeded the NIAAA daily safe drinking limit at least once per month at the time of the survey had a 2.14 year (p=7.2*10^-6) earlier age of onset. Stratified analyses revealed an interaction between alcohol consumption and GRS (p=0.0134). Exceeding the NIAAA daily limit was associated with a 3 year earlier age of onset in subjects in the top half of the GRS distribution, whereas age of onset was only 0.8 years earlier in the bottom half. Our findings suggest that both genetic and environmental factors influence the age of onset of AMD, and suggest that it may be possible through lifestyle changes to delay the onset of AMD in subjects at high genetic risk.

858W

Perfect pitch in Costa Rica: contribution of genetic and environmental factors. *G. Chavarria-Soley.* Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica.

Absolute pitch (AP), also known as perfect pitch, is the unusual ability to recognize a pitch without an external reference. The most popular current view is that both environmental (such as an early age at the beginning of musical training) and genetic (as evidenced by familial aggregation and twin studies) factors are involved in the acquisition of the trait. With the goal of attempting to dissect the influence of genes versus environment on the manifestation of AP, 89 musicians were subjected to an interview and musical tone identification test. Subjects were recruited using a mixed strategy, which included university music students as well as volunteers who responded to a newspaper article. Therefore, prevalence cannot be estimated. The test consisted of the identification of 40 piano and 40 'pure' tones, with four seconds between tones. Subjects who correctly identified 75 or more tones were considered AP possessors. A total of 16 AP possessors were identified, eight of whom correctly named all 80 tones. None of the AP possessors reported family members with perfect pitch. In most of the cases, however, the parents and siblings of the subjects had no musical training. A new test for AP, which does not rely on the naming of specific tones, could help elucidate whether familial aggregation is present. Interestingly, the average age at the beginning of musical training did not differ between AP (7.8 ± 3.9 years) and non AP (8.8 ± 4.1 years) subjects (t test= 0.81, d.f.= 87, p=0.42), challenging the assumption that early musical training is a requirement for the development of AP. A total of 56% (9 out of 16) of AP possessors began their musical training after age 7; 6 of these after age 10. A possible interpretation is that in the presence of a strong genetic predisposition, musical training can start at more advanced ages. Additionally, there is an interesting group of 16 non-AP possessors who correctly identified more tones than predicted by chance and for whom most incorrectly identified tones were off by only a semitone. They represent a middle category between true AP and true non-AP possessors. In agreement with the likely complex nature of perfect pitch, this category possibly includes individuals in whom genetic predisposition was high and environmental exposure was not enough to develop the trait, or the other way around.

859T

Interplay of Genetic Risk (CHRNA5) and Environmental Risk (partner smoking) on Smoking Cessation Success. *L. Chen[1], T.B. Baker[2], M. Munafò[3], L.J. Bierut[1].* 1) Psychiatry, Washington University School of Medicine, St. Louis, MO, USA; 2) 2Tobacco Research and Intervention, University of Wisconsin, School of Medicine, Madison, WI 53711, USA; 3) 3Department of Biological Psychology, School of Experimental Psychology, University of Bristol, 12a Priory Road, BRISTOL BS8 1TU, United Kingdom.

Objective: Identifying predictors affecting individual variation in quitting success is important for both clinical response prediction and public health. Smoking cessation success can be predicted by both genetic factors such as variants in the nicotinic receptor gene (*CHRNA5*) and environmental factors such as partner smoking. This study tests whether a specific genetic risk for cessation success is moderated by an environmental feature. Method: In a community-based, longitudinal study (N=1,856) of pregnant women who were smokers before pregnancy, and a randomized comparative effectiveness smoking cessation trial (N=1,065), we examine if the effect of partner smoking on smoking cessation varies with a genetic variant in *CHRNA5*. Smoking cessation was defined as the trajectory of amount smoked over time in the observational study of pregnant women, and as the trajectory of alveolar carbon monoxide (CO) levels in the cessation trial. Results: While the pregnant women decreased their smoking quantity over time, both the genetic risk (rs16969968(A)) and partner smoking predict heavier smoking during pregnancy. However, rs16969968 interacted with partner smoking: the genetic risk is significantly increased for expectant mothers who have a partner who smokes (b=0.071, 95% CI=0.013-0.13, p=0.017). Similarly, among the smoking cessation trial participants receiving the placebo, a similar interaction was found between rs16969968 and partner smoking on cessation: The genetic risk is significantly increased for smokers with a partner who smokes (b=0.20, 95%CI=0.049-0.36, p=0.010). This interaction between genetic and environmental factors occurs only in the placebo group and not in the active pharmacotherapy group (significant interaction, b=-0.25, 95% CI=-0.42 to -.091, p=0.0023). Conclusions: The risk of partner smoking and the *CHRNA5* genetic marker interact in both studies of smoking cessation. The risk associated with *CHRNA5* is moderated by the environment, suggesting that the genetic vulnerability may be mitigated by altering environmental factors such as partner smoking. The increased risk of cessation failure with combined genetic and environmental risks can be ameliorated by cessation pharmacotherapy. Incorporating both genetic and environmental factors is critical in designing successful smoking cessation treatments.

860F

Genetic Predisposition for Hypertriglyceridemia is Modified by Extremes of Adiposity. *C. Cole[1], M. Nikpay[1], R. Dent[2], R. McPherson[1].* 1) Atherogenomics Laboratory, Univ Ottawa Heart Institute, Ottawa, ON, Canada; 2) Weight Management Clinic, The Ottawa Hospital, Ottawa, Canada.

We determined the predictive value of a genetic risk score (GRS) for plasma lipids in 2 cohorts of obese and lean subjects genotyped on Affy 6.0. OBLE: 959 OB & 869 LE, BMI 43.1/20.3. CC: 830 OB & 1,044 LE; BMI 35.8/21.5 kg/m2. GRS predicted TG variance better for OB vs LE in both cohorts, whereas the opposite was observed for HDL-C.

Triglycerides					
OBLE cohort	beta GRS mmol/L	adjusted R2	C C cohort	beta GRS mmol/L	adjusted R2
lean	0.094 (5.8e-10)	0.042	lean	0.102 (4.2e-06)	0.024
obese	0.154 (<2e-16)	0.090	obese	0.153 (1.5e-07)	0.044
HDL-C					
OBLE cohort	beta GRS mmol/L	adjusted R2	C C cohort	beta GRS mmol/L	adjusted R2
lean	0.589 (<2e-16)	0.083	lean	0.684 (1.0e-08)	0.036
obese	0.495 (4.3e-14)	0.060	obese	0.535 (6.6e-05)	0.023

.

861W

Education Influences the Association between Genetic Variants and Refractive Error: A Meta-analysis of Five Singapore Studies. *Q. Fan[1], R. Wojciechowski[2], M.K. Ikram[1,3], C-Y. Cheng[1,3], P. Chen[1], X. Zhou[1], C-W. Pan[1], C-C. Khor[1,4], E.S. Tai[5], T. Aung[1,3], T-T. Wong[1,3], Y-Y. Teo[1,6], S-M. Saw[1,3].* 1) SSH School of Public Health, National University of Singapore, Singapore, Singapore; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 3) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 4) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore; 5) Department of Medicine, National University of Singapore, Singapore, Singapore; 6) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

Refractive error is a complex ocular trait governed by both genetic and environmental factors and possibly their interplay. Thus far, data on the interaction between genetic variants and environmental risk factors for refractive errors are largely lacking. By using findings from recent genome-wide association studies, we investigated whether the main environmental factor, education, modifies the effect of 40 single nucleotide polymorphisms (SNP) on refractive error among 8,461 adults from five studies including ethnic Chinese, Malay and Indian residents of Singapore. Three genetic loci exhibited a strong association with myopic refractive error in individuals with tertiary or university education ($P = 1.68 \times 10^{-4}$ to 3.6×10^{-6}), whereas the association at these loci was non-significant or of borderline significance in those with secondary education or lower (P for interaction: 3.82×10^{-3} to 4.78×10^{-4}). The evidence for interaction was strengthened when combining the genetic effects of these three loci (P for interaction = 4.40×10^{-8}), and significant interactions with education were also observed for axial length and myopia. Our study shows that low level of education may attenuate the effect of risk alleles on myopia. These findings further underline the role of gene-environment interactions in the pathophysiology of myopia.

862T

A genome-wide interaction study suggests contrasting interaction effects with smoking in ACPA-positive versus ACPA-negative rheumatoid arthritis. *X. Jiang[1], H. källberg[1], L. Ärlestig[2], S. Rantapää-Dahlqvist[2], L. Klareskog[3], L. Padyukov[3], L. Alfredsson[1], The Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) Study Group.* 1) Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Public Health and Clinical Medicine, Rheumatology, Umeå University; 3) Rheumatology Unit, Department of Medicine, Karolinska Institute, Stockholm, Sweden.

Background: Rheumatoid arthritis (RA) is believed to have a multifactorial etiology, involving both genetic and environmental components, and can be divided into two major subsets based on the presence/absence of anti-citrullinated protein/peptide antibodies (ACPA). Smoking is the most established environmental risk factor. Despite progress from genome-wide association studies (GWAS), identified genetic variants only explain a small proportion of RA occurrence, and gene-environment interaction could add etiologic understanding of the disease. The aim of current study is to investigate large scale gene-environment interaction between smoking and SNPs for each RA subset. Methods: We analyzed data from the Swedish EIRA case-control study using logistic regression models. Smoking history (never/ever smoking) was collected through questionnaires. Genetic markers was obtained from a custom made Illumina Immunochip scan. Interaction between smoking and 133648 genetic markers that passed quality control were examined for the two RA subsets (1590 ACPA positive cases, 891 ACPA negative cases; compared with 1856 controls). Attributable proportion (AP) due to interaction was evaluated for each smoking-SNP pair and corrected for multiple comparisons. We performed replication in a case-control study of RA from northern Sweden. To further validate the results we also performed interaction analysis using GWAS data for the EIRA study. Results: In ACPA positive RA, 102 SNPs were significantly interacting with smoking after Bonferroni correction, all located in the HLA region (one in HLA class I region, the rest in HLA class II region); 51 were replicated in the independent case-control study from northern Sweden. No additional loci besides from chromosome 6 turned up in the GWAS validation. After adjusting for HLA-DRB1 shared epitope (SE), 15 SNPs remained significant for ACPA positive RA, with 8 of them replicated. For ACPA negative RA, no SNP passed threshold for significance. 10 genes were identified for ACPA positive RA, with dominance on antigen presentation pathways: HLA-DOB, HLA-DQA1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB5, TAP2. Conclusion: Our study presents the most explicit picture to date regarding the patterns of gene-smoking interaction in ACPA positive/negative RA, suggesting contrasting etiology of the two subsets. Except for HLA-DR, the study additionally linked RA risk to the class I HLA, implicating the function of cytotoxic T cells in RA pathogenesis.

863F

A study on genetic variation associated with visceral adipose tissue and interaction of life style on the expression of genetic variation. *H. Kwon*[1,2], *K. Sohn*[1], *B. Cho*[1], *H. Choi*[1]. 1) Seoul National University Hospital, Seoul, South Korea; 2) Healthcare Research Institute, Seoul, South Korea.

Introduction : Obesity, especially abdominal obesity is known as a risk factor for various diseases, and also to increase mortality. Among components of abdominal obesity, visceral adipose tissue (VAT) is a risk for various metabolic diseases and mortality. Genetic variations affecting abdominal obesity have been identified, and interactions between genetic variations and lifestyle factors also have been studied. But there has been no study on genetic variations altering VAT in Koreans, nor study on interaction between genetic variations and life style factors affecting VAT in Koreans. In this study, I have tried to find the association of VAT and SNPs which were previously known to be associated with obesity, and also to test the interaction of life style such as smoking, alcohol drinking and physical activity on the association of VAT and SNPs. Methods : I have selected 17 SNPs previously known to be associated with obesity, and analyzed selected SNPs through realtime PCR from the blood of Korean men aged 20 to 65 who took comprehensive health checkup programs including abdominal fat analysis with CT scan. I have tested the association between selected SNPs and obesity traits including VAT with linear regression analysis, and also checked interaction of life style factors on the association. I also tried stratified analysis with the criteria of body mass index 25kg/m2. Results : Rs9939609 was associated with VAT in dominant model, with AT/AA allele type tends to 6.8cm2 higher in VAT, 0.6kg/m2 higher in body mass index, 1.4cm thicker in waist circumference, 0.7% more in body fat, and 11.2cm2 higher in abdominal subcutaneous fat. On interaction analysis, rs1514175, rs2112347, rs10767664 were affected by alcohol drinking, and rs1718537 and rs3817334 were affected by physical activity on their association with VAT, but on stratified analysis by life style factors there was no significant difference in VAT according to allele type. On stratified analysis by body mass index, rs713586 and rs3810291 were associated with VAT only in subjects with body mass index higher than 24kg/m2 Conclusions : Rs9939609, rs713586 and rs3810291 were associated with VAT in Korean male, and this finding could be implemented in practice and study in the field of obesity.

864W

Lineages based genome-wide association analysis in tuberculosis. *S. Mahasirimongkol*[1], *N. Smittipat*[2], *T. Juthayothin*[2], *T. Mushiroda*[3], *S. Wattanapokayakit*[1], *N. Wichukchinda*[1], *S. Nedsuwan*[4], *K. Dokladda*[2], *K. Ruksseree*[2], *P. Billamas*[2], *P. Palittapongarnpim*[5], *B. Chaiyasirinroj*[6], *A. Chaiprasert*[7], *H. Yanai*[8,9], *K. Tokunaga*[9]. 1) Medical genetics section, National Institute of Health, Nonthaburi, Thailand; 2) National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumtani; 3) Laboratory for Pharmacogenetics, Center for Integrative Medical Sciences, RIKEN, Japan; 4) Chiangrai prachanukroh hospital, Chiangrai, Thailand; 5) Department of Microbiology, Faculty of Science, Mahidol University; 6) TB/HIV research foundation, Chiangrai, Thailand; 7) Department of Microbiology, Faculty of Medicine Siriraj Hospital, Bangkok, Thailand; 8) Division of Medical Diagnostics, Department of Clinical Laboratory, Fukujuji Hospital, Japan; 9) Department of Human Genetics, Graduate School of International Health, Tokyo, Japan.

Objectives: Genome wide association of tuberculosis in Asians identified a locus near *MAFB* associated with young tuberculosis (Mahasirimongkol, Yanai et al. 2012). Beijing or Modern lineages of *M.tuberculosis* is prevalent in young tuberculosis in Thailand, and genetic heterogeneity based on pathogen diversity in M. tuberculosis is a potential confounding factor contributing to inconsistent association evidences in genetic epidemiology of tuberculosis across populations. Global spatial distribution of *M.tuberculosis* lineages support plausible co-evolution with human populations(Gagneux 2012). Two major lineages of ***Beijing***(belong to Modern) and East African Indians (***EAI*** belong to Ancient) circulating in the Thai tuberculosis populations, allowing study of lineages based association in tuberculosis; we did genome wide association analysis based on two major *M.tuberculosis* lineages to determine the lineage specific association in tuberculosis.

Methods: *M. tuberculosis* were genotyped using the large sequence polymorphisms (LSP)-based PCR and direct repeats (DR)-based spoligotyping methods. The genome-wide genotyping was carried out with the Illumina 610 array following manufacturer protocol, data analysis was carried out with GenABEL. List of 242 TB candidate genes were retrieved from the HugeNet database. The association analysis specific to modern strains and ancient strains were carried out and reported.

Results: In this analysis, 201 Modern TB and 204 Ancient TB were available for the genome-wide association analysis. The comparisons were made against each group of TB, in each comparison; the additional 835 controls genetically matched samples were used as control group. For this analysis, QQ plots, Manhattan plots and regional association plots were analyzed and presented.

Conclusion: Candidate genes analysis supported genetic heterogeneity based on lineages of TB. The most interesting association from this analysis is the heterogeneity within MHC class II in TB. Among the top 50 SNPs, a locus associated with Ancient TB near Leukocyte receptor cluster (LRC) was identified, replication analysis of these findings are required for concrete confirmatory evidences.

865T

FTO genetic variants, dietary intake, and body mass index: results from 177,330 individuals. *Q. Qi[1], T. Kilpeläinen[2,3], M. Downer[1], T. Tanaka[4], C. Smith[5], T. Sørense[3], F. Hu[1], R. Loos[6], J. Nettleton[7], L. Qi[1], FTO Gene and Dietary Intake Consortium.* 1) Department of Nutrition, Harvard School of Public, Boston, MA; 2) MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, United Kingdom; 3) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 4) Translational Gerontology Branch, National Institute on Aging, Baltimore, Maryland, United States of America; 5) Nutrition and Genomics Laboratory, Jean Mayer USDA HNRCA at Tufts University, Boston, Massachusetts, United States of America; 6) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, Department of Preventive Medicine, Mount Sinai School of Medicine, New York City, N; 7) Division of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas Health Science Center, Houston, Texas, United States of America.

FTO is the strongest known genetic susceptibility locus for obesity. Experimental studies in animals suggest the potential roles of FTO in regulating food intake and energy expenditure. The interactive relation among FTO variants, dietary intake of energy-dense macronutrients, and body mass index (BMI) are complex and results from previous often small-scale studies in humans are highly inconsistent. We performed large-scale analyses based on cross-sectional data from 177,330 adults from 40 studies according to a standardized analytical plan to examine: 1) the association between the FTO-rs9939609 variant (or a proxy SNP) and total energy and macronutrient intake; and 2) the interaction between FTO variant and dietary intake on BMI. Macronutrient intake was expressed as the percentage of total energy intake. Dietary variables were dichotomized into two categorical variables based on the median intake in each study for the interaction analysis. The BMI-increasing allele of the FTO variant showed a significant association with higher dietary protein intake (effect per allele =0.08 [95% CI 0.06, 0.10]%, P=2.4×10-16), lower total energy intake (-6.4 [-10.1, -2.6]kcal/day, P=0.001), and lower dietary carbohydrate intake (-0.07 [-0.11, -0.02]%, P= 0.004). The associations of the FTO variant with total energy intake (P = 0.002) and protein intake (P=7.5×10-9) remained significant after adjustment for BMI. We did not find significant interactions between the FTO variant and dietary intake of total energy (P=0.25), protein (P=0.87), carbohydrate (P=0.80), or fat (P=0.13) on BMI. The magnitude of the association between the FTO variant and BMI was similar between participants in the low intake and the high intake groups of these dietary factors. In conclusion, the BMI-increasing allele of FTO seems to be associated with increased dietary protein intake. There is no obvious evidence supporting interactions between the FTO genetic variant and intakes of total energy or macronutrients in relation to BMI.

866F

Genome-wide environmental interaction (GWEI) analysis using multi-dimensional data reduction principles to identify asthma pharmacogenetic loci in relation to corticosteroid therapy. *F. Van Lishout[1,2], K. Bessonov[1,2], QL. Duan[3], K. Tantishira[3,4], K. Van Steen[1,2].* 1) Systems and Modeling Unit, Montefiore Institute, University of Liège, 4000 Liège, Belgium; 2) Bioinformatics and Modeling, GIGA-R, University of Liège, 4000 Liège, Belgium; 3) Channing Division of Network Medicine, Brigham & Women's Hospital and Harvard Medical School; 4) Pulmonary Division, Brigham & Women's Hospital and Harvard Medical School.

Genome-wide gene-environment (GxE) and gene-gene (GxG) interaction studies share a lot of challenges via the common genetic component they involve. GWEI studies may therefore benefit from the abundance of methodologies that are available in the context of genome-wide epistasis detection methods. One of these is Model-Based Multifactor Dimensionality Reduction (MB-MDR), which does not make any assumption about the genetic inheritance model. MB-MDR involves reducing a high-dimensional GxE space to GxE factor levels that either exhibit high or low or no evidence for their association to disease outcome. In contrast to logistic regression and random forests, MB-MDR can be used to detect GxE interactions in the absence of any main effects or when sample sizes are too small to be able to model all main and GxE interaction effects. In this ongoing study we demonstrate the opportunities and challenges of MB-MDR for genome-wide GxE interaction analysis and analyzed the difference in prebronchodilator FEV1 following 8 weeks of inhaled corticosteroid therapy, for 565 Caucasians from the Childhood Asthma Management Program (CAMP). In particular, we first followed standard marker quality control procedures. Missing genotypes were imputed with MaCH using 1000 Genomes Project Reference Panels (8,221,073 SNPs). Second, residuals were computed based on a polygenic regression model to correct for different types of relatedness and population stratification and to account for possible confounders such as age and height at baseline and gender. Third, LD pruning with maximum correlation threshold (r2) between markers of 0.5 was applied yielding 687,859 SNPs. Third, we curtailed MB-MDR to perform a GxE interaction analysis, with E a dichotomous variable coding for corticosteroid therapy (1: inhaled corticosteroids 0: other). Amongst our top results, we found 3 SNPs mapped to potentially interesting genes: a) rs12442193 mapped to SLCO3A1, a member of the solute carrier organic anion transporter family identified as participant of neurodifferentiation. b) rs2515471 mapped to ANGPT2 and MCPH1, related to the vascular remodeling and DNA damage repair. c) rs10898318 mapped to DLG2, a member of the membrane-associated guanylate kinase family involved in signal propagation in synapses and cell signaling. Follow-up of these results include replication in independent datasets and formal testing and validation of identified interactions as corticosteroid-related pharamacogenetic loci.

867W

Empirical comparison of Multi-association for Type 2 diabetes in Koean population-based cohort. *J. Lee[1], J. Lee[1], S. Park[1], B. Kim[1], B. Han[1], S. Won[2].* 1) Structural & Genomics, Centers Disease Control and prevention , #187 Osong sangmyung2ro, South Korea; 2) Department of Applied Statistics, Chung-Ang University, #5ga Anam-dong, South Korea.

Type 2 diabetes mellitus(T2D) is one of world-wide disease, its prevalence is above 10 percent in Korea. T2D is very highly correlated risk factors for epidemic and clinical, but is not sufficiently significant in Genome -wide association analyses. In genome wide association study, many Single nucleotide polymorphisms(SNPs) were significant for various risk factors, but these SNPs account for small fraction. We have constructed community and twin-family based on cohort, which is an ongoing prospective studies and surveyed samples were drawn from the Korean Genome and Epidemiology Study (KoGES) and Korea Genome Analysis Project (KoGAP) in Korea. From community-based cohort(KARE; Korea Association REsource), we selected 8,842 subjects who were in their various age(n=10,038) in this cohort and surveyed their self-report questionnaires about 1,400 items and genotyped using Affy 5.0. The aim of this study was to estimate their multi-association(multi trait-multi genetic information). These were estimated and optimized Quantitative genetic analysis adjust age and sex. Then we compared empirically association and multi-association results.

868T

Influence of physical activity on body mass index in relation to well-replicated obesity loci in African-American adults: The ARIC Study.
K.L. Young[1, 2], A.E. Justice[2], E.W. Demerath[4], W-H.L. Kao[5], E. Boerwinkle[6], K.E. North[2,3], K.L. Monda[7]. 1) Carolina Population Center, University of North Carolina, Chapel Hill, NC; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 4) Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; 5) Department of Epidemiology, Johns Hopkins University, Baltimore, MD; 6) Division of Epidemiology, Schoool of Public Health, University of Texas Health Science Center, Houston, TX; 7) Center for Observational Research, Amgen, Inc, Thousand Oaks, CA.

Obesity is a known risk factor for many chronic diseases, and physical activity (PA) is often used to lower those risks by reducing weight. However, the impact of PA on obesity exhibits considerable variation, which is likely influenced by genetics. African Americans (AA) are at particularly high risk of obesity, but remain understudied. We investigate the modifying influence of PA on 24 BMI loci which generalize or for which a better SNP in the same locus has been reported in AA using the Atherosclerosis Risk in Communities (ARIC) Study, a population-based sample of adults aged 45-64 years at baseline. Our analyses were restricted to genotyped individuals of self-reported African descent (N=2,373). BMI (kg/m2) was computed from measured height and weight. PA was measured using a modified Baecke questionnaire, resulting in an ordinal score (range: 1.0-10.0). We assumed an additive genetic model, regressing BMI on SNP, PA level, SNP by PA, sex, age, current smoking, education, center, alcohol consumption, and principal components (PCs) to account for ancestry. Additionally, we performed main effects analyses controlling for sex, age, current smoking, center, and PCs. The majority of our sample was female (63%), with a mean age of 53.3 (SD 5.7) years, and at least a high school education. Mean BMI was 29.7 (SD 6.0) kg/m2, and PA was 6.98 units (SD 1.19, range: 2.0-9.0). SNPs were distributed on 15 of the 22 autosomes and ranged in minor allele frequency from 5.8% (PTBP2) to 42.4% (MIR148A/NFE2L3). In the main effects analysis, seventeen of the 24 loci were directionally-consistent with earlier findings. While none of the effect modification p-values were significant after Bonferroni correction ($p<0.002$), MAP2K5 reached suggestive significance ($p<0.05$). Each additional copy of the risk allele in MAP2K5 resulted in an increase of 0.04 kg/m2 for low PA individuals, but a decrease of 0.12 kg/m2 for high PA individuals (p-interaction=0.045). Mean BMIs for those with low PA increase with additional copies of the G (risk) allele (mean BMI for AA=30.8, AG=31.5, GG=31.1 kg/m2). Ten of 24 SNPs were directionally consistent with the expected effect modification with the effect of the BMI increasing allele decreasing in the context of high PA. Overall, as has been reported in European descent populations, our results suggest higher levels of physical activity may modify the influence of selected obesity susceptibility variants on BMI in African Americans.

869F

Transcriptional response to acute respiratory virus infection - a prospective cohort study.
Y. Zhai[1,4], J. Belmont[1,4,6], R. Atmar[2,3], J. Quarles[5], N. Arden[5], K. Bucasas[1], J. Wells[3], D. Niño[3], X. Wang[4], G. Zapata[4], C. Shaw[1], L. Franco[1,2], R. Couch[3]. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX; 4) Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX; 5) Department of Microbial and Molecular Pathogenesis, Texas A&M University System Health Science Center, College Station, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX.

Objectives: Acute viral respiratory infections (ARIs) are responsible for a large number of outpatient visits and hospitalizations in the U.S., and the threat of pandemic influenza will likely add to these numbers. A new influenza A/H1N1 (pH1N1) virus, which emerged in April 2009, spread worldwide and continued to circulate in the following year. In this study we aim to identify functional gene networks and pathways that are correlated with immune response to influenza infection by time-course analysis of the host transcriptional profiles. Methods: 1618 healthy adults, enrolled in fall 2009 and 2010, were followed for acute viral respiratory disease. Subjects reporting moderate to severe acute respiratory illness had virus quantitation for 3 weeks. Peripheral blood samples for RNA extraction were obtained on the first day of illness (day 0) and then on 2, 4, 6, and 21 days after first symptoms. In addition to an RNA sample obtained at enrollment, a final RNA sample was obtained at the end of the study. RNA samples were analyzed using expression microarrays and the patterns of gene expression were analyzed to identify differentially expressed transcripts. Result: Among the 133 subjects who suffered viral respiratory and completed 5 illness day study visits, 73 had influenza infection, 64 influenza A and 9 influenza B. Of these, 24 had at least one other known viral respiratory pathogen e.g. human rhinovirus (HRV) or respiratory syncytial virus (RSV). Gene expression profiles showed highly similar patterns among the major subgroups. There was a dramatic up-regulation of interferon pathway and innate immunity on the first day of infection. This usually persisted for 2 days. A convalescent phase was observed on days 4 and 6 after infection. By day 21 the gene expression pattern had returned to levels indistinguishable from baseline. Using lineage and activation state specific transcripts to produce cell decomposition scores, patterns of acute depression of the B and T lymphocytes were observed accompanied by the evidence of dramatic activation of dendritic cells, NK cells, and, to a lesser extent, activated T cells. Conclusions: Transcriptional profiling gives a genome wide view of a coordinated systemic response to acute viral respiratory infection. There are two clear phases of gene expression, corresponding to intense activation of innate immunity pathways followed by a convalescent phase marked by cell proliferation and repair.

870W

Insights into the molecular arms race between the Malaria parasite and its human host from genomic analysis of over 15,000 African individuals.
C.C.A. Spencer[1], L.S. Quang[1], G. Band[1], K. Rockett[1], D. Kwiatkowski[1,2], MalariaGEN. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

The continuing burden of death and illness that malaria inflicts in endemic regions underlines its likely important role in human evolution. To gain insights into the key molecular determinants of human susceptibility to severe malaria we studied over 15,000 individuals from eight countries across sub-Saharan Africa, using genome-wide SNP genotyping, as part of the MalariaGEN consortium (www.malariagen.net). We use these data, in combination with novel statistical methodology, to jointly describe patterns of association and natural selection across the genome. By systematically coupling signals of association with models of recent evolution across genes of different function we can test specific hypotheses. We assess the evidence for two broad classes of protective effects: the first, protective alleles that are maintained under balancing selection through deleterious effect (for example sickle cell disease) which have consistent effects across populations; the second, protective alleles that are under frequency-dependent selection, and are expected to have different effects on susceptibility across populations, due to an interaction with parasite diversity. These new observations reinforce the role of red blood cell surface proteins as important in severe malaria susceptibility, putatively by inhibiting parasite invasion mechanisms. In addition, our genome-wide data can be used to assess the relationship between recently reported shared human and chimpanzee polymorphisms and infectious disease caused by the malaria parasite. As well as helping to quantify the impact of malaria susceptibility alleles on human diversity in Africa, we are also able to use signals of natural selection to aid identification of new susceptibility loci.

871T

The *MTHFD1* 1958 G>A (R653Q) variant is associated with elevated C-reactive protein and body mass index in a Canadian premature birth cohort. *K.E. Christensen*[1,2], *M. Dahhou*[2], *M.S. Kramer*[2,3], *R. Rozen*[1,2]. 1) Dept. of Human Genetics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 2) Dept. of Pediatrics, McGill University, McGill University Health Centre, Montreal, Quebec, Canada; 3) Dept. of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada.

MTHFD1 1958 G>A (rs2236225) is a non-synonymous variant in the synthetase domain of the trifunctional MTHFD1 enzyme. It reduces stability and impairs *de novo* purine synthesis. *In vitro* data suggest that the variant's effects may be modulated by folate. This SNP has been associated with increased risk for birth defects (heart and neural tube) and pregnancy loss, particularly in the Irish population, which is not folate fortified. We have recently generated a mouse with partial synthetase deficiency (*Mthfd1S*$^{+/-}$) as a model for this variant. The majority of the mice were healthy and normal; however, the rate of developmental defects was higher in offspring of *Mthfd1S*$^{+/-}$ dams. In addition, nonpregnant female *Mthfd1S*$^{+/-}$ mice were heavier than their wild-type littermates, and white blood cell counts (specifically neutrophils) were reduced, particularly during pregnancy. To examine immune function and weight in women with the 1958 G>A SNP, we genotyped 206 women with spontaneous preterm births and 443 controls from a large multicentre cohort of Quebec women. There was no association with prematurity, with or without stratification by median plasma folate (29.6 nM). In grouped cases and controls, plasma C-reactive protein (CRP, an immune marker associated with low-grade inflammation) was increased due to the A allele (p=0.053, ANOVA; p for trend = 0.041) in women with folate < median. Using logistic regression of high versus low CRP including genotype, maternal age, smoking, BMI and education in the model, CRP was increased in GG compared to AA women with folate < median (p=0.055). There was also a strong association between pre-pregnancy BMI >= 30 and CRP in this analysis, both above and below median folate (p<0.001). A nonsignificant dose-response relationship was observed between mean BMI and genotype, especially in the low-folate group, when analyzed by ANOVA. However, there was a significant association between the 1958G>A variant and BMI, when examined for low (<18.5), normal (18.5-25), overweight (25-30) and obese (>=30) groups by chi-square analysis of combined cases and controls (p=0.0113). These results suggest that the *MTHFD1* 1958G>A variant may influence immune function and obesity, particularly if folate intake is low. Additional cohorts should be evaluated, particularly since the median folate level in this cohort was quite high.

872F

Analysis of APOE-e4 and Facial Profile Convexity in Obstructive Sleep Apnea. *J.K. Hartsfield*[1,2,3], *J.J. Roedig*[1], *B.A. Phillips*[2], *L.A. Morford*[1,3], *J.E. Van Sickels*[1], *G. Falcão-Alencar*[3], *D.W. Fardo*[4], *G.T. Kluemper*[1]. 1) College Dentistry, Univ Kentucky, Lexington, KY; 2) College of Medicine, Univ Kentucky, Lexington, KY; 3) Center for Oral Health Research, Univ Kentucky, Lexington, KY; 4) College of Public Health, Univ Kentucky, Lexington, KY.

Objective: The inheritance of obstructive sleep apnea (OSA) is complex. Previously, APOE-e4 and increased facial convexity (mandibular retrognathism, i.e., a Class II (CII) sagittal skeletal jaw relationship) have independently been associated with OSA. We hypothesized that in OSA there is an association between APOE-e4 and non-CII compared to CII subjects. Associations between APOE-e4 with BMI, and Apnea-Hypopnea-Index (AHI); and jaw relationship with BMI and AHI were also studied. Materials and Methods: Seventy-six Caucasian OSA subjects with an AHI > 15 were classified into different sagittal skeletal jaw relationships by oral exam and profile photos. In addition to CII, Non-CII includes Class I with straight profile, and Class III with mandibular prognathism. DNA via saliva was extracted. Two SNPs (rs429358 and rs7412) were genotyped to determine APOE allele(s). Chi-square analysis assessed Hardy-Weinberg-Equilibrium (significance at p<0.05). Fisher's exact test was used to look for association between APOE-e4 status and the skeletal classification of OSA subjects. ANOVA compared BMI and AHI among the skeletal types. Results: Twenty-five CII subjects (mean BMI 30.7, AHI 33.6) and fifty-one non-CII subjects (mean BMI 37.4, AHI 44.2)were examined. Seventy subjects were obese, and six subjects had a normal BMI. All six individuals with a normal BMI were CII. APOE-e4 was not associated with different sagittal facial profiles. CII subjects had significantly lower BMI than CI 37.2 or CIII 37.8 subjects (p<.001). There was no difference among the facial profile groups for AHI (CII 33.6, CIII 39.5, CI 46.0). There was a significant difference between CII and CI for AHI. Conclusion: Patients who have OSA and CII have a lower BMI, and 1/5 are not even obese, suggesting that CII is a contributing factor to OSA with or w/out increased BMI. No association between non CII and APOE-e4 was found in this preliminary study.

873W

Smoking and asthma: effect modification by phospholipase C, beta 1 and peroxiredoxin 6 polymorphisms. *I. Ahmed*[1,2], *M. Rava*[2,3], *M-H. Dizier*[4,5,6], *V. Siroux*[7,8], *M. Kogevinas*[9], *N. Probst-Hensch*[10,11], *F. Demenais*[4,5,6], *P. Tubert-Bitter*[1,2], *R. Nadif*[2,3]. 1) Biostatistics Team, Inserm U1018, Villejuif, France; 2) Univ Paris-Sud; 3) Respiratory and Environmental Epidemiology Team, Inserm U1018, Centre for Research in Epidemiology and Population Health, Villejuif, France; 4) INSERM UMRS-946, Paris, France; 5) Institut Universitaire d'Hématologie, Univ Paris Diderot, France; 6) Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain, Paris, France; 7) Environmental epidemiology applied to reproduction and respiratory health Team, Inserm U823, Grenoble, France; 8) Univ Joseph Fourier, Grenoble, France; 9) Centre for Research in Environmental Epidemiology (CREAL), Barcelona, France; 10) Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute Swiss TPH, Switzerland; 11) University of Basel, Switzerland.

The number of biological proofs of a role of oxidative/nitrosative stress in asthma is increasing. Tobacco smoke, which plays a role in asthma, is an environmental factors related to oxidative/nitrosative stress. Our goal was to identify polymorphisms (SNPs) in genes involved in the response to oxidative/nitrosative stress interacting with smoking on asthma. This study takes place in the framework of the French Epidemiological study of the Genetic and Environmental factors of Asthma (EGEA), which combines a case control study and a family study. 182 genes (5498 SNPs) were selected according to a pathway-based strategy integrating biological knowledge related to tobacco smoke metabolism. Tobacco smoke was expressed as current, ex- and non-smokers. SNP discovery was performed by combining marginal genetic associations and gene by smoking interactions as proposed by Dai (Dai, AJE, 2012). Both the marginal associations and gene interactions were estimated using GEE logistic regression models, adjusted for age, sex and principal components for population ancestry in 1570 adults (43 years old, 36.3% with active asthma, 24.8% smokers). 6 SNPs (5 in PARK2 (Parkinson protein 2) and 1 in PRKCA (protein kinase C, alpha)) had Dai P values lower than 1×10-3. All of them showed marginal associations with active asthma (p<5×10-4) but no evidence for interaction with tobacco smoke. Considering SNPs showing suggestive evidence for interaction with tobacco smoke revealed 14 signals (p<1×10-3) in PLCB1 (phospholipase C, beta 1), PRDX6 (peroxiredoxin 6) and NFE2L2 (nuclear factor (erythroid-derived 2)-like 2), a transcription activator of antioxidant genes. Although replication of our results is needed, it is interesting to note that oxidized phospholipids are associated to impair function of macrophages in mice exposed to chronic cigarette smoke (CS), and that chronic 6 months of CS exposure resulted in increased lung inflammatory response in prdx6 -/- mice.

874T

Pathogenic Role of MicroRNA in Pediatric Asthma. *X-Y. Dong*[1, 2], *N. Zhong*[1,2,3]. 1) Children's Hospital of Shanghai, Jiaotong University, Shanghai, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, China; 3) New York State Institute for Basic Research in DD, Staten Island, NY.

Pediatric asthma is one of the most common children's respiratory diseases. The morbidity of asthma is stably increasing worldwide. Presently the etiology and pathogenesis of asthma are yet unclear but complicated. Hypothesis of that asthma is resulted from interactions of genetics, immune regulation, and environmental factor(s) has been proposed. This has led us to explore the epigenetic mechanism underlying pediatric asthma. For which, we have undertaken a stud of epigenomic regulation of genome wide microRNAs (miRNA) for the pediatric asthma. miRNA is one of the non-coding small RNAs, with a length of 21-25 nucleotides. Mature miRNA may bind to 3' UTR (or other regions) of target mRNA and result in degradation or translational repression of mRNA during development, cell proliferation, variation, and apoptosis. Recent studies determined that miRNA plays an important role in the pathogenesis of adults' asthma. We therefore proposed that miRNA might also involve in the pathogenesis of pediatric asthma. In out pilot study, 12 cases of children's asthma that had been confirmed with dust mite allergy, compared to 12 cases of age- and gender-matched normal children, were applied to Exiqon's genome wide microarray assays. Our preliminary results showed three miRNAs, two (miRNA 5588-5p, miRNA 4255) up-regulated and one (miRNA 27b-3p) down-regulated, are associated with inflammatory signaling pathway with impact on regulating mRNA expression of IL-6, -10, -13, TNF and INF. Functional analysis is being undertaken to understand how the miRNA regulates the mRNA expression within the inflammatory signaling pathway. Our study has opened a new avenue to uncover the molecular pathogenic mechanism underlying pediatric asthma.

875F

Occupational exposures to potential irritants and asthma: effect modification by glutathione S-transferase Z1 and ATP-binding cassette transporters polymorphisms. *M. RAVA[1, 2], I. AHMED[2,3], O. DUMAS[1,2], M. KOGEVINAS[4], N. PROBST-HENSCH[5,6], P. TUBERT-BITTER[2,3], N. LE MOUAL[1,2], F. DEMENAIS[7,8,9], R. NADIF[1,2].* 1) INSERM U1018, Centre for research in Epidemiology and Population Health (CESP), Respiratory and Environmental Epidemiology Team, Villejuif, France; 2) Univ Paris-Sud 11, UMRS 1018; 3) INSERM U1018, Centre for research in Epidemiology and Population Health (CESP), Biostatistics Team, F-94807, Villejuif, France; 4) Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; 5) Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute Swiss TPH, Switzerland; 6) University of Basel, Switzerland; 7) INSERM, UMRS-946, F-75010, Paris, France; 8) Univ Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, F-75007, Paris, France; 9) Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), Paris, F-75010, France.

Occupational asthma (OA) is a good model to study the pathophysiology of asthma. Around 15% of adult asthma would be caused by occupational exposures, and more than 400 distinct agents have been identified as causing OA. The mechanisms of asthma induced by potential irritants, such as cleaning and chemical products, are unclear but could be related to oxidative/nitrosative stress. Our goal was to identify polymorphisms (SNPs) in genes involved in the response to oxidative/nitrosative stress interacting with occupational exposures known to have an effect on asthma, in the familial case-control French Epidemiological study of the Genetic and Environmental factors of Asthma (EGEA, https://egeanet.vjf.inserm.fr/). A large set of genes (162 genes, 4979 SNPs) was selected according to a pathway-based strategy, that integrates biological knowledge related to occupational exposures. Occupational exposure to potential irritants was evaluated using the asthma job exposure matrix (Asthma JEM, http://cesp.vjf.inserm.fr/asthmajem/) and job-specific questionnaires for cleaners and healthcare workers. First, the marginal association between each SNP and current asthma, and the SNP by occupational exposures interactions were estimated separately using GEE logistic regression models, adjusted for age, sex and principal components for population ancestry in 1254 adults (45 years old, 38% with current asthma, 32% exposed). Then, the marginal genetic associations and the gene by environment interactions were simultaneously tested as proposed by Dai (Dai, AJE, 2012). 17 SNPs had Dai P values lower than 1×10-3. Among these SNPs, one in ABCC1 (ATP-binding cassette, sub-family C (CFTR/MRP), member 1) and one in GSTZ1 (glutathione S-transferase Z1) showed suggestive evidence for interaction with occupational exposures (p= 2×10-4 and 7×10-4 respectively). In addition, SNPs in PARK2 and NDUFS4 showed association signals with current asthma (p<1×10-3). Although replication of our results is needed, it is interesting to note that GSTZ1 is a member of the GSTs super-family which encodes multifunctional enzymes important in the detoxification of electrophilic molecules by conjugation with glutathione, and that ABCC1 mediates ATP-dependent transport of glutathione and glutathione conjugates, and other xenobiotics.

876W

Comparative Analysis of T2D Risk Allele Load in African Americans and European Americans. *J.M. Keaton[1,2,3], J.N. Cooke Bailey[2,3,4], N.D. Palmer[2,3,7], B.I. Freedman[5], C.D. Langefeld[6], M.C.Y. Ng[2,3], D.W. Bowden[2,3,7,8].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Molecular Medicine and Translational Science, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC; 6) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 7) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 8) Internal Medicine-Endocrinology, Wake Forest School of Medicine, Winston-Salem, NC.

The burden of type 2 diabetes (T2D) is greater in populations of African descent compared to other ethnicities. In the United States prevalence of T2D is 12.6% among African American adults compared to 7.1% for European Americans. Although lifestyle and environmental factors contribute to T2D risk, genetic contributors to T2D disparities have not been well documented. Genome-wide association studies (GWAS) have identified ≥60 common genetic variants that contribute to T2D risk in populations of European, Asian, African, and Hispanic descent. Previous studies have not comprehensively examined population ancestry-specific differences in cumulative risk allele load. Genetic risk load was analyzed in 1,990 African Americans (n=963 T2D cases, n=1027 controls) and 1,654 European Americans (n=726 T2D cases, n=928 controls) ascertained and recruited using a common protocol in the southeast US. 49 T2D index single nucleotide polymorphisms (SNPs) from 45 loci from GWAS in European, East Asian, South Asian, and African derived populations were analyzed. In African American subjects, risk allele frequencies ranged from 0.052 to 0.976. Risk alleles from 28 SNPs demonstrated directional consistency with previous studies, and 4 SNPs from *ADAMTS9*, *TCF7L2*, *INS-IGF2*, and *ZFAND6* showed evidence of association (P<0.05). African American individuals carried 38-69 risk alleles, and had an average risk allele load of 55. In European American subjects, risk allele frequencies ranged from 0.001 to 0.996. Risk alleles from 37 SNPs demonstrated directional consistency, and 10 from *BCL11A*, *PSMD6*, *ADAMTS9*, *ZFAND3*, *ANK1*, *CDKN2A/B*, *TCF7L2*, *PRC1*, *FTO*, and *BCAR1* showed evidence of association (P<0.05). European American individuals carried 38-65 risk alleles, and had an average risk allele load of 51.African Americans have a highly significant mean increase of 4.3 risk alleles ($P=1.6\times10^{-177}$; two sided t-test, two samples with unequal variances) compared to European Americans. This cumulative increase in T2D risk allele load may contribute to the observed disparity in disease prevalence between population ancestries.

877T

Melanocortin 1-receptor haplotypes associated to pigmentation phenotypes in the brazilian population. *L.A. Marano[1], A.L. Simoes[1], E.A. Donadi[2], C.T. Mendes-Junior[3].* 1) Departamento de Genetica, FMRP, Universidade de Sao Paulo, Ribeirao Preto, SP, Brazil; 2) Departamento de Clinica Medica, FMRP, Universidade de Sao Paulo, Ribeirao Preto, SP, Brazil; 3) Departamento de Quimica, FFCLRP, Universidade de Sao Paulo, Ribeirao Preto, SP, Brazil.

The MC1R (melanocortin 1-receptor) gene is one of the best characterized so far among the known genes influencing eye, skin and hair pigmentation variation. This gene activity happens due the coding of a transmembrane protein in melanocytes, responsible for regulating the production of melanin. MC1R importance relies on the determination of the ratio of eumelanin (brown/black) and pheomelanin (yellow/red) pigments present in melanocytes, accordingly to its higher or lower signaling activity. This study aims to use MC1R data obtained from 29 known SNPs in order to reconstruct haplotypes and evaluate their diversity influence on features like freckles and pigmentation variation of eyes, skin and hair in humans. We analyzed 29 known SNPs in the coding region of MC1R gene in 296 individuals from the region of Ribeirao Preto, Brazil. DNA extraction was performed using the salting-out technique. The MC1R gene coding region (951pb) was amplified in a single PCR reaction, which was sequenced on a ABI PRISM-310 genetic analyzer by capillary electrophoresis, using the same primers used for amplification. From the 29 SNPs evaluated, only 22 showed variation in the samples studied, half of them showing to be associated with pigmentation characteristics. Haplotype reconstruction generated 31 haplotypes. Four of them were associated with dark skin and other two had significantly high frequencies in fair skin. One haplotype was associated with green eyes, while two other had association with dark brown eyes. Darker hair color was associated with six different haplotypes, whereas red hair was associated with two and blonde hair with one haplotype. Finally, the absence or presence of freckles was significantly related to three haplotypes. This study shows substantial associations between MC1R haplotypes and external visible traits, confirming that this gene also plays a relevant role in the pigmentation variation in the Brazilian population.

878F

The IRF5-TNPO3 association has two components in systemic lupus erythematosus (SLE), which are shared with other autoimmune disorders. *L.C. Kottyan[1,2], E.Z. Zoller[1], K.M. Kaufam[1,2], J.B. Harley[1,2].* 1) Center for Autoimmune Genomics and Etiology, Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; 2) Cincinnati Veterans Affairs Med. Ctr., Cincinnati, Ohio, USA.

Recent improvements in genetic methods allow for more complete identification of candidate causal variants. Exploiting genotyping, DNA sequencing, imputation, and trans-ancestral mapping, we modeled the IRF5-TNPO3 genetic association on chromosome 7, now implicated in two immunotherapies and seven autoimmune diseases. We capture 7,650 IRF5-TNPO3 genetic variants across five ethnicities in 8,395 systemic lupus erythematosus (SLE) cases and 7,367 controls. We resolve separate lupus-risk associations in the IRF5 promoter (all ancestries) and an extended European haplotype- strongly associated with lupus in subjects of European decent but non-polymorphic in subjects from Asia or Sub-Saharan Africa. Adjusting for two variants tagging each genetic effect removes the association of other variants in the region. Using both frequentist and Bayesian approaches, we identify the mostly likely causative variants to three IRF5 promoter variants (confined to 5.7 kb) and 22 variants of the 85.5 kb European haplotype that spans IRF5 and TNPO3. The possible statistical models from this sample virtually eliminate the previously purported IRF5 functional variants as causal. Strikingly, this model also appears to operate in Sjögren's syndrome and systemic sclerosis (both components) and primary biliary cirrhosis (haplotype only), demonstrating the nuances of similarity and difference in autoimmune disease risk mechanisms at IRF5-TNPO3. Finally, we performed case-only subphenotypic analyses to identify variants in the IRF5-TNPO3 region useful in predicting the disease progression of patients with lupus and found strong association with age-of-onset (variants in the haplotype) and the antigen specificity of autoantibodies anti-Ro and anti-dsDNA (variants in the IRF5 promoter). In conclusion, genetic variants in the IRF5-TNPO3 region critically regulate the development of multiple autoimmune diseases and the clinical presentation of lupus. Given the importance of this locus to multiple immune disorders, future studies will be aimed at understanding the mechanisms driving the genetic associations. (72 other collaborators- providers of samples, genetic, analytical, and financial resources will be provided at presentation.).

879W

Epiregulin (*EREG*) and human V-ATPase (*TCIRG1*) are not associated with pulmonary tuberculosis in West African populations from Guinea-Bissau and The Gambia. *S.M. Williams[1,2], M.J. White[1,2], A. Tacconelli[3], C. Wejse[4], P.C. Hill[5], G. Novelli[3], G. Sirugo[3].* 1) Department of Genetics, Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Centro di Genetica, Ospedale San Pietro FBF, Rome, Italy; 4) Bandim Health Project, Danish Epidemiology Science Centre and Statens Serum Institute, Bissau; 5) Centre for International Health, University of Otago School of Medicine, New Zealand.

Tuberculosis (TB) is caused by *Mycobacterium tuberculosis* (M. tuberculosis) and infects approximately one third of the world's population, with the majority of infection occurring in Africa and South-east Asia. Active TB, of which pulmonary TB (PTB) accounts for approximately 80%, has one of the highest mortality rates for an infectious disease. There is a large discrepancy between latent TB infection and incidence of active TB, with only a small percentage of those infected developing active disease, suggesting genetic factors influence host response to *M. tuberculosis* infection. Recent studies have identified two novel TB susceptibility candidate genes, *EREG* (Epiregulin) and *TCIRG1* (human ATPase, a3 isoform), in Vietnamese and Italian populations, respectively. The main aim of this study was to validate the associations between *EREG* and *TCIRG1* with PTB in 'at risk' populations from West African endemic areas. Using a discovery cohort from Guinea Bissau (n = 289 cases, 322 controls) and a replication cohort from the Gambia (n = 240 cases, 248 controls), we assessed association with 14 SNPs in *EREG* and *TCIRG1*; using single and multi-locus analysis to determine whether previously described associations generalized to African populations. Logistic regression analysis revealed one SNP in *TCIRG1*, rs10896289, as significantly associated with PTB (p=0.042) and another SNP in *EREG*, rs1563826, that was marginally associated with PTB (p= 0.056) in our discovery cohort; these associations were not observed in the Gambian replication cohort. Haplotype analyses were also performed to investigate associations with PTB; no significant results were found in either cohort. We also performed multi-factor dimensionality (MDR) analysis to elucidate possible interaction effects, between SNPs and demographic factors. Interestingly, although MDR identified the same variables in both our discovery and replication cohorts, further investigation revealed that the risk models differed. In conclusion, despite being adequately powered to detect the effect sizes previously reported, our single and multi-locus analyses were unable to validate *EREG* and *TCIRG1* as PTB susceptibility candidate genes in West African populations. Our findings suggest that associations of PTB with genetic variation in these genes may be influenced by inter-population differences in genetic or environmental context and/or the mycobacterial lineages that cause active disease.

880T

Exploring genetic load of known multiple sclerosis risk alleles in Hispanic whites. *J. McCauley[1], A. Hadjixenofontos[1], C.P. Manrique[1], A.H. Beecham[1], I. Konidari[1], P.L. Whitehead[1], P-A. Gourraud[2], M.A. Pericak-Vance[1], L. Tornes[3], M. Ortega[3], K.W. Rammohan[3], S.R. Delgado[3].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, School of Medicine, University of California at San Francisco, San Francisco, CA, USA; 3) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA.

Multiple Sclerosis (MS) is a demyelinating disease with autoimmune etiology and variable expression across populations. Studies on the genetic susceptibility to MS have primarily focused on Northern European (NE) populations, largely due to the increased prevalence in these countries. However, Hispanic patients are underrepresented in well-powered studies of disease presentation, and even more so in investigations into the underlying genetic architecture. Our previous studies have revealed potentially important clinical differences between Hispanic white and non-Hispanic white patients that warrant further investigation. As a first step we have calculated an MS risk score to capture the genetic load attributed to variants identified in NE populations, therefore we hypothesize that an admixed population will be characterized by a partly-overlapping set of genetic variants. The most recent study by the International MS Genetics Consortium has significantly increased the list of known MS risk variants while also fine-mapping previously discovered signals. We use 113 of these variants to construct a weighted MS genetic risk score. Our goal is to assess the extent to which the NE derived MS risk score is able to distinguish Hispanic white cases (n = 186) from non-Hispanic white cases (n = 161). Preliminary results suggest that the genetic load due to the variants included in the MS risk score is similar in Hispanic white cases (mean = 12.94, SD = 0.75) and non-Hispanic white cases (mean = 12.91, SD = 0.79) (t-test p-value = 0.67). A more comprehensive examination of the variation present in the known regions using information on local ancestry is underway. Our study highlights the need for further studies of MS risk factors among Hispanics with the goal of understanding the causes of the differences we see in MS presentation in this population.

881F

Genetic Basis of Height and Skin Pigmentation in Southern Africa. *B.M. Henn[1,2], J.M. Granka[3], A.R. Martin[2], C.R. Gignoux[4], M. Lin[1], J.M. Kidd[5], E.G. Hoal[6], M.W. Feldman[2], C.D. Bustmante[1].* 1) Dept of Ecology and Evolution, Stony Brook University, Stony Brook, NY; 2) Dept. of Genetics, Stanford University, Stanford CA; 3) Dept. of Biology, Stanford University, Stanford, CA; 4) University California, San Francisco, CA; 5) Dept. of Human Genetics, University of Michigan, Ann Arbor, MI; 6) Molecular Biology and Human Genetics, Stellenbosch University, Tygerberg, South Africa.

No two traits are as immediately recognizable for their variability among and within human populations as height and skin pigmentation. Many evolutionary hypotheses have been proposed to explain their variability, but relatively little is known about the genetic basis of these traits across human populations. In particular, the genetic basis of these phenotypes in the KhoeSan populations of southern Africa remain completely unknown. For the first time, we analyze the heritability of height and skin pigmentation in over 200 ≠Khomani Bushmen and Nama individuals from South Africa, and perform genome-wide association analyses for these traits with both high-coverage exome and SNP array data. We contrast estimates of heritability for height and skin pigmentation and find that most variation in innate skin pigmentation is potentially explained by common single nucleotide polymorphisms (SNPs); height is only moderately heritable and tanning is largely environmental. While varying amounts of European, west African and KhoeSan ancestry strongly correlate with skin pigmentation, ancestry does not appear to contribute significantly to variation in height in our samples. Finally, after controlling for kinship relationships in our data, we identify several large effect loci associated with height located near genes previously identified to be associated with body mass index/height (e.g. SLC1A3, VWA8, DGKH) Our results emphasize that analyses of phenotypically and genetically diverse endogamous populations have the potential to reveal novel insights into the genetic basis and evolutionary history of complex traits.

882W

Explicit modeling of genetic ancestry improves polygenic prediction accuracy. *C. Chen[1], J. Han[1,3,4], D. Hunter[1,4,5], P. Kraft[1,2,4,5], A. Price[1,2,5].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Department of Dermatology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 5) Broad Institute of Harvard and MIT, Cambridge, MA.

Polygenic prediction using genome-wide SNPs can provide higher prediction accuracy than prediction using only known associated SNPs. It is widely known that the prediction accuracy of polygenic model in structured populations may be partly due to genetic ancestry. However, we hypothesized that explicitly modeling ancestry could improve polygenic prediction accuracy, by providing unbiased estimates for the polygenic component while incorporating associations between ancestry and phenotype as a separate component in the prediction model. This prevents ancestry effects from entering into each SNP effect and being over-weighted. We analyzed three large GWAS of hair color, tanning ability and basal cell carcinoma (BCC) in European Americans (sample sizes: 7,400 to 9,800) to demonstrate our approach, restricting to a set of 70,000 independent SNPs. We compared polygenic prediction without correction for ancestry to polygenic prediction with ancestry as a separate component in the model, using the top 5 principal components to model ancestry. In 10-fold cross-validation, the R^2 for hair color increased by 66% (from 0.0456 to 0.0755; p-value $< 10^{-12}$) when explicitly modeling ancestry in the polygenic prediction model. Similarly, the R^2 for tanning ability increased by 123% (from 0.0154 to 0.0344; p-value $< 10^{-12}$) and the liability-scale R^2 for BCC increased by 68% (from 0.0138 to 0.0232; p-value $< 10^{-12}$) when explicitly modeling ancestry in the polygenic model. For each of the three traits, an improvement in prediction accuracy remains when including known associated SNPs in the polygenic model. In summary, our results show that explicitly modeling ancestry can be important in polygenic prediction.

883T

Powerful detection of osteoarthritis susceptibility loci by comprehensive examination of clinically important endophenotypes. *K. Panoutsopoulou[1], S. Thiagarajah[2], A.G. Day-Williams[1,3], L. Southam[1], K. Hatzikotoulas[1], A. Matchan[1], M. Doherty[4], J.M. Wilkinson[2], E. Zeggini[1], arcOGEN Consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Academic Unit of Bone Metabolism, Department of Human Metabolism, University of Sheffield, Sheffield, United Kingdom; 3) Biogen Idec, Cambridge, MA, USA; 4) Academic Rheumatology, Nottingham City Hospital, Nottingham, UK.

Osteoarthritis (OA) is a highly heterogeneous disease characterised by variable clinical features with possibly different genetic aetiologies. Thus far, the few genetic variants that have been robustly associated with OA (n=13 in Europeans) explain only a small proportion of its heritability. Studies performed to date have used very broad phenotypic definitions of OA which do not take into account the differences in pattern of joint involvement between individuals that may represent different physiological processes, and may thus have different genetic aetiologies. In a disease like OA the use of expanded and tighter phenotype definitions closer to the biology of the disease is likely to significantly enhance power to identify associations. We have been part of the largest genome-wide association study for OA (arcOGEN GWAS) and have previously established 9 OA associated loci using broad definitions of OA. Here we have comprehensively analysed an expanded set of narrower, clinically relevant OA endophenotypes derived from radiographs of 2,000 knee and 2,000 hip OA cases from arcOGEN. Variables studied relate to joint morphology, specific anatomic pattern of joint involvement, severity and bone response. Following 1000 Genomes Project-based imputation and stringent quality control, >7 million variants were tested for association with each phenotype. Our results indicate that the study of endophenotypes in OA has the potential to dramatically enhance power to detect OA-relevant associations. For example analysis by knee compartment involvement vs population-based controls yielded 25 independent loci for knee OA at $p<1\times10^{-6}$ vs 1 locus detected for knee OA vs controls in the equivalent binary trait GWAS. In hip OA endophenotype analyses several promising signals were identified some of which are found near genes that are very plausible biological candidates for OA: in the analysis of atrophic vs hypertrophic hip OA response a strong signal (OR[95% CI]=2.03[1.57-2.63], $p=2.5\times10^{-8}$) was detected in *GPR98*. *gpr98* knockout mice have a low bone mass phenotype, and a *GPR98* polymorphism has recently been associated with osteoporotic fracture in Japanese women. Pattern of hip migration shows strong association with variants in *LRCH1* ($p=2.9\times10^{-7}$) previously suggestively associated with OA and *BMP1* ($p=2.9\times10^{-7}$) which induces bone and cartilage development. This work has the potential to yield an unprecedented amount of information on OA susceptibility genes.

884F

Two novel susceptibility loci to Takayasu arteritis and synergistic role of the *IL12B* and *HLA-B* regions in a Japanese population. *C. TERAO[1,2], H. YOSHIFUJI[2], A. KIMURA[3], T. MATSUMURA[4], K. OHMURA[2], M. TAKAHASHI[1], M. SHIMIZU[1], T. KAWAGUCHI[1], Z. CHEN[3], T. NARUSE[3], A. SATO-OTSUBO[5], Y. EBANA[6], Y. MAEJIMA[7], Y. WADA[8], I. NARITA[8], Y. KAWAGUCHI[9], H. YAMANAKA[9], S. OGAWA[5], I. KOMURO[4], R. NAGAI[10], R. YAMADA[1], Y. TABARA[2], M. ISOBE[7], T. MIMORI[2], F. MATSUDA[1], Kyoto University Takayasu Arteritis Consortium.* 1) Center for Genomic Medicine, Kyoto University Graduate school of Medicine, Kyoto, Japan; 2) Department of Rheumatology and Clinical Immunology, Kyoto University Graduate school of Medicine, Kyoto, Japan; 3) Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 4) Department of Cardiovascular Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 5) Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo; 6) Department of Bio-informational Pharmacology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 7) Department of Cardiovascular Medicine, Tokyo Medical and Dental University, Tokyo, Japan; 8) Division of Clinical Nephrology and Rheumatology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; 9) Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan; 10) Jichi Medical University, Tochigi, Japan.

Takayasu arteritis (TAK) is an autoimmune systemic vasculitis of which etiology is unknown. While previous studies have revealed that HLA-B*52:01 has an effect on TAK susceptibility, no other genetic determinants have been established so far. Here, we performed genome-scanning of 167 patients with TAK and 663 healthy controls using Illumina Human-Exome array, followed by a replication study consisting of 212 patients with TAK and 1,322 controls. As a result, we found that the *IL12B* region on chromosome 5 (overallp=1.7×10^{-13}, OR:1.75, 95%CI:1.42-2.16) and the *MLX* region on chromosome 17 (overallp=5.2×10^{-7}, OR:1.50, 95%CI:1.28-1.76) as well as the *HLA-B* region (overallp=2.8×10^{-21}, OR:2.44, 95%CI:2.03-2.93) exhibited significant associations. A significant synergistic effect of the SNP in the *IL12B* region and HLA-B*52:01 was found with a relative excess risk of 3.45, attributable proportion of 0.58, and synergy index of 3.24 (p≤0.00028) in addition to a suggestive synergistic effect between the SNP in the *MLX* region and HLA-B*52:01 (p≤0.027). We also found that the SNP in the *IL12B* region showed a significant association with clinical manifestations of TAK, including increased risk and severity of aortic regurgitation, a representative severe complication of TAK. Detection of these susceptibility loci in the current study will provide a new insight to the basic mechanisms of TAK pathogenesis. Our findings indicate that *IL12B* plays a fundamental role on the pathophysiology of TAK in combination with HLA-B*52:01 and that common autoimmune mechanisms underlie the pathology of TAK and other autoimmune disorders such as psoriasis and inflammatory bowel diseases in which *IL12B* is involved as a genetic predisposing factor.

885W

Identification of CNVs association in saliva flow using PennCNV. *M. Lee[1], K. T.Cuenco[1], X. Zheng[1], E. Feingold[1], D.E. Weeks[1], R.J. Weyant[1], R.J. Crout[2], D.W. McNeil[2], M.L. Marazita[1].* 1) University of Pittsburgh, Pittsburgh, PA; 2) West Virginia University, Morgantown, WV.

Background/Objectives: Appalachian populations experience elevated amounts of poor health outcomes like oral disease. Families recruited through the Center for Oral Health Research in Appalachia (COHRA) participated in the NIH Gene, Environment Association Studies Consortium (GENEVA) genome-wide association study (GWAS) of dental caries. Caries is a complex trait also impacted by low saliva flow which helps establish the local environment that interacts with the teeth. The estimated heritability of saliva flow is 49% in COHRA subjects. Past GENEVA GWAS of saliva flow have not considered the potential impact of copy number variations (CNVs) as part of the genetic variability explaining saliva flow differences. The purpose of this study is to identify CNVs that are associated with saliva flow in the COHRA study population and to conduct a comprehensive investigation of these genetic variants. Methods: 1506 Caucasian subjects of age 5 to 75 years were available with saliva flow measures from the COHRA GENEVA population. Saliva flow was categorized as low (≤0.6 ml/min, n_1=925) or high (>0.6 ml/min, n_2=581). The Center for Inherited Disease Research generated genotypes and markers for CNVs with the Illumina 610-Quad platform. We generated CNV calls with the GC model wave adjustment procedure using the PennCNV software. Poor-quality samples identified with log R ratio standard deviation greater than 0.3 are excluded. CNVs with copy number <2 were defined as deletions, and those with copy number >2 were deemed duplications. Genome-wide CNV association analyses with saliva flow are being conducted. We also compare CNV association results with previous hits from single SNP GWAS to check long-range linkage disequilibrium between genome regions. Results: We are currently assessing CNV calls and their association with saliva flow. Identified associations of CNVs with saliva flow will enrich understanding of genetic variability in caries-related traits. Support: DE020127 (K T.Cuenco), and DE018903 and DE014899 (ML Marazita).

886T

Analysis of Copy Number Variation on Down- Syndrome Associated Atrioventricular Septal Defects. *D. Ramachandran[1], J. Mulle[2], A.E. Locke[3], P. Bose[1], L.J. Bean[1], S. Le[1], T. Rosser[1], K. Dooley[4], D.J. Cutler[1], E. Feingold[5], S.Y. Cheong[6], C.L. Cua[7], G.T. Capone[8], C.L. Maslen[9], R.H. Reeves[10], S.L. Sherman[1], M.E. Zwick[1].* 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 3) Center for Statistical Genetics and Department of Biostatistics, University of Michigan School of Public Health, MI; 4) Sibley Heart Center Cardiology, Children's Hospital of Atlanta, GA; 5) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 6) Battelle Center Mathematical Medicine, Nationwide Children's Hospital, Columbus, OH; 7) Heart Center, Nationwide Children's Hospital, Columbus, OH; 8) Down Syndrome Center, Kennedy Krieger Institute, Baltimore, MD; 9) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR; 10) Department of Physiology, McKusick Nathans Institute for Genetic Medicine, School Of Medicine, John Hopkins University, MD.

Atrioventricular septal defects (AVSD), a severe congenital heart defect (CHD), occur in the general population in ~1 in 10,000 births. Nearly 20% of infants with Down Syndrome (DS) have an AVSD, representing a ca. 2000-fold increased risk compared to the euploid population. We hypothesize that in the presence of an extra chromosome 21, otherwise benign copy number variants (CNVs) act in an additive manner to explain the increased penetrance of DS-associated AVSD. We have used the Affymetrix SNP 6.0 genotyping platform to comprehensively characterize CNVs in 538 DS samples, consisting of 243 cases (DS + complete AVSD) and 295 controls (DS - CHD). We implemented strict quality control filters to minimize false positive calls, including 3 algorithms to make the CNV calls (BEAST, GADA and GLAD). We also required each putative CNV call to contain > 10 SNPs within the interval. After excluding CNVs overlapping centromeres, we identified 1,711 total deletions (781 in cases and 930 in controls) and 1,302 duplications (607 in cases and 695 in controls). Association analyses in PLINK uncovered a ~275 kb deletion overlapping the MIR1324 gene on chr 3, which after permutation correction for multiple testing is genome-wide significant ($p<0.049$). Additionally, we have identified large rare duplications overlapping genes previously associated with CHD, including MAP3K3, POSTN, MYH11, CTSB and PLCB1, although standard burden tests in PLINK did not identify statistically significant differences between cases and controls in the number or size of the CNVs. This analysis suggests that the genetic factors contributing to the increased prevalence of DS- associated AVSD are complex and heterogeneous. Our efforts are currently focused on the analyses of rare variants in the data set that might contribute to AVSD susceptibility in this sensitized population.

887F

Analysis of Copy Number Variation on Chromosome 21 in Down Syndrome Associated Congenital Heart Defects. *B.L. Rambo-Martin[1], D. Ramachandran[1], J. Mulle[2], A.E. Locke[3], P. Bose[1], L.J. Bean[1], S. Le[1], T. Rosser[1], K. Dooley[4], D.J. Cutler[1], S.Y. Cheong[5], C.L. Cua[6], C.L. Maslen[7], R.H. Reeves[8], S.L. Sherman[1], M.E. Zwick[1].* 1) Human Genetics, Emory University, Atlanta, GA; 2) Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 3) Center for Statistical Genetics and Department of Biostatistics, University of Michigan School of Public Health, MI; 4) Sibley Heart Center Cardiology, Children's Hospital of Atlanta, GA; 5) Battelle Center Mathematical Medicine, Nationwide Children's Hospital, Columbus, OH; 6) Heart Center, Nationwide Children's Hospital, Columbus, OH; 7) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR; 8) Department of Physiology, McKusick Nathans Institute for Genetic Medicine, School Of Medicine, John Hopkins University, MD.

Atrioventricular septal defects (AVSD) are a life threatening congenital heart defect (CHD) inflicting substantial social and healthcare costs. Traditional genetic mapping strategies have revealed a limited but highly heterogeneous, set of aberrant genes responsible for AVSD, and suggest that many AVSD risk loci remain undiscovered. Children born with Down Syndrome (DS) are at a 2000-fold increased risk for AVSD over the general population, resulting in AVSD prevalence in DS of 1 in 5. These data suggest that individuals with trisomy 21 compose a sensitized cohort with respect to AVSD, and may help reveal novel AVSD susceptibility genes. In particular, we hypothesize that copy number variants (CNVs) on chromosome 21 influence the risk of AVSD in individuals with DS, and therefore explain the elevated risk for AVSD found in the DS population. We have performed high-density array comparative genomic hybridization of chromosome 21 in 221 cases (individuals with DS and complete AVSD) and 247 controls (individuals with DS without any CHD). We used the ADM2 algorithm that identifies a CNV as a region with a mean Log2 score significantly departing from the expectation of Log2 = 0. In cases and controls, 668 deletions and 303 duplications were detected. Duplications ranged in size from 1.9 to 782 kb and averaged 108 kb. Deletions ranged in size from 1.8 kb to 391 kb and average 31 kb. Initial analyses suggested no statistically significant differences in the overall rate or size of CNVs between our cases and controls. After filtering out common variants found in the Database of Genomic Variants, cases were found to have a significantly higher fraction of genes duplicated when adjusting for total duplication coverage in each sample (empirical p-value = 0.05). Continuing efforts include: evaluating additional CNV calling algorithms, testing for CNV enrichment over AVSD candidate genes on chromosome 21 and performing pathway analyses. Candidate CNV regions identified from association testing will be validated with Taqman PCR and breakpoint sequencing. Preliminary analyses suggest that the effect of a single or a few common CNVs on chromosome 21 alone do not explain the 2000-fold increased risk for AVSD in the DS population. Our data bolster arguments that there is a high level of heterogeneity found in the etiology of AVSD.

888W

Identification of *SPOCK2* as a susceptibility gene for bronchopulmonary dysplasia. *A. Hadchouel[1, 2, 3], X. Durrmeyer[4], E. Bouzigon[5, 6, 7], R. Incitti[8], J. Huusko[9], P.H. Jarreau[10], R. Lenclen[11], F. Demenais[5, 6, 7], M.L. Franco-Montoya[3], I. Layouni[4], J. Patkai[10], J. Bourbon[3], M. Hallman[9], C. Danan[4, 12], C. Delacourt[1, 2, 3].* 1) Pneumologie Pédiatrique, Hôpital Necker Enfants Malades, APHP, Paris, France; 2) Université Paris Descartes - Sorbonne Paris Cité, Institut Imagine; 3) Inserm, U955, Créteil, France; 4) Réanimation Néonatale, Centre Hospitalier Intercommunal, Créteil, France; 5) INSERM, U946, Paris, France; 6) Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, Paris, France; 7) Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain, Paris, France; 8) Institut Mondor de Recherche Biomédicale, INSERM, U955, Plateforme Microarray, Faculté de Médecine, Créteil, France; 9) Institute of Clinical Medicine, Department of Pediatrics, University of Oulu, Oulu, Finland; 10) Service de Réanimation Néonatale, CHU Cochin-Port Royal, Paris, France; 11) Service de Réanimation Néonatale, CH Poissy-Saint Germain, Poissy, France; 12) Unité Fonctionnelle de Recherche Clinique, Centre Hospitalier Intercommunal, Créteil, France.

Bronchopulmonary dysplasia (BPD), defined as a requirement for oxygen supplementation at 36 weeks of postmenstrual age (PMA), is the most common chronic respiratory disease in premature infants and its treatment places major demands on health services. Despite considerable advances in the care of very-low-birth-weight (VLBW) infants, BPD still occurs among 20-40% of survivors. Besides the recognized detrimental effects of environmental factors, VLBW twin concordance studies suggested a role of genetic factors. Candidate-gene studies failed to identify robust associations in this setting. The objective of our work was to perform a first genome-wide association study (GWAS) in BPD. We prospectively evaluated 418 premature neonates (PMA lower than 28 weeks), of whom 22% developed BPD. Two discovery series were created using a DNA pooling strategy in neonates from white and African ancestries. Association between single nucleotide polymorphisms (SNPs) and BPD were investigated by two different analysis methods. SNPs associated with the disease were confirmed in an independent replication population. Selected genes were then explored by fine mapping and associations were replicated in an external Finnish population of 213 neonates. Validated genes expression patterns were studied in rat lung during physiologic development and after hyperoxia exposure. The *SPOCK2* locus was the only region identified by the two analysis methods in both discovery series. The most significant polymorphism (rs1245560; p=1.66 × 10-7) was confirmed by individual genotyping and in the replication population (p=0.002). Fine mapping confirmed the association of rs1245560 with BPD in both white and African populations with adjusted odds ratios (ORs) of 2.96 (95% confidence interval [CI], 1.37-6.40) and 4.87 (95% CI, 1.88-12.63), respectively. In white neonates, rs1049269 was also associated with the disease (OR=3.21; 95% CI, 1.51-6.82). These associations were replicated in the Finnish population. In newborn rat lungs, *SPOCK2* mRNA levels markedly increased during alveolarization. After rat exposure to hyperoxia, *SPOCK2* expression significantly increased relative to air-exposed controls. Immunofluorescent studies showed that the SPOCK2 protein was expressed throughout the pulmonary extra-cellular matrix. Thus, a GWAS identified *SPOCK2* as a new candidate susceptibility locus for BPD. Its pattern of expression during lung development points to a potential role in alveolarization.

889T

Dipeptidyl peptidase 10 (*DPP10*) is associated with severe retinopathy in type 1 diabetes. *S.M. Hosseini[1], K. Howard[2], L. Sun[3], A.P. Boright[4], D.A. Tregouet[5], N. Sandholm[6], M.S. Lajer[7], K. Hietala[6], C. Forsblom[6], M. Marre[8], P. Rossing[7], P.H. Groop[6], A.J. Canty[9], S. Hadjadj[10], B.E. Klein[2], S.B. Bull[3,11], R. Klein[2], A.D. Paterson[1,3], the DCCT/EDIC Research Group.* 1) Gen & Genomic Biol, Hosp Sick Children, Toronto, ON, Canada; 2) Dept of Ophthalmology & Visual Sciences, Univ Wisconsin, Madison, WI, USA; 3) Dalla Lana School of Public Health, Univ Toronto, Toronto, Canada; 4) Dept of Medicine, Univ Toronto and LMC Diabetes & Endocrinology, Toronto, Canada; 5) INSERM UMR_S 937, ICAN Institute for Cardiometabolism and Nutrition, Pierre & Marie Curie Univ, Paris, France; 6) Folkhälsan Inst of Genetics, Univ of Helsinki, Helsinki, Finland; 7) Steno Diabetes Center, Gentofte, Denmark; 8) INSERM U695 & Dept of Endocrinology, Bichat-Claude Bernard Univ Hosp, Paris, France; 9) Dept of Mathematics & Statistics, McMaster Univ, Hamilton, Canada; 10) Dept of Endocrinology & Diabetology, Univ of Poitiers, Poitiers, France; 11) Samuel Lunenfeld Research Institute, Toronto, Canada.

Background: Diabetic retinopathy (DR) is a major health problem affecting most patients with diabetes, yet only a subset progress to severe DR (SDR). Duration of diabetes and glycemia are the main risk factors for DR. Several lines of evidence suggest a genetic contribution to the risk of SDR; however, no genetic variant has shown convincing association with DR in genome-wide association studies (GWAS). **Purpose:** To identify common polymorphisms associated with SDR. **Methods:** White persons with type 1 diabetes (T1D) participating in the Epidemiology of Diabetes Interventions and Complications (EDIC, n=1304) and Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR, n=603) were genotyped by Illumina Human1M and Omni1-Quad assays, respectively. Following quality control procedures, genotypes of ~2.5M autosomal SNPs from phase 2 and 3 of HapMap were imputed in both studies. SDR status [level 53/<53 or worse on ETDRS scale or panretinal laser photocoagulation treatment] was identified by grading 7-field retinal photographs obtained at the most recent follow-up visit. Multivariable models accounted for glycemia (measured by A1C), diabetes duration and other relevant covariates in the association analysis of SDR with additive genotype. Four GWAS were performed separately in WESDR and the 3 subgroups of EDIC (primary cohort, secondary cohort by treatment group - conventional or intensive) prior to combining the results in a meta-analysis. The association of top SNPs was evaluated in three independent white T1D cohorts: Genesis-GeneDiab (n=502), Steno (n=936), FinnDiane (n=2194). **Results:** Several SNPs in *DPP10* showed genome-wide significant association with SDR in the meta-analysis (3 SNPs in strong LD with P<5E-8, 14 with FDR<0.05). In meta-analysis, each copy of the risk allele of any of the top 3 SNPs at this locus doubled the odds of SDR (OR=2.22, 95%CI:1.66-2.98). Although, the top SNPs did not show significant association with SDR in the three replication studies (P>0.05), the direction of effect remained consistent in all but one of the examined populations. *DPP10* is strongly expressed in the brain and pancreas and alters the expression and biophysical activity of voltage gated potassium channels. It has been associated with asthma by linkage/association mapping. Unlike DDP4, target for anti-diabetic gliptins, DPP10 does not show protease activity. **Conclusion:** *DPP10* is the first locus to show strong evidence for association with SDR in a meta-GWAS.

890F

Common Regulatory Variants Mapped in Human Adipose and Skeletal Muscle But Not Lymphoblastoid Cell Lines Tissue Explain Much of the Heritability of Type 2 Diabetes. *J.M. Torres, E.R. Gamazon, N.J. Cox.* Medicne/Genetic Medicine, The University of Chicago, Chicago, IL, IL.

Previous studies have shown that top signals from genome-wide association studies (GWAS) on type 2 diabetes (T2D) are enriched for expression quantitative trait loci (eQTLs) identified in skeletal muscle and adipose tissue. We therefore hypothesized that such eQTLs might account for a disproportionate share of the phenotypic variance in liability to type 2 diabetes (T2D) estimated from all SNPs interrogated through GWAS. To test this hypothesis, we applied genome-wide complex trait analysis (GCTA) to the WTCCC-T2D GWAS dataset representing subjects from the United Kingdom (1924 cases, 2938 controls). We estimated the proportion of phenotypic variance attributable to additive effects of all variants interrogated in these GWAS (i.e. chip-based heritability), as well as from a much smaller set of variants identified as eQTLs for human skeletal muscle and adipose tissue. The estimate of chip-based heritability explained by the total set of GWAS variants is 60% (SE = 6%). The estimates of heritability explained by SNP subsets enriched for eQTLs mapped in human adipose and skeletal muscle tissue are 21% (SE = 7%) and 22% (SE = 6%), respectively. The combined set of adipose and skeletal muscle eQTL-enriched SNP subsets explained 33% (SE = 7% p-value = 4×10-6) of the phenotypic variation whereas the complement set of genome-wide SNPs explained less heritability (30%, SE = 7%), despite representing 3X more variants. The heritability estimate corresponding to the combined eQTL-enriched set was also greater and more statistically significant than the estimate of 6% (SE = 5%, p-value = 0.1) corresponding to a SNP set enriched for eQTLs mapped in human lymphoblastoid cell lines (LCLs) despite there being a greater number of SNPs (8% more) represented in the LCL eQTL-enriched set. Taken together, these results support our hypothesis that common eQTLs mapped in insulin-responsive tissues account for a substantial portion of the variance in liability to T2D.

891W

Higher BMI genetic risk score is associated with presence of radiographic knee osteoarthritis but not progression in the Osteoarthritis Initiative Study. *M.S. Yau[1], R.D. Jackson[2], M.C. Hochberg[1], S. Krishnan[3], D.J. Duggan[3], B.D. Mitchell[1], L.M. Yerges-Armstrong[1].* 1) University of Maryland, Baltimore, MD; 2) Ohio State University, Columbus, OH; 3) Translational Genomic Research Institute, Phoenix, AZ.

Purpose: Increased BMI is highly associated with knee osteoarthritis, but the association could be a cause or effect of the condition. To distinguish between these alternatives, we assessed the association of BMI genetic risk score (GRS) with radiographic knee osteoarthritis (rKOA) and structural progression, reasoning that higher BMI GRS would be associated with greater rKOA presence if BMI were indeed a causal risk factor. **Methods:** We created a weighted GRS in the Osteoarthritis Initiative (OAI) based on published effect sizes for 32 robustly associated BMI risk alleles ($P<5\times10^{-8}$) and tested its association with rKOA presence and progression using logistic regression, adjusting for sex and age. The OAI is a longitudinal cohort of racially diverse high-risk individuals ages 45-79 years at baseline. The current analysis is restricted to Caucasian participants with at least two consensus radiographs over four years. Genome-wide genotyping was conducted on the Illumina 2.5M platform and imputed to the 1000 genomes (CEU/June2011) using Minimac. We defined rKOA cases as individuals who have a KL grade ≥ 2 (n=1,912) in one or more knees and rKOA controls as individuals who have a KL grade ≤ 1 in both knees (n=1,225). Progressive cases have an increase in KL grade from baseline KL grade=1, 2, or 3 in one or more knees or total joint replacement (n=608). Controls for progression have no change in KL grade from baseline KL grade=1, 2, or 3 in both knees (n=1,239). **Results:** BMI GRS was significantly associated with presence of rKOA (adjusted OR=1.22, 95% CI=1.07-1.40), but not with rKOA progression (adjusted OR=1.08, 95% CI=0.90-1.29). Adjusting for baseline BMI substantially attenuated the association between GRS alleles on rKOA (adjusted OR=1.13, 95% CI=0.98-1.30), further supporting a causal association between overweight/obesity and rKOA. Also, *FTO* rs1558902 alone was associated with rKOA (adjusted OR=1.37, 95% CI=1.05-1.79), adding to previous reports that *FTO* is associated with osteoarthritis. BMI GRS calculated without *FTO* rs1558902 also remained significant (adjusted OR=1.17, 95% CI=1.00-1.37). **Conclusions:** Higher BMI GRS is associated with rKOA, but not rKOA progression, even without the most robustly associated *FTO* allele. This supports the hypothesis that the effect of BMI precedes disease onset and plays a role in the onset of rKOA, but not in the trajectory of the disease. These results may have practical implications for management of rKOA.

892T

Investigating Age-related Macular Degeneration in the Amish. *J.D. Hoffman[1], L.N. D'Aoust[1], J.N. Cooke Bailey[1], L. Jiang[1], R. Laux[1], A. Agarwal[2], W.K. Scott[3], M.A. Pericak-Vance[3], J.L. Haines[1,2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Ophthalmology and Visual Sciences, Vanderbilt University, Nashville, TN, United States; 3) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL.

Age-related macular degeneration (AMD) is the leading cause of blindness among the adult population in the developed world. To further the understanding of this disease, we have studied the genetically isolated Amish population of Ohio and Indiana. The Amish are more genetically homogeneous than the general U.S. population and live a stricter lifestyle, thus reducing the variability of environmental effects on AMD in this population. For this analysis we set out to perform a genome-wide association scan for AMD and to characterize how known AMD associated loci contribute to disease risk in the Amish. Additionally, we examined how that risk compares to a general population based sample. We used genotyping data generated with the Affymetrix Human SNP Array 6.0 and Sequenom MassARRAY platforms. In total we analyzed 1,022 Amish individuals (96 with self-reported AMD). Case-control association analysis with adjustment for relatedness based on the complete 13-generation pedigree was carried out using MQLS (Thornton & McPeek 2007). Heterogeneity LOD (HLOD) scores were computed for affecteds-only parametric and nonparametric 2-point linkage using Merlin (Abecasis et al, 2002), with a disease allele frequency of 10%. Mean cumulative genetic risk scores were calculated using a previously described logistic regression model. In our cumulative genetic risk score analysis we observe a mean risk score of 1.12 (95% CI [1.10, 1.13]) in our Amish controls and 1.18 (95% CI [1.13, 1.22]) in our Amish cases. This mean difference in risk scores is statistically significant (p-value of 0.0042). We also observe a significant decrease in cumulative mean risk scores in the Amish when compared to the general sample population for both control (p-value of 0.012) and case datasets (p-value < 0.00001. Our most significant uncorrected MQLS p-value observed in the genome-wide scan is $8.36*10^{-8}$ for the intronic SNP rs16988305 on chromosome 20 in the gene encoding UBOX5. Additional signals are observed at rs17049426 (p-value of $1.67*10^{-7}$) on chromosome 3, and rs1451915 (p-value of $7.76*10^{-7}$) on chromosome 1. These results demonstrate that although the Amish have a similar burden of disease as compared to general populations, the lower cumulative genetic risk indicates that other AMD loci remain to be identified in the Amish.

893F

A genome-wide association study by using imputed genotypes identifies a susceptibility locus for Crohn's disease in a Japanese population. *K. Yamazaki[1], J. Umeno[1,2], A. Takahashi[3], A. Hirano[1,2], T. Johnson[4], T. Morizono[4], T. Kawaguchi[5], M. Takazoe[5], T. Yamada[6], Y. Suzuki[6], H. Tanaka[7], S. Motoya[7,8], M. Hosokawa[8], Y. Arimura[8], Y. Shinomura[8], T. Matsui[9], T. Matsumoto[2], M. Iida[2], T. Tsunoda[4], Y. Nakamura[10], M. Kubo[1].* 1) Laboratory for Genotyping Development, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; 2) Department of Medicine and Clinical Science, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 3) Laboratory for Statistical Analysis, Center for Integrative Medical Science, RIKEN, Tokyo, Japan; 4) Laboratory for Medical Science Mathematics, Center for Integrative Medical Science, RIKEN, Yokohama, Japan; 5) Department of Medicine, Division of Gastroenterology, Social Insurance Chuo General Hospital, Tokyo, Japan; 6) Department of Internal Medicine, Faculty of Medicine, Toho University, Chiba, Japan; 7) Department of Gastroenterology, Sapporo Kosei Hospital, Sapporo, Japan; 8) First Department of Internal Medicine, Sapporo Medical University School of Medicine, Sapporo, Japan; 9) Department of Gastroenterology, Fukuoka University Chikushi Hospital, Fukuoka, Japan; 10) Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

Crohn's disease is (CD) an inflammatory bowel disease (IBD) induced by multiple genetic and environmental factors. The prevalence of IBDs is much lower in Asian countries, including Japan, than in Western countries, but it is rapidly increasing. Genome-wide association studies (GWAS) have identified many genetic factors for CD in the European population, but information in other ethnic groups is scarce. Recently, we reported two novel susceptibility loci for CD in a Japanese population. To search for additional candidate loci, we performed genome-wide genotype imputation in the GWAS cohort (372 cases and 3,366 controls) using 286 East-Asian subjects from the 1000 Genomes Project March 2012 release (phase 1, version 3) as a reference. From analysis of the imputed GWAS data, we found 44 additional candidate lead SNPs representing putative novel susceptibility loci for CD ($P < 1 \times 10^{-4}$). After applying quality control, we performed genotyping of 34 SNPs with sufficient accuracy. The replication study conducted real genotype data in 1,158 cases and imputed genotype data in 15,061 controls. We calculated combined association statistics using meta-analysis between the imputed GWAS data and mixed replication data (cases using real genotypes with imputed control data). As a result, rs11894081 on 2p25 showed a strong signal for CD in the Japanese population ($P = 3.58 \times 10^{-9}$). The nearest gene to rs11894081 was *SOX11* (SRY (sex determining region Y)-box 11). Although aberrant expression of *SOX11* activates histone marks in some B-cell neoplasms, the relationship between *SOX11* and CD is unknown. Moreover, this association was obtained from analysis of real genotype data for cases and imputed genotype data for controls. To clarify the association, we performed high-density mapping using real genotypes and identified the most responsible SNP for CD susceptibility.

894W

Genome-wide association study identifying genetic risk loci for full-thickness rotator cuff tears. *J.M. Farnham[1], R.Z. Tashjian[2,3], C.C. Teerlink[1], L.A. Cannon-Albright[1,2].* 1) Div. of Genetic Epidemiology, School of Medicine, University of Utah, Salt Lake City, UT; 2) George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah; 3) Dept. ofOrthopedics, School ofMedicine, University of Utah, Salt Lake City, UT.

Introduction. Limited data exits on the genetic predisposition for the development of rotator cuff tearing. Previous evidence suggests a familial predisposition for the development of rotator cuff tearing utilizing a large population database(1) The purpose of the present study was to further investigate the genetic predisposition by identifying potential genetic variants associated with the development of rotator cuff tears. Materials and Methods. A total of 175 patients were identified as having full-thickness posterosuperior rotator cuff tears as defined by MRI. DNA was obtained through a blood sample from these individuals. Samples were genotyped on the Illumina Omni5.0 SNP platform. We used 3293 Caucasian controls from the Illumina icontrols database genotyped on the Illumina 550K platform. The intersection of the two platforms contains 287,563 markers across the autosomal and X chromosomes. We performed typical quality control procedures using Plink software resulting in 272,641 markers for analysis. We used GEMMA software to conduct tests of association at each marker. GEMMA software empirically estimates a kinship covariance matrix for all pairs of individuals, which is used in the analysis to mitigate potential effects of population stratification and cryptic relatedness. Results. Two SNPs were identified that exceeded the threshold for significance (p < 1.8e-7) after adjusting for multiple testing. One SNP (rs12491628; p = 6.9e-12; OR = 2.77) resides on the 3p24.1, approximately 100Kb upstream of the LRRC3B gene. The other significant SNP (rs161811; p =1.2e-8; OR = 3.17) resides on 1p36.2 in the TNFRSF9 gene. Discussion We have identified two potential genetic variants associated with the development of rotator cuff tearing. To our knowledge, these findings represent the first attempt to identify genetic factors influencing the development of rotator cuff tears via a genome-wide association study. Further research is being conducted to further characterize these regions with the goal of identifying specific genes associated with tearing. References. 1. Tashjian RZ, Farnham JM, Albright FS, Teerlink CC, Cannon-Albright LA. J Bone Joint Surg Am 2009; 91:1136-42.

895T

Evaluation of the common genetic architecture of problematic peer relationships. *B. St Pourcain[1,2], C.M.A. Haworth[3], O.S.P. Davis[3], N.J. Timpson[1], G. McMahon[1], J. Kemp[1], D. Evans[1], S.M. Ring[4], W. McArdle[4], J. Golding[4,5], R. Plomin[3], G. Davey Smith[1].* 1) MRC CAiTE, University of Bristol, Bristol, United Kingdom; 2) School of Oral and Dental Sciences, University of Bristol, UK; 3) MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, UK; 4) School of Social and Community Medicine, University of Bristol, UK; 5) Centre for Child and Adolescent Health, University of Bristol, UK.

Peer interaction plays an important role in the development of social competence, and problematic childhood peer relationships often precede later maladjustment. Some links between early peer rejection and later maladaptive functioning however might be mediated through an underlying pathology. For example, impairments in social interaction skills are characteristic and heritable symptoms of the autistic dimension. Our study performed a twin study and a genome-wide analysis, to investigate genetic influences contributing to impaired peer relationships during childhood and adolescence. Heritability was estimated using ≤7366 UK twin pairs (TEDS) with parent-report on peer problems at 4, 7, 9 and 11 years. Using a multivariate (ACE) model, we found that a considerable proportion of the phenotypic variance is attributable to genetic factors ($0.60< h^2<0.71$), and that some genes affecting peer problems during development remain the same ($0.32< r_g<0.69$). Across development, we observed primarily non-shared environmental influences ($0.27< e^2<0.32$), with minimal evidence for shared environmental effects ($0.02< c^2<0.09$). A subsequent genome-wide screen was conducted in ≤6000 children from a large UK birth-cohort (ALSPAC) with parent-report on peer problem at 4, 7, 8, 10, 12, 13 and 17 years. The association between the skewed phenotypes and allele-dosage was assessed using a Poisson family model, and carried out for each time-point individually, thus allowing for genetic heterogeneity. The strongest signals ($2\times10^{-7}P\leq10^{-5}$) were eventually modelled longitudinally and further investigated in 793 autism pedigrees (AGRE). Two population-based signals contributed to risk for autism and were consistent in their direction of effect. However, combined evidence (based on P-values) only reached suggestive levels of significance. This included an age-independent SNP effect within the vicinity of *GLIS3* at 9p24.2 ($P_{ALSPAC}=4.9\times10^{-6}$, $P_{AGRE}=0.0030$, Fisher-Combined-$P=2.8\times10^{-7}$) and an age-dependent SNP effect (Age-SNP-$P_{ALSPAC}= 4.9\times10^{-5}$) near *GHR* at 5p12, which was strongest in later adolescence ($P_{ALSPAC,\ 17\ years}=2.5\times10^{-5}$, $P_{AGRE}=0.00058$, Fisher-combined-$P= 2.7\times10^{-7}$). Signals are currently being followed-up in ≤2835 independent individuals in TEDS (age 4-11 years). Together, our findings support the contribution of common genetic effects to variation in problematic peer relationships, though many more samples might be required to reliably identify individual SNP signals.

896F

Reclassification in genetic risk prediction over time. *R. Barfield[1], J. Krier[2], R. Green[3], P. Kraft[1,4].* 1) Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; 2) Harvard Medical School Genetics Training Program, Boston, Massachusetts; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; 4) Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts.

Genome wide association studies have identified hundreds of common alleles with modest effects on risk of complex disease. The clinical utility of genetic risk profiles based on these variants depends crucially on the number and effect size of identified loci, and how stable the predicted risks are as additional loci are discovered. Individuals flagged as high risk at one time may be reclassified as low risk or vice versa as more loci are identified. Motivated by the development of a risk report for common complex disease as part of the MedSeq project on the integration of whole genome sequencing into clinical medicine, we quantify this reclassification using breast cancer (BrCa) and heart disease (CHD) as examples. Published results for potential causal SNPs were taken from the NHGRI catalogue. We simulated genotypes for a EUR population (n=100000) for these SNPs from the 1000 genomes projects. We calculated the predicted odds of disease for an individual as the product of the genotype specific odds ratios across all the risk markers known in 2007, 2009, 2011 and 2013. The odds were then normalized by the mean odds for all members of the simulated cohort for that year. The range of the predicted risks increased from 2007 to 2013: the 95th risk percentiles rose from 1.34 and 1.33 times the population average to 2.11 and 2.22 for CHD and BrCa, respectively. This caused 6% of the population to be reclassified from lower than 2x average CHD risk in 2007 to higher in 2013. The reclassification proportion for BrCa was 7%. The proportion of subjects at very high risk (greater 5x population average) increased from 2007 to 2013 (0 to 4×10^{-4} and 7×10^{-4} for CHD and BrCa respectively). We used the distribution of GWAS-identified risk markers to estimate the number of as-yet-unknown common risk alleles and their effect distribution, and used this to project the number of risk markers identified in future studies. The future reclassification from above 2x risk to below (or vice versa) projected from doubling available GWAS samples was notably smaller. This suggests that risk estimates in the 0.5-2.0 times average range--a widely used clinical benchmark for common diseases--are stable for diseases that have been as extensively studied as CHD and BrCa. Finally, our projections place upper bounds on the contribution of rare variants to total heritability, giving insight into clinical utility of rare variants.

897W

Genome-wide association study on Morbid Obesity in Taiwan Han Chienese. *H.C. Chang[1], W.J. Lee[2], H.C Yang[3], W.H. Pan[1].* 1) Institution of Biomedical Science, Academia Sinica, Taipei, Taiwan; 2) Ming-Sheng General Hospital, Taoyuan, Taiwan; 3) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

Development of obesity and successful weight loss are influenced by multiple genes, the combinations of which are most likely ethnic-specific. While there is yet to be any systematic research regarding the obesity genome of the Chinese Han ethnic group, the obesity rate in Taiwan has been increasing steadily over recent years. The need for such information in obesity prevention and management not only has medical urgency, but should also have commercial value. In this study, we intend to recruit 570 obese persons who were body mass index (BMI: weight in kilogram divided height in meter square) equal or over 35 (as morbid obesity, MO) from Bariatric Surgery center in Ming-Sheng General Hospital , and compare with the 570 control subject with BMI less than 24 from TGWAS (Taiwan genome wide association study). The variables considered in patient matching include: (1) gender, (2) age (±5 years old). Inclusion criteria include: (1) between 20-55 years of age and (2) parents claimed to be of Han ethnicity. Via Axiom Genome-Wide CHB Array Plate which has been specific designed for CHB populations from Affymetrix Company, we will conduct the genotyping. And then, we hope to perform a whole genome screening (via genome-wide association studies, GWAS) to investigate related DNA single-nucleotide polymorphism (SNP) focusing on individual BMI phenotype. Final, we aim to construct a sensitive and specific genome composed of obesity-related genes entirely specific to Taiwan's Han ethnic group. The results can be used in the future to screen for high-risk populations.

898T

Genome-wide association study identifies novel loci associated with abdominal obesity in Africans. *A. Doumatey[1], A. Adeyemo[1], G. Chen[1], F. Tekola-Ayele[1], AR. Bentley[1], J. Zhou[1], H. Huang[1], D. Shriner[1], D. Ngare[8], O. Fansanmade[2], T. Johnson[2], J. Oli[3], G. Okafar[3], BA. Eghan[4], K. Agyenim-Boatang[4], J. Adeleye[5], W. Balogun[5], C. Adebamowo[5,7], A. Amoah[6], J. Acheampong[4], CN. Rotimi[1].* 1) NHGRI/CRGGH, National Institutes of Health, Bethesda, MD; 2) Department of Medicine, University of Lagos, Lagos, Nigeria; 3) Department of Medicine, University of Nigeria Teaching Hospital, Enugu, Nigeria; 4) Department of Medicine, University of Science and Technology, Kumasi, Ghana; 5) Department of Medicine, University College Hospital, Ibadan, Nigeria; 6) Department of Medicine and Therapeutics, University of Ghana Medical School, Accra, Ghana; 7) Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD; 8) Department of Mental Health and Behavioral Sciences, Moi University of Medicine, Eldoret, Kenya.

Obesity, especially abdominal obesity (AOb), is a risk factor for cardiovascular diseases and type 2 diabetes (T2D). AOb is heritable and several loci have been shown to influence its distribution at the population level. However, most prior studies of genes influencing AOb were performed in populations of non-African ancestry, a limitation given known differences in AOb by ethnicity. In this study, we conducted the first genome-wide association study of AOb in Africans. The study participants comprised 1808 Africans enrolled in the Africa America Diabetes Mellitus (AADM) study. Waist circumference (WC) and Waist-to-hip ratio (WHR) were used as measures of abdominal adiposity. Genotyping was carried out using the Affymetrix Axiom PANAFR SNP array (~2.1 million SNPs) imputed using the latest 1000 Genomes cosmopolitan reference for a total of 15 million SNPs included in the analysis. Association tests were adjusted for age, gender, T2D, BMI, and the first 3 principal components (PCs) of the genotypes. A cluster of SNPs on chromosome 2 (2p25.3, near FAM110C) and rs116158389 on chromosome 19 (19p13.2, near RDH8) were associated with WC at genome-wide significance level ($p<5\times10\text{-}8$). The top-ranked SNP on chromosome 2 (rs114441856, FAM110C, p=5.1×10-10) is in moderate- to complete LD($0.64<r2<1$) with the other associated SNPs in the region, indicating the likelihood of a single signal at the locus. FAM110C is expressed in adipocytes and has been implicated in metabolic processes including microtubule remodeling and Akt1 activation. The associated chromosome 2 variants are polymorphic in sub-Saharan Africans but are monomorphic in other continental populations. Several loci were nominally (7.6×10-8<p-value< 1.6×10-6) associated with WHR, including variants on chromosomes 19 (near RDH8, ZNF814, and MIR5589), 1 (RALGPS2), 2 (near REG3G), and 11 (in MMP13). This first GWAS of AOb in Africans identified novel loci that may provide new insights into the genetic architecture of obesity. While additional studies are needed to confirm and extend the findings, these results underscore the importance of studying multiple global populations in genetic studies of complex traits.

899F

Genome-wide association analysis shows the highly polygenic character of age-related hearing impairment. *E. Fransen[1, 2], S. Bonneux[1], J.J. Corneveaux[3], I. Schrauwen[1, 3], F. Di Berardino[4], C.H. White[5], J.D. Ohmen[5], P. Van de Heyning[6], U. Ambrosetti[4], M.J. Huentelman[3], G. Van Camp[1], R.A. Friedman[5].* 1) Center Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) StatUa center for statistics, University of Antwerp, B-2000 Antwerp; 3) Neurogenomics Division, The Translational Genomics Research Institute, Phoenix, Arizona 85004 USA; 4) Audiology Unit, Dept. of Clinical Sciences & Community Health, University of Milan; Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; 5) Cell Biology and Genetics Division, House Research Institute, Los Angeles, CA 90057; 6) Department of Otolaryngology, University Hospital of Antwerp, B-2650 Edegem, Belgium.

We have performed a genome-wide association study (GWAS) to identify the genes responsible for age-related hearing impairment (ARHI), the most common form of hearing impairment in the elderly. Analysis of common variants, with and without adjustment for stratification and environmental covariates, rare variants and interactions, as well as gene-set enrichment analysis showed no variants with genome-wide significance. No evidence for replication of any previously reported genes was found. The phenotype depends on the aggregated effect of a large number of SNPs, of which the individual effects are undetectable in a modestly powered GWAS. We estimate that 22% of the variance in our dataset can be explained by the collective effect of all genotyped SNPs. A score analysis showed a modest enrichment in causative SNPs among the SNPs with a p-value below 0.01. This is the first time that ARHI GWAS data are analyzed in such a manner, and the results prove that ARHI is highly polygenic in nature, with probably no major genes involved.

900W

Metabolic pathways in relation to obesity: Untargeted metabolomic profiling in a large population-based study. *J. Kumar[1], A. Ganna[2], T. Fall[1], J. Prenni[3], C. Broeckling[3], J. Prince[2], L. Lind[4], E. Ingelsson[1].* 1) Department of Medical Sciences, Molecular Epidemiology, Uppsala University and Science for Life Laboratory, Uppsala, Sweden; 2) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 3) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, U.S.A; 4) Department of Medical Sciences, Uppsala University, Uppsala, Sweden.

Introduction: Obesity is the main cause of various disorders such as type 2 diabetes and cardiovascular diseases and has become a serious public health problem of increasing prevalence. In the present study, we investigated the associations of body mass index (BMI) with circulating metabolites in a non-targeted metabolomics experiment and then performed a genome-wide association study (GWAS) of associated metabolites. Materials and Methods: Ultra-performance liquid chromatography coupled with tandem mass spectrometry was employed to perform a non-targeted metabolite profiling in two population-based cohort studies (TwinGene; N=1520 and PIVUS; N=970). Linear regression analyses were performed to test the association of BMI with different metabolites adjusting for potential confounding factors (age, sex, seasonal variation, storage, handling time and time since last meal). Significant findings were selected by controlling a false discovery rate (FDR) of 1%. GWAS was performed on the significant metabolite features to identify their associated loci. Results: A total of 8,185 and 11,056 molecular features were observed after quality control in PIVUS and TwinGene, respectively. In total, 2,281 molecular features were found to be significantly associated with BMI in PIVUS (FDR<1%) and 1,216 of these features could be replicated in TwinGene (p<0.05). These 1,216 molecular features represent 113 different metabolites (based on similar retention time and high correlation) of which seven could be annotated with high confidence (1-oleoyl-2-hydroxy-sn-glycero-3-phosphocholine, Glycerophosphocholine, negatively associated with BMI; Alpha-Linolenic acid, Caffeine, Deoxycholic acid glycine conjugate, L-Carnitine and L-Tyrosine, positively associated with BMI) utilizing a private compound library. GWAS of these seven metabolites from PIVUS and TWINGENE showed that SNPs from carnitine gene (rs1171617, $p=2.41*10^{-30}$) were found to be significantly associated with levels of circulating carnitine. Conclusion: In conclusion, the application of non-targeted metabolite profiling to evaluate two large population-based cohorts revealed 113 metabolites to be significantly associated with BMI. Seven of these metabolites were annotated with high confidence. Genetic variants from the carnitine gene were found to be significantly associated with circulating levels of L-Carnitine. Further efforts to improve annotation of the remaining are underway through additional experiments.

901T

Genomewide Association Studies of Lipids in Samoans. *R.L. Minster[1], N.L. Hawley[2,3], G. Sun[4], H. Cheng[4], S. Viali[5], R. Deka[4], D.E. Weeks[1,6], S.T. McGarvey[7].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Weight Control and Diabetes Research Center, the Miriam Hospital, Providence, RI, USA; 3) The Alpert Medical School, Brown University, Providence, RI, USA; 4) Department of Environmental Health, School of Medicine, University of Cincinnati, Cincinnati, OH; 5) Medical Specialist Clinic and National Health Services, Government of Samoa, Apia, Samoa; 6) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 7) Department of Epidemiology, International Health Institute, School of Public Health, Brown University, Providence, RI.

Genomewide association studies (GWAS) of quantitative traits in population isolates affords a unique opportunity to identify quantitative trait loci. Studies in populations like Samoans can detect genes that are novel and population-specific as well as those that affect traits across populations. Obesity has greater prevalence among Samoans compared to many other ancestry groups, and identifying variation that affects lipids levels against this background could reveal novel mechanisms for intervention. Here we report the results of GWAS of levels of total cholesterol, high-density lipoprotein, low-density lipoprotein and triglycerides in 3,122 adults from Samoa. Age, age^2, sex, BMI, smoking and alcohol use were considered as covariates for each of the four phenotypes. A total of 871,188 autosomal single-nucleotide polymorphisms (SNPs) were genotyped and extensive quality control of the genotypes was performed. Five loci were identified with genomewide significance ($p < 5 \times 10^{-8}$), all of which have been reported previously (*APOE*, *ZNF259*, *CETP*, *GCKR* and *LIPG*). Six loci, three of them novel, were identified with suggestive associations between $p = 10^{-6}$ and $p = 5 \times 10^{-8}$. An additional twenty-eight loci had suggestive associations with p values between 10^{-5} and 10^{-6} with one or more of the four lipid traits. Additional studies in related Polynesian populations are necessary to verify these new findings. This work was supported by U.S. National Institutes of Health grant R01-HL093093 (P.I. Stephen McGarvey).

902F

Identification of genetic factors underlying asthma age-of-onset sub-phenotypes. *C. Sarnowski[1,2,3], M-H. Dizier[1,3], I. Ahmed[2,4], P. Margaritte-Jeannin[1,3], M. Lathrop[5,6], F. Demenais[1,3], E. Bouzigon[1,3], the EGEA cooperative group.* 1) UMR946, INSERM, PARIS, France; 2) Univ. Paris Sud, Paris, France; 3) Univ. Paris Diderot, Paris, France; 4) CESP, INSERM, UMRS 1018, Villejuif, France; 5) McGill Univ., Montréal, Canada; 6) CNG/CEA, Evry, France.

Asthma is a heterogeneous disease with variable clinical expression over the life span. The disease age of onset is one of the simplest features that can be used to differentiate asthma phenotypes. To characterize the genetic factors influencing asthma in age-of-onset specific manner, we conducted a GWAS using a multinomial regression model applied to 750 asthmatics categorized according to their age-of-onset and 1,085 non-asthmatics from the French EGEA study with HapMap2 imputed data. Asthmatics were split into four specific age-of-onset sub-phenotypes: A) age-of-onset ≤ 4yrs (early-onset), B) 5-12yrs (before puberty), C) 13-20yrs (between puberty and adulthood) and D) > 20yrs (adult-onset). First, we applied an association test allowing heterogeneity of SNP effect between sub-phenotypes (Morris et al. Genet Epidemiol 2010) and detected 60 SNPs with P-value $\leq 10^{-5}$. Then, we tested whether these SNPs had a heterogeneous effect among the four sub-phenotypes. We identified 53 SNPs located in 16 regions with an interclass heterogeneity P-value $\leq 10^{-3}$. Among these regions, six had intra-class association P-values $\leq 10^{-5}$. We confirmed the specific association between 17q12-q21 genetic variants and early-onset asthma ($P=10^{-6}$) (Bouzigon et al. N Engl J Med 2008). We also detected five new regions among which four loci with SNP effect restricted to one asthma age-of-onset sub-phenotype: 9q34 with phenotype A ($P=5\times10^{-6}$), 3q25 with phenotype B ($P=2\times10^{-7}$), 1p13-p12 and 3q27-q28 with phenotype C ($P\leq3\times10^{-7}$), and one locus (3p22) associated with both phenotypes C and D ($P=2\times10^{-7}$ and $P=8\times10^{-3}$ respectively). This analysis will be extended to GABRIEL Asthma consortium datasets. Thus, taking into account the age of onset in a multinomial regression framework can be a powerful approach to identify new loci underlying complex diseases. Funded by: Région Ile de France, Fonds de Dotation "Recherche en Santé Respiratoire" & ANR-GEWIS-AM, GABRIEL.

903W

Genetic Association of Serum Magnesium Levels in African Americans: the Atherosclerosis Risk in Communities (ARIC) Study. *A. Tin[1], A.R. Folsom[2], N. Maruthur[3], C.A. Friedrich[4], J. Coresh[1], E. Boerwinkle[5], W.H. Kao[1].* 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) University of Minnesota School of Public Health, Minneapolis, MN; 3) Johns Hopkins University School of Medicine, Baltimore, MD; 4) University of Mississippi Medical Center, Jackson, MS; 5) University of Texas School of Public Health, Houston, TX.

Background. Magnesium (Mg) is important in enzymatic reactions and the regulation of mitochondrial function and vascular tone. Low serum Mg is associated with a greater burden of cardiovascular disease, hypertension, and diabetes. African Americans (AAs) have both lower serum Mg and dietary Mg intake versus European Americans (EAs). Genome-wide association studies (GWAS) in EA identified six loci associated with serum Mg, but no GWAS of serum Mg have been conducted in AA. Methods. We conducted the first GWAS of serum Mg in 2758 AA ARIC participants using linear regression controlled for age, sex, and population substructure. We interrogated the flanking region (+/-500kb or the two closest recombination hot spots) of the six published serum Mg index SNPs from EA studies in our AA sample (significance level p <0.05 / number of independent SNPs in each locus). We then determined the impact of a genotype score (the sum of the alleles of the index SNPs associated with decreased Mg in each population) on the association between serum Mg and self-reported race in 2758 AA and 8626 EA ARIC participants.We tested the correlation of serum Mg with African ancestry. Results. No locus reached genome-wide significance in AA. The most significant association was on chromosome 14 ($p=1.4\times10^{-7}$). Of the six known genome-wide significant loci in EA, five were nominally significant in AA (MUC1 at chr1:155184975, $p=1.9\times10^{-7}$; MDS1 at chr3:168830777, $p=5.5\times10^{-4}$; SHROOM3 at chr4:77638013, $p=1.2\times10^{-3}$; TRPM6 at chr9:77481834, $p=2.2\times10^{-6}$; DCDC51 at chr11:30979139, $p=6.1\times10^{-4}$). In AA, the index allele associated with lower Mg had larger effect sizes (5/5) but lower frequencies (4/5) compared with those in EA. The five index SNPs explained 2.7% of the variance in serum Mg in AA. After accounting for age and sex, self-reported AA status was associated with 0.04 mmol/L lower serum Mg ($p=1.7\times10^{-122}$), which became 0.06 mmol/L ($p=2.0\times10^{-190}$) after accounting for the genotype score. The correlation between serum Mg and African ancestry was -0.05 (p=0.02). Conclusions. In a large sample of EA and AA middle-aged adults, most of the top loci for serum Mg in EA were also associated with serum Mg in AA, but these loci did not explain the disparity in serum Mg levels between the two populations. Additional fine mapping of these loci to identify the causal variants will be important for understanding the genetic basis of hypomagnesemia and its possible relationships to other conditions.

904T

Genome-wide search for age- and sex-dependent genetic loci for human anthropometric traits: Methods and results from genome-wide meta-analyses across 310,000 Individuals. *T.W. Winkler[1], Z. Kutalik[2,3,4], M. Graff[5], A. Justice[5], L. Barata[6], M. Feitosa[6], S. Chu[7], R. Mägi[8], J. Czajkowski[9], T. Fall[10], Y. Lu[11], T.O. Kilpeläinen[12], I.M. Heid[1,13], I. Borecki[6], K.E. North[5], R.J.F. Loos[11,14], GIANT (Genetic Investigation of ANthropometric Traits) Consortium.* 1) Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 2) Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; 3) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 4) Institute of Social and Preventive Medicine, CHUV-UNIL, 1010 Lausanne, Switzerland; 5) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA; 6) Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA; 7) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA; 8) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 9) Division of Statistical Genomics, Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, MO 63110, USA; 10) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 11) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 12) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Copenhagen, Denmark; 13) Genetic Epidemiology, Helmholtz Zentrum Muenchen-German Research Center for Environmental Health, Neuherberg, Germany; 14) MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK.

Height, adiposity and fat distribution differ between men and women and change over time (e.g. following menopause). Previously, genome-wide association meta-analyses (GWAMAs) of anthropometric traits revealed sexually dimorphic loci for waist-hip ratio (WHR), a measure of fat distribution, but little is known of whether genetic effects on anthropometric traits change with age, and whether such changes differ between men and women. To detect age and sex dependent genetic effects, we conducted GWAMAs stratified by age (cut-off at 50 years, i.e. average age of onset of menopause) and by sex. We used 91 imputed GWA studies and 23 MetaboChip studies comprising >310,000 individuals of European ancestry from the GIANT consortium. Each study tested up to 2.8M SNPs for association with height, BMI and WHR in four strata (men≤50y, women≤50y, men>50y, women>50y). For each trait, we conducted inverse-variance weighted meta-analyses (one for each stratum), subsequently tested the pooled stratum-specific estimates for difference between the age-groups, between the sexes and for heterogeneity between the four strata and controlled each test at 5% FDR. To boost statistical power, we focused only on the SNPs that showed some overall association in the four strata combined (P<1E-5). For BMI, our analysis yielded 13 loci with significant age-difference (near *COBLL1, DDC, NEGR1, TNNI3K, SEC16B, TMEM18, ADCY3, AC016194.1, TCF7L2, STK33, FTO, MC4R* and *APOC1*), of which 10 showed a stronger effect in the younger subgroup. For WHR, our analysis yielded 25 loci with significant sex-difference (16 novel sexually dimorphic loci near *MSC, GANAB, RPS6KA5, GORAB, MEIS1, PLXND1, FAM13A, MAP3K1, NSD1, LY86, NFE2L3, NKX3-1, DNAH10, KCNJ2, EDEM2* and *EYA2*; and all nine previously established sexually dimorphic variants for WHR). Of the 25 loci, 23 showed a stronger effect in women. Although we did not find any loci with significant sex-difference for BMI or age-difference for WHR, our scan on the heterogeneity between the four strata identified two additional loci for BMI (near *CXXC5* and *TSHZ2*) and one additional locus for WHR (near *BMP2*), which might reflect potential age or sex dependency. Our analysis did not yield any age or sex dependent genetic effects for height. Our results underscore the importance of sex- and age-stratified analyses to illustrate a sexually dimorphic and age-dependent genetic underpinning for anthropometric traits.

905F

Targeted resequencing of genome wide associated candidate regions for pediatric venous thrombosis. *A. Witten[1], A. Arning[1], A. Barysenka[1], Ch. Grote[1], M. Hiersche[1], F. Ruehle[1], U. Nowak-Goettl[2], M. Stoll[1].* 1) Genetic Epidemiology of vascular disorders, Leibniz-Institute for Arteriosclerosis Research at the University of Münster, Münster, NRW, Germany; 2) Institute of Clinical Chemistry, University Hospital Schleswig-Holstein, Germany.

Venous thrombosis is a common multifactorial disease, which is influenced by environmental and genetic factors. Here we present a genome wide association study (GWAS) in 212 trios with pediatric venous thrombosis comprising affected children and their parents. For two SNPs exceeding the threshold for genome wide association (p>10-5) determined by 1.000.000 bootstrap permutations replication was conducted in 201 trios with thromboembolic stroke (TS). Among these, rs1304029 and a SNP curtly missing the threshold for permutated p-value (rs2748331) reside in a region on chromosome 6q13 comprising the gene for beta-1,3-glucoronyltransferase 2 (B3GAT2), and are associated with pediatric VTE (rs1304029; p=1.42×10-6, rs2748331; p=6.11×10-6). rs2748331 was replicated (p=0.00719) in our GWAS on pediatric TS (combined p=7.88×10-7). Twenty-seven additional SNPs were associated at confident permutated P-values (p<1×10-4). To investigate possible causative gene variants in the resulting linkage disequilibrium based candidate regions, we performed a next generation resequencing approach in 24 affected children and 24 unaffected siblings. The selected target regions of about 11 Mb in total comprise 30 gene regions from 16 chromosomes. Custom target enrichment was performed using the NimbleGen SeqCap EZ Choice technology. The resulting DNA libraries were paired-end sequenced (100 cycles) on an Illumina HiScanSQ instrument yielding in 142.5 Gb sequence data in total and 91.8% bases with a QScore > 30. Sequence reads were mapped by using the BWA algorithm and analyzed by GATK yielding in 83% median target specificity and median target region coverage of 196x. Variant annotation was done by using SNPEff and Annovar software tools. A sibling TDT test was applied on the 41,478 identified variants, 4,062 of which were novel to compare the two sample groups. 23 significant (p<0.05) coding non-synonymous or UTR SNPs in 10 genes were identified and selected for subsequent validation using capillary sequencing and genotyping within the full cohort of 257 nuclear families. Future studies elucidating the functional relevance of these genetic variants are warranted to further elucidate the role of these in the pathogenesis of DVT.

906W

A meta-analysis of genome-wide association studies for adiponectin level in East Asians identifies a novel locus near *WDR11-FGFR2*.
Y. Wu[1], H. Gao[2,3,4], H. Li[5], Y. Tabara[6], M. Nakatochi[7], YF. Chiu[8,9], EJ. Park[10], S. Vadlamudi[1], M. Fogarty[1], W. Wen[11], XO. Shu[11], C. Shin[12], SH. Jee[10], LM. Chuang[13,14], T. Miki[15], M. Yokota[16], X. Lin[5], KL. Mohlke[1], ES. Tai[2,17,18], The Asian Genetic Epidemiology Network (AGEN) adiponectin working group. 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore; 3) NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 6) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 7) Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 8) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Taiwan; 9) Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan; 10) Institute for Health Promotion and Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Republic of Korea; 11) Division of Epidemiology, Department of Medicine; Vanderbilt Epidemiology Center; and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; 12) Department of Internal Medicine, Korea University Ansan Hospital, Ansan, South Korea; 13) Division of Endocrinology and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 14) Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan; 15) Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Ehime, Japan; 16) Department of Genome Science, Aichi-Gakuin University School of Dentistry, Nagoya, Japan; 17) Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore; 18) Duke-NUS Graduate Medical School, National University of Singapore, Singapore.

Blood levels of adiponectin, an adipocyte-secreted protein correlated with metabolic and cardiovascular risk, are highly heritable. Genome-wide association (GWA) studies have identified 14 loci harboring variants associated with adiponectin level. To identify novel adiponectin-associated loci, particularly those of importance in East Asians, we conducted a meta-analysis of GWA studies for adiponectin in 7,827 individuals, followed by *in silico* and *de novo* replication in 4,298 and 5,724 additional individuals of the Asian Genetic Epidemiology Network (AGEN). Of the adiponectin-associated loci previously reported, we confirmed the association at six loci including *CHD13*, *ADIPOQ*, *PEPD*, *CMIP*, *ZNF664* and *GPR109A*. We identified a novel signal on chromosome 10, ~300 kb from *WDR11* and ~300 kb from *FGFR2* ($P = 1.9\times10^{-14}$). We also provided suggestive evidence of another locus near *OR8S1-LALBA* on chromosome 12 ($P = 9.4\times10^{-8}$). *FGFR2* is a strong candidate gene implicated in adipocyte hyperplasia and hypertrophy; the novel signal at *WDR11-FGFR2* explained 0.6% of the total variation in adiponectin. Despite a consistent direction of allelic effect, the index SNP at *WDR11-FGFR2* did not show strong evidence of association ($P = 0.093$) in > 29,000 Europeans in the publicly released ADIPOGen data. The difference in the trait-increasing allele frequency (AGEN: 0.57; ADIPOGen: 0.24) and the significant level of association across populations suggested that a variant at *WDR11-FGFR2* might have a larger genetic effect on adiponectin level in East Asians than Europeans, or the pairwise LD between the index SNP and the un-typed causal variant vary across different populations. The adiponectin-increasing allele of the index SNP exhibited evidence of association with a decreased BMI-adjusted waist-hip-ratio in 17,560 East Asians ($P = 9.8\times10^{-3}$). At *WDR11-FGFR2*, twenty-two candidate variants that span 65 kb are in moderate to high LD ($r^2 > 0.6$) with the index SNP, and at least four candidate variants are located at or near enhancer marks in adipose nuclei and predicted to possibly alter the transcriptional activity of nearby genes. However, the reporter assays in differentiated adipocytes showed no allelic difference in transcriptional activity. Our findings improve knowledge of the genetic basis of adiponectin variation, demonstrate the shared allelic architecture of adiponectin and central obesity, and motivate further studies of underlying biological mechanisms.

907T

Genome-wide association analysis of skeletal muscle fiber types.
T. Karaderi[1], N. Oskolkov[2], C. Ladenvall[2], S. Keildson[1], A. Mahajan[1], L. Lind[3], E. Ingelsson[4,1], L. Groop[2], P. Franks[5], A.P. Morris[1], O. Hansson[2], C.M. Lindgren[1]. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden; 3) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 4) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Department of Clinical Sciences, Genetic and Molecular Epidemiology, Lund University, Malmö, Sweden.

The human body contains over 400 skeletal muscles, together constituting 40-50% of the total body weight. Skeletal muscle is a key organ in many basic bodily functions, such as movement, respiration and thermogenesis. Individual muscles consist of a combination of 3 types of fibers: type 1 fibers (red, slow and oxidative), type 2A (red, fast and oxidative) and type 2X (white, fast and glycolytic). It has been previously shown that muscle fiber composition is associated with obesity, weight loss and type 2 diabetes. Therefore, it is important to explore the genetic make-up of muscle fiber composition to better understand the factors affecting the risk of such conditions. We conducted a genome-wide association study (GWAS) of these three skeletal muscle fiber types in a total of 656 individuals from three independent cohorts from Sweden. Genotype data in each cohort were imputed up to 35 million variants from the 1000 Genomes 'all ancestries' reference panel (March 2012). Within each cohort, phenotypes were inverse rank normalized and tested for association with each high-quality imputed variant (info>0.4) under an additive model after adjustment for age and body-mass index. After a conservative double genomic control correction, the data was merged using fixed-effects meta-analysis, and filtered for total minor allele count ≥5. Four independent loci showed association at genome-wide significance ($p<5\times10^{-8}$) with either type 2A or 2X fibers, but none with type 1 fiber. Fiber type 2A is associated with rs74979762 (minor allele frequency (MAF)=0.005, $p=8.04\times10^{-9}$), which is located downstream of *SIPA1L2*, within a potential transcription factor binding site. There were also three loci associated with fiber type 2X. The first is intronic (rs145631867, MAF=0.02, $p=2.23\times10^{-8}$) and maps to *NYAP2* and the second locus is intragenic (rs7776803, near *STEAP1*, MAF=0.27, $p=2.29\times10^{-8}$). Finally, one associated SNP (rs149081100, MAF=0.02, $p=2.96\times10^{-8}$) maps in *ADRA1B*, within a DNase I hypersensitivity region and a possible strong enhancer-binding site for glucocortocoid receptors functioning in anti-inflammatory immune responses and regulation of glucose metabolism. In conclusion, we have performed the first GWAS of detailed physiological measures of human skeletal muscle fibre. The genetic associations we observe suggest that the make-up of fiber types in skeletal muscle are under at least partly genetic control.

908F

Genome-wide association study of skin pigmentation and tanning in African Americans. *K. Batai[1, 2], E. Shah[2, 3], R.A. Kittles[2, 3, 4].* 1) Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL; 2) Institute of Human Genetics, College of Medicine, University of Illinois at Chicago, Chicago, IL; 3) Section of Hematology/Oncology, Department of Medicine, University of Illinois at Chicago, Chicago, IL; 4) Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, IL.

Background: Skin pigmentation is a complex trait that is strongly controlled by genes and sunlight exposure. The identification of pigmentation genes is important for understanding of skin cancer susceptibility and vitamin D status. Genome-wide association studies (GWAS) in European, Asian, and African-European admixed Cape Verdean populations have identified multiple loci associated with pigmentation traits. Many of these GWAS loci are similar across populations while several are population specific. To date, there has not been any GWAS performed in African Americans (AAs) that attempt to identify loci associated with pigmentation traits. Here, we report results of the first GWAS of skin pigmentation and tanning ability in AAs.

Methods: Skin pigmentation (M-Index) of 215 AA men from Washington, DC was measured on subjects' forehead and upper inner arm using reflectometer and samples were genotyped using Illumina Infinium 1M-Duo bead array. After quality control, 994,618 SNPs were available for analyses. We tested association for the inner-arm (constitutive) M-Index using linear regression model adjusting for age and first 4 principal components (PCs)s and for tanning ability (inner arm M-Index subtracted from forehead M-Index) adjusting for age and first 3 PCs.

Results: A cluster of *SLC24A5* SNPs were significantly associated with M-Index and three of them reached genome wide significant P-value ($P<5.02 \times 10^{-8}$). The strongest association was found for rs2675345 ($P=2.90 \times 10^{-8}$). SNP, rs3828107, in *NR5A2* (1q32) was the most strongly associated with tanning ability ($P=2.16 \times 10^{-7}$). When UV exposure was included in the regression model, association between this SNP and tanning ability was attenuated, but rs9392299 in *FLJ43763* located in 6p25, about 200kb upstream of *IRF4*, a gene that was associated with tanning ability and sun sensitivity in European descent populations showed marginal significance ($P=6.52 \times 10^{-6}$).

Conclusion: We found that *SLC24A5* SNP showed stronger association with M-Index than previously reported GWAS SNPs in South Asian and admixed Cape Verdean populations. We successfully confirmed previous GWAS findings of 6p25 SNPs and tanning ability in our AA population. Our study also provides evidence for a novel loci on chromosome 1 associated with tanning ability.

909W

The First Genome-wide Association Study of Serum Lipids among Africans. *A.R. Bentley[1], D. Shriner[1], G.J. Chen[1], F. Tekola-Ayele[1], A.P. Doumatey[1], H. Huang[1], J. Zhou[1], O. Fasanmade[2], T. Johnson[2], J. Oli[3], G. Okafor[3], B.A. Eghan, Jr[4], K. Argyenim-Boateng[4], J. Adeleye[5], W. Balogun[5], C. Adebamowo[6], A. Amoah[7], J. Acheampong[4], D. Ngare[8], A. Adeyemo[1], C.N. Rotimi[1].* 1) Center for Research in Genomics and Global Health, Natl Human Genome Research Institute, Bethesda, MD; 2) Department of Medicine, University of Lagos, Lagos, Nigeria; 3) Department of Medicine, University of Nigeria Teaching Hospital, Enugu, Nigeria; 4) Department of Medicine, University of Science and Technology, Kumasi, Ghana; 5) Department of Medicine, University College Hospital, Ibadan, Nigeria; 6) Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD; 7) Department of Medicine and Therapeutics, University of Ghana Medical School, Accra, Ghana; 8) Department of Mental Health and Behavioral Sciences, Moi University School of Medicine.

The burden of chronic disease in Africa is increasing rapidly, with sharper increases predicted, yet the genetic determinants underlying relevant traits among Africans are largely unstudied. We have conducted the first genome-wide association study of serum lipids in Africa. The study participants comprised 1808 primarily West Africans enrolled in the Africa America Diabetes Mellitus (AADM) Study, characterized for total cholesterol (TC), HDL-cholesterol (HDLC), LDL-cholesterol (LDLC), and triglycerides (TG) and genotyped on the Affymetrix Axiom PANAFR array (~2.1 million SNPs). Imputation using the 1000 Genomes cosmopolitan reference yielded ~15 million SNPs. Association models were adjusted for gender, age, BMI, type 2 diabetes, and the first 3 principal components of ancestry. The most statistically significant results were for variants in the *APOE* region with LDLC (-0.5 mg/dl, lowest $p=5.5 \times 10^{-17}$ [rs7412]). A larger effect size was observed for a nearby rare variant that has not been reported among non-African populations, rs192607279 (-4.5 mg/dl LDLC, $p=1.5 \times 10^{-10}$). Variants in this region were also associated with lower TC (lowest $p=4.4 \times 10^{-10}$ [rs61679753]). Common *LDLR* variants were associated with ~0.3 mg/dl lower LDLC. rs111590558 near *HTR2A* was associated with LDLC (0.6 mg/dl, $p=4.0 \times 10^{-9}$); this SNP is monomorphic among those of European ancestry. Although not reaching genome-wide significance, a lead association for HDLC was in a well-known HDLC locus, *CETP*. No TG association reached statistical significance. Replication of previous findings was attempted using both an exact method and an LD-based method. Of the 248 SNPs investigated, 61 replicated. This is the first replication for a high proportion (56%) of these results in any African ancestry population. This first African GWAS of serum lipids confirmed the relevance of some known lipids loci (*APOE*, *LDLR*, and *CETP*) in this population, as well as revealing some novel associations, particularly among variants not present among non-African ancestry populations. Replication attempts of known loci were successful for 25% of SNPs investigated. The large proportion of variants for which this is the first replication in an African ancestry population is consistent with the hypothesis that a reduced influence of environmental risk factors in Africans residing on the continent may simplify detection of genetic associations compared to Africans in the Diaspora with more Western lifestyles.

910T

Identification of asthma-related trans-acting epistatic eQTL using Model-Based Multifactor Dimensionality Reduction (MB-MDR). *K. Bessonov[1], D. Croteau-Chonka[2], W. Qi[2], VJ. Carey[2], BA. Raby[2], K. Van Steen[1].* 1) Department of Electrical Engineering and Computer Science (Institut Montefiore), University of Liege, Liege, Belgium; 2) Channing Division of Network Medicine,Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States.

Genetic epistasis likely underlies most complex traits, including gene expression, yet is very difficult to detect using standard approaches. When considering the interactive effects on expression Quantitative Trait Loci (eQTL), interactions between genes, scattered throughout the genome, are referred to as trans-epistasis. In this work we aimed at identifying transcripts whose expression is regulated by SNP-SNP interactions using Model-Based Multifactor Dimensionality Reduction (MB-MDR). This model-free approach for epistasis detection involves reducing a high-dimensional GxG space to GxG factor levels that either exhibit high, low evidence or no evidence at all for their association to gene intensities of interest. We applied this method to a dataset consisting of 19,451 CD4+ lymphocyte expression phenotypes and genotypes from ~516,000 SNPs from 174 Caucasian non-smoking subjects from CAMP asthma cohort. Univariate analysis highlighted 1,844 genome-wide significant cis-acting eQTL. With MB-MDR, each of these was tested for potential synergetic interaction with the remaining ~516,000 trans-SNPs, excluding cis-SNPs located in genetic range between 2 Mb upstream and downstream of the eQTL gene. Our preliminary results reveal evidence for the existence of epistatic interactions in the regulation of gene expression. Out of the total of 1,086,601 cis-SNP/trans-SNP pairs with marginal p-value ≤ 0.001 based on 999 permutations, the top 10 trans-genes were all related to signaling and cell-cell adhesion pathways (PTPRD, CDH13, CNTN5, LRP1B). For example, the top 2 cis-trans SNP pairs ranked on the basis of their associated MB-MDR statistic (shown in brackets) - rs622614/rs6039399 (49.526) and rs5756379/rs11069218 (46.219)-regulating lipid hydrolysis/protein degradation and host immune system activation, respectively, both implicated in PI(3) kinase-mediated signaling. Some of the previously literature reported loci with genome-wide significance for Asthma included rs7544426/rs3771166 (17.12), rs2241632/rs1342326 (11.004), rs17051336/rs744910 (15.214),and rs11244079/rs2284033 (13.041). When variants were ranked by the number of interacting partners, those with the highest number of trans-SNP effect modifiers belonged to pathways involving signaling, detoxification and DNA repair. Follow-up of these results include replication of the epistatic interactions in independent eQTL datasets, and formal testing of identified pairs as asthma-susceptibility loci.

911F

Genome-wide association study (GWAS) of major depressive disorder in N=9300 Han Chinese women using low-pass sequencing data. *T. Bigdeli[1], R.E. Peterson[1], Y. Li[2], W. Kretzschmar[2], F. Yang[3], H.H. Maes[1], A.H. Fanous[1], B.T. Webb[1], B.P. Riley[1], J. Wang[4], S. Shi[5], Y. Chen[6], J. Marchini[2], R. Mott[2], S.A. Bacanu[1], K.S. Kendler[1], J. Flint[2], CONVERGE consortium.* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, PO Box 980126, Richmond VA, 23298, USA; 2) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, Oxfordshire, United Kingdom; 3) Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, No. 600 South Wanping Road, Shanghai, P.R. China; 4) Beijing Genomics Institute, Floor 9 Complex Building, Beishan Industrial Zone, Yantian District, Shenzhen 518083, P.R. China; 5) Huashan Hospital of Fudan University, No.12 Middle Wulumuqi Road, Shanghai, P.R. China; 6) CTSU, Richard Doll Building, Old Road Campus, University of Oxford, Headington, Oxford OX3 7LF, United Kingdom.

Background: Major depressive disorder (MDD) is a common, complex psychiatric disorder and a leading cause of disability worldwide. Although modestly heritable (~30-40%), a complex genetic architecture has hindered efforts to identify robustly associated genetic risk variants. We sought to evaluate the evidence for common genetic variation in the etiology of MDD) in a large, ethnically homogenous Chinese sample.

Methods: Using single nucleotide polymorphism (SNP) data imputed from low-pass (1.2X) sequencing data, we tested for association between common SNPs and MDD in 4168 cases and 4614 controls from the CONVERGE (China, Oxford and VCU Experimental Research on Genetic Epidemiology) project. We estimated the proportion of variation in disease susceptibility captured by common SNPs using the Genome-wide Complex Trait Analysis (GCTA) utility and considered the predictive value of scores constructed from GWAS results.

Results: In our preliminary analysis of ~7.5*M* SNPs, no single variant attained genome-wide significance (5×10^{-8}). Based on an estimated prevalence of 12%, common SNPs accounted for an estimated 15.7% of the variance in MDD risk (95%CI=[10.3,21.1], $P=2.0\times10^{-10}$). A polygenic score based on results from the Psychiatric Genomics Consortium (PGC) GWAS of MDD accounted for less than a tenth of a percent of the variance in disease risk (P=0.03). However, when then CONVERGE sample was randomly divided into training and testing sets, and SNP effects estimated by the best linear unbiased prediction (BLUP) method in the training set, the resultant aggregate score was modestly associated with MDD in the testing set ($P= 6.4\times10^{-8}$), accounting for 0.79%; of the variability in disease risk.

Discussion: We have conducted a large GWAS of MDD in an ethnically homogenous sample using SNP data imputed from low-pass sequencing data. In our preliminary analysis, no single variant demonstrated genome-wide significant association with MDD. Aggregate risk scores based on GWAS in European and Chinese populations were found to be of differential predictive value. These observations support a complex etiology for MDD, and possible population differences in predisposing genetic factors.

912W

Genome-wide Association Study of serum minerals in pediatric African-American and Caucasian cohorts. *X. Chang, P. Sleiman, H. Hakonarson.* Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Calcium, magnesium, potassium and sodium cations as well as the chloride and phosphorus anions are major dietary minerals which can not be manufactured by the human body. They play important roles in various biological functions and are essential to sustain life and maintain optimal health, and thus are commonly measured in blood serum to monitor a range of health conditions. The mineral intake of children and adolescents are especially important due to their rapid growth. In this study, we performed a genome-wide association study to explore the contribution of potential common genetic variations to serum concentrations of the six major dietary minerals in Children as well as the ratio of concentrations between minerals. A total of 10,308 children, 5602 Caucasians and 4706 African-Americans were included in the analysis. All samples were genotyped on the Illumina HH550 or 610 arrays. Maximum serum concentration values of the six minerals, and the mineral ratios, were tested for association by linear regression as implemented in plink. In the African-American pediatric cohort, we report association of the serum calcium concentration with a missense variant in the calcium-sensing receptor (CASR) gene on 3q13 (rs1801725 5.6×10E-3. The SNP was previously reported to correlate with the level of serum calcium in both Caucasian and Indian-Asian populations, but no previous studies show a calcium concentration association in the African-American population or pediatric populations. The association between SNP rs1801725 and calcium concentration was also confirmed in our Caucasian pediatric cohort with a p-value of 1.03×10E-4; combined P-val 2.64×10E-6. Our results extended the understanding of CASR mediated calcium regulation in African-American and Caucasian pediatric populations. In the Caucasian pediatric population, our results also confirmed another previously reported serum magnesium concentration associated SNP rs4072037 with a p-value of 1.74×10E-5 located in gene MUC1 on 1q21. We also investigated the ratio of Calcium/Magnesium, Sodium/Potassium and Chloride/Phosphorus concentrations in blood serum. We report significant association of Calcium/Magnesium ratio with variants located that map to a locus on 10q26 that contains the Fibroblast growth factor receptor 2 gene (FGFR2). Results of the GWAS for the six minerals and the mineral ratios will be presented at the meeting.

913T

GWAS meta-analysis identified a novel locus associated with corneal curvature in Asian populations. *P. Chen[1], C.Y. Cheng[1,2,3,8], C.C. Khor[1,3,4,6], T. Aung[2], S.M. Saw[1,2], T.Y. Wong[2,3], E.S. Tai[1,5,7], Y.Y. Teo[1,6,9,10,11].* 1) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 2) Singapore Eye Research Institute, Singapore National Eye Center, Singapore; 3) Department of Ophthalmology, National University of Singapore, Singapore; 4) Department of Paediatrics National University Health Systems, National University of Singapore, Singapore; 5) Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 6) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore; 7) Duke-National University of Singapore Graduate Medical School, Singapore; 8) Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore; 9) NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore; 10) Life Sciences Institute, National University of Singapore, Singapore; 11) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

Corneal curvature (CC) is an ophthalmic biometric that has connection to various eye diseases. CC is highly heritable but there are only two loci discovered currently, *FRAP1* and *PDGFRA*. Both of the two genes have been reported in Asian populations for the first time in an Asian genome-wide association (GWAS) meta-analysis. Although *PDGFRA* has been well established to be associated with CC, *FRAP1* has not been replicated in other populations. Meanwhile in Asian populations, the proportion of variance which is explained by these two genes ranges from 1.8% to 11.1%. We conducted a genome-wide association study meta-analysis of CC to provide further evidence for the association of *FRAP1* with CC in additional samples, and discover more genes that are still hiding behind the scene. In total, we recruited 5326 Chinese, 2138 Malay and 2124 Indian individuals in Singapore from 7 GWAS. This study adopted a two-stage meta-analysis scheme. The discovery stage studies (SP2, SiMES, SINDI and SCORM) have been recruited by the previous meta-analysis. Two novel studies, SCES and STARS comprising 2473 Chinese individuals were employed in the replication stage. The radius of CC was defined as the average reading of right and left eye. Within each study, the radius was normalized by an inverse-normal transformation. Imputation of the SNP array genotypes was done using 1000 Genomes project phase 1 (March 2012, GRCh37) cosmopolitan haplotypes as reference panel. Effect of the allele dosage was evaluated in a linear regression model which adjusted for age and gender. Principle components were also adjusted for Malay and Indians to take care of the population stratification. The meta-analysis was done using a fixed effect inverse-variance weighting scheme. We successfully replicated the signals at *FRAP1* and *PDGFRA* gene region. Notably, the association of *FRAP1* was again replicated in our replication studies. Besides the two known genes, we found that *CMPK1* gene was associated with CC for the first time (array SNP meta-analysis P value = 2.81E-9), whereas the imputed genotypes brought us even more power with more significant signal (imputed top SNP meta-analysis P value = 3.27E-11). Our meta-analysis replicated the two known genes and identified a novel locus to be associated with CC. We provided further evidence that *FRAP1* is associated with CC in Asian populations. However, whether it is an Asian-specific CC susceptibility gene remains to be assessed in other populations.

914F

Uncovering loci associated with urinary incontinence in African and Hispanic American women. *C. Chen[1], A. Rajkovic[2], A. Park[3], G. Heiss[4], S. Hendrix[5], N. Franceschini[4].* 1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) department of obstetrics, gynecology and reproductive science, University of Pittsburgh,Pittsburgh, PA; 3) Department of Obstetrics & Gynecology, Medstar Washington Hospital Center, Georgetown University School of medicine, Washington, DC; 4) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 5) Department of Surgical Services, Michigan State University College of Osteopathic Medicine, East lansing, MI.

Urinary incontinence is an under-recognized health problem that negatively impacts women's quality of life. To date, genes contributing to urinary incontinence susceptibility have not been identified. We present the first genome-wide association study aimed to identify single nucleotide polymorphisms associated with overall urinary incontinence and two subtypes of urinary incontinence (stress/urge) in 11,526 US minority women from Women's health Initiative SHARe cohort. These women included 8,088 African American and 3,438 Hispanic American women. We report that the prevalence of overall urinary incontinence in these women aged 50-79 years was 55%, with 32% of women experiencing urge urinary incontinence, 29% of women having stress urinary incontinence and 8% of women had both. More African women reported having urge urinary incontinence while more Hispanic women reported having stress urinary incontinence. We show that rs2086297, located in the intron of PRCP, was significantly associated with stress UI (p-value<4.4×10-8) in trans-ethnic analysis. Each copy of the rs2086297 T allele showed around 20% reduction in the odds of stress urinary incontinence. Future studies on the UI of minority women may identify additional loci associated with urinary incontinence and will be helpful in broadening our understanding of urinary incontinence associated genetic pathways.

915W

Genome-wide association study of thyrotoxic periodic paralysis. *K.C. Chen[1], P.L. Chen[1,2,3,4], T.C. Chang[3,5].* 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan; 5) Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan.

Thyrotoxic periodic paralysis (TPP) is a unique disease characterized by episodic attacks of muscle weakness and hypokalemia in certain thyrotoxic individuals. TPP shows male predominance and huge inter-population prevalence difference; up to 10% of thyrotoxic males in Asia develop TPP, while only less than 0.1% thyrotoxic Caucasian males do. In 2010, Ryan *et al.* (Cell 140(1): 88-98) reported that mutations of the *KCNJ18* gene could be found in 33% of their patients (mainly of European descent), but not in controls. However, in 2012, Cheung *et al.* (Nat Genet 44(9): 1026-1029) published a genome-wide association study (GWAS) in ethnic southern Chinese and identify a single susceptibility locus at 17q24.3 near *KCNJ2*, with estimated odds ratio of 3.3. It is still an unsolved issue whether these two loci are genuine TPP-related loci, and whether there are more susceptibility loci to be identified. To address these issues, we conducted a GWAS using 41 TPP patients as cases and 725 thyrotoxic, non-TPP individuals as controls, all of them are ethnic Chinese in Taiwan. The genotyping platform was Axiom Genome-Wide CHB Array Plate, which contains 642,832 common (MAF>5%) SNPs in Han Chinese genome. We replicated the 17q24.3 association signals reported by Cheung *et al.*, with several SNPs showed p values smaller than 5×10^{-7} in the trend test and allelic test. There were also association signals at 2q31.2, 2q36.3, 9q22.1 and 18q21. Fine mapping of the novel loci is currently underway. On the other hand, we did not find association signals at or near *KCNJ18*. We are now performing Sanger sequencing of *KCNJ18* to search for possible multiple rare variants. Our results demonstrate that 17q24.3 is the genuine TPP susceptibility locus, at least in ethnic Chinese Han population. We also report several novel susceptibility loci awaiting replication.

916T

Common variations at chromosome 21q22 influence the risk of age-related nuclear cataract in Asians: the Singapore Epidemiology of Eye Diseases (SEED) Study. *C.Y. Cheng[1,2,3], J. Liao[1,3], P. Chen[2], X. Li[3,4], X. Wang[2], A.G. Tan[5], J.J. Wang[5,6], P. Mitchell[5], J.B. Jonas[7], S.M. Saw[2,3], C.C. Khor[1,8], E.S. Tai[2,9], T. Aung[1,3,10], Y.Y. Teo[2,4,11,12], T.Y. Wong[1,3,10].* 1) Department of Ophthalmology, National University of Singapore, Singapore; 2) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 3) Singapore Eye Research Institute, Singapore; 4) Department of Statistics and Applied Probability, National University of Singapore, Singapore; 5) Centre for Vision Research, Department of Ophthalmology, University of Sydney, New South Wales, Australia; 6) Center for Eye Research Australia, University of Melbourne, Melbourne, Australia; 7) Department of Ophthalmology, Medical Faculty Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany; 8) Human Genetics, Genome Institute of Singapore, Singapore; 9) Department of Medicine, National University of Singapore, Singapore; 10) Singapore National Eye Centre, Singapore; 11) Centre for Molecular Epidemiology, National University of Singapore, Singapore; 12) Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore.

Age-related cataract is the leading cause of visual impairment and blindness in the world. Although several mutations were identified as causative for congenital and juvenile cataracts, little is known about the genetic variants that influence the susceptibility of age-related cataract. We conducted a meta-analysis of genome-wide association studies (GWAS) on 2,369 Malays and 2,200 Indians aged 40 to 80 years enrolled in the Singaporean Epidemiology of Eye Diseases (SEED) Study. Participants underwent a comprehensive eye examination including slit-lamp lens photography and provided a blood sample for genotyping. Lens photographs were graded using the Wisconsin Cataract Grading System, with the grades ranging from 0.1 (the lowest grade) to 5.0 (the highest). Genotyping was performed with Illumina HumanHap610-Quad chips. We performed SNP imputation using the genotype data, together with the reference panels in the 1000 Genomes project (March 2012, GRCh37). The association between SNPs and the severity (i.e., grade) of nuclear cataract was assessed using linear regression analysis with adjustment for age, sex and genetic principal components. In the meta-analysis of GWAS in the Malay and Indian samples (n = 4,569), the strongest evidence was observed at chromosome 21q22 (β [the increase in nuclear cataract grade per risk allele] = 0.11, $P = 5.2 \times 10^{-9}$), reaching genome-wide significance. The results were similar after further adjustment for smoking status. Replication tests in two independent Chinese cohorts (n = 2,571) were performed. Using random effect models, the pooled estimates of the four cohorts confirmed the association at the 21q22 locus ($\beta = 0.08$, $P = 2.0 \times 10^{-9}$). To our knowledge, this is among the first GWAS to investigate age-related nuclear cataract. The results yield insights at the gene level for the pathogenesis of age-related cataract.

917F

Copy Number Variations are associated with Bone Mineral Density: A large-scale genome-wide analysis in the Framingham Study. *W. Chou[1], K. Nandakumar[1], D. Karasik[1], C. Liu[2], L. Cupples[2], D. Kiel[1], Y. Hsu[1].* 1) Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Boston, MA; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Bone mineral density (BMD) is a complex phenotype with high heritability. Our previous genome-wide SNP association study identified more than 56 BMD loci in Caucasian populations, explaining < 6% of BMD variation. Copy number variation (CNV), a type of genomic structural variation, accounts for > 20% variation of the human genomic structure between individuals. CNVs may explain the missing heritability of BMD. Thus, we conducted a genome-wide CNV association analysis with BMD in Framingham Study participants. This study included 7,451 adult Caucasians (4126 men and 3325 women) with mean age of 55 years. BMD at lumbar spine (LS) and femur neck (FN) was measured by dual energy X-ray absorptiometry. CNVs were estimated using Affymetrix 500K genotyping array and PennCNV package. A CNV was defined as a DNA segment longer than 1 kbp and composed of at least three consecutive genotyping probes. Under an additive genetic model, we employed a linear mixed effects model to account for family relatedness, and also adjusted for age, sex, estrogen usage, menopause status, cohorts within the Framingham Study and principal components for population stratification. To correct for multiple-testing, a genome-wide significant cutoff ($p < 4 \times 10^{-6}$) was defined by a false-discovery rate as less than one false-positive result among the genome-wide significant findings. Up to 6,398 individuals had valid CNV calls and information on covariates. Two CNVs (at chrom 2q14.2 and 6q25.3) and four (at chrom 5q23.1, 15q22.31, 17p13.1 and 17q21.2) were significantly associated with LS BMD and FN BMD, respectively. The length of the CNVs ranged from 133 kbp to 1,025 kbp. The most significantly associated CNV with $p = 1.9 \times 10^{-20}$ was located at chromosome 15q22.31 (*SMAD6*). The Smad6 protein inhibits signaling of bone morphogenetic proteins. Among 6,398 individuals included in the association analysis of the CNV, 5 had this CNV. Individuals with deleted CNVs had lower FN BMD. In addition, we identified three genome-wide suggestive CNVs associated with BMD at $p < 2.7 \times 10^{-6}$. In summary, we identified six genome-wide significant CNVs and three genome-wide suggestive CNVs associated with FN BMD or LS BMD in the Framingham Study. To replicate the CNV-BMD associations, we are performing analyses in independent samples. The newly identified CNVs in this study may provide additional biological information regarding the genetic determinants of BMD and may explain some of the missing heritability.

918W

Genetic associates of childhood wheezing phenotypes. *J.A. Curtin[1], D.C.M Belgrave[1,2], A. Custovic[1], A. Simpson[1].* 1) Centre for Respiratory Medicine and Allergy,Institute of Inflammation and Repair, University of Manchester, Manchester, Manchester, United Kingdom; 2) Centre for Health Informatics, Institute of Population Health, University of Manchester, UK.

Background: Much of the heritability of complex traits (e.g. asthma) remains unexplained by genome wide association studies (GWAS). One reason for this may be poor phenotype definition; most studies investigating genetic associates use phenotypes that encompass a number of different endotypes (e.g. doctor-diagnosed asthma). We recently identified four distinct childhood wheezing phenotypes (Transient early wheeze-TEW, Late-onset wheeze-LOW, Persistent controlled wheeze-PCW and Persistent troublesome wheeze-PTW) by joint modelling of parentally-reported and physician-confirmed wheeze. Methods: DNA was genotyped using Illumina 610 quad chips. Following standard QC we imputed additional genotypes (IMPUTE version 2.2.2) with the '1000 Genomes Phase I integrated variant set' reference genotypes; we excluded SNPs with INFO <0.4 and MAF <0.01. Genome-wide association study (GWAS) was performed for these four wheezing phenotypes (TEW, n=122; PTW, n=29; PCW; n=106; LOW, n=146), using Non Wheezers (n=469) as controls, and also using the traditional asthma definition of 'doctor diagnosed asthma' (192 cases, 610 controls). For each phenotype we used Ingenuity pathway analysis (IPA) to identify biological pathways common to genes found within 50KB of associated Single Nucleotide Polymorphisms (SNPs). Results: We identified distinct genome-wide significant ($p<5\times 10^{-8}$) associates for each of the phenotypes: TEW (86 SNPs), PTW (193 SNPs), PCW (126 SNPs), LOW (47 SNPs) and doctor diagnosed asthma (19 SNPs). The associated SNPs were unique for each phenotype except for 'chr12:98730665:D' which was significant in TEW and PCW. Using IPA we generated a network of 21 interconnected proteins important in mediating immune responses in relation to TEW, and found that many of the associated SNPs of PTW and PCW reside in genes related to respiratory disease. Conclusions: Childhood wheezing phenotypes may represent different diseases with unique genetic associates.

919T

Exploring the causes of heterogeneity in meta-analysis of genome-wide association studies. *H. Deng[1,2], Y. Pei[1,2], L. Zhang[1,2].* 1) Biostatistics and Bioinformatics, Tulane Univ, New Orleans, LA; 2) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Meta-analysis of Genome wide association (GWA) data has been widely used to detect genetic variants responsible for complex diseases. Between-study heterogeneity is an important issue when conducting meta-analysis. Exploring the extent and possible sources for heterogeneity between studies is important to the result interpretation. To exemplify these issues, we used data from seven GWA studies on obesity. We first investigated the extent and distribution of the heterogeneity across the whole genome. Among the 4,325,550 SNPs tested, the I2 inconsistency metric was different from 0 (no detectable heterogeneity) for 43.8% of the total SNPs; inconsistency was moderate to very large (I2>25%) for 18.4% of the total SNP. Significant heterogeneity (pQ<0.1) existed for 11.0% of the total SNPs. Heterogeneity was enriched in SNPs with small p-values. Among SNPs with p-value less than 1.0×10-5, 190 (41.4%) SNPs showed moderate or larger heterogeneity. Meta-regression analyses were performed for the 190 SNPs to explore potential causes of heterogeneity. Five moderators, which included average age, ethnicity, gender composition, study base and imputation accuracy, were constructed to test their effects on heterogeneity. Among the tested factors, ethnicity, average age and imputation accuracy had significant effects on the heterogeneity. After adjusting for all the five moderators, almost all of the significant heterogeneity (96.2%) was removed. Our results have significant implications for the study design and results interpretation of GWA meta-analyses.

920F

Cross-disease analysis using Immunochip reveals four new loci for celiac disease and rheumatoid arthritis. *J. Gutierrez-Achury[1], G. Trynka[2,3], S. Raychaudhuri[2,3,4], J. Greenberg[5], D. Diogo[2,3], R. McManus[6], R.M Plenge[2,3], C. Wijmenga[1], A. Zhernakova[1], Celiac Disease Immunochip Consortium, RACI consortium.* 1) Department of Genetics, University Medical Centre Groningen - University of Groningen, Groningen, Netherlands; 2) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 3) Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA; 4) Institute of Inflammation and Repair, University of Manchester, Manchester, UK; 5) New York University School of Medicine, New York, NY; 6) Department of Clinical Medicine and Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland.

Background: Autoimmune diseases are complex and heterogeneous but are known to share a subset of loci and pathways. Celiac disease (CeD) and rheumatoid arthritis (RA) are two examples in which GWAS and the Immunochip have identified strong associations to each trait and substantial overlap between loci. The Immunochip array was specifically designed to refine GWAS results. We aimed to jointly analyze the largest CeD and RA datasets available to date and genotyped using the Immunochip to identify new unique or shared loci and to fine-map the known regions. Methods: We meta-analyzed 6 European CeD cohorts (12,400 cases, 14,257 controls) and 7 European and North American RA cohorts (13,877 cases, 18,120 controls). We determined common SNPs between the datasets, aligned them in the same strand, removed duplicates and related individuals, and performed a principal component analysis (PCA) per cohort. We performed logistic regression by adjusting for first five PCAs per cohort. We then conducted meta-analysis using a fixed-effect model with inverse variance weighting. Results: From the joint meta-analysis of all 13 cohorts, 28 loci reached the threshold p-value of association, 1×10^{-8}, and 17 more reached a suggestive p-value of 1×10^{-5}. Of those that reached genome-wide significance, we report the new locus *NCF2* ($OR_{meta} = 1.25$, $p\text{-}val_{meta} = 1.24\times10^{-14}$) represented by a probably damaging missense mutation with a polyphen-2 score of 0.98. Three other loci (*AFF3*, *PTPN22* and *CXCR5*) had been previously reported in one disease but were now identified in the second phenotype and in the combined analysis. Using co-expression analysis based on approximately 80,000 microarrays, it was possible to determine that the new gene *NCF2* is involved in the defense response to Gram-positive bacteria, which points to its importance in the innate immune response and microbiota regulation. These aspects should be further analyzed by functional studies.

921W

A Large Scale Genome Wide Association Study of Asthma in the 23andMe Cohort. *D. Hinds, C. Tian, A.K. Kiefer, J.L. Mountain, N. Eriksson, J.Y. Tung.* 23andMe, Inc., Mountain View, CA.

Asthma and allergic disease are common, chronic conditions with substantial public health burdens. We carried out a genome-wide association study of self-reported asthma in the 23andMe participant cohort, including more than 17,000 cases and 87,000 controls with European ancestry, imputed against 1000 Genomes reference haplotypes. In addition to replicating established asthma GWAS findings, we see genome-wide-significant associations with several loci that have been associated with other forms of atopic disease, including rs17673553 in *CLEC16A*, rs12542017 near *ZBTB10*, and rs3001426 near *STAT6*. Using conditional logistic regression, we see evidence for multiple independent associations at the *IL1RL1* locus on chromosome 2, the *TSLP* locus on chromosome 5, and the *IL33* locus on chromosome 9.

We find novel associations with rs2070902 in *FCER1G*, encoding the gamma subunit of the high affinity IgE receptor; with rs4707609 in *BACH2*, or basic leucine zipper transcription factor 2; and with rs62192043 near *D2HGDH*, or mitochondrial D-2-hydroxyglutarate dehydrogenase, and *GAL3ST2*, or galactose-3-O-sulfotransferase. Variation in *BACH2* has previously been associated with a variety of autoimmune diseases. *GAL3ST2* is expressed in goblet cells and is involved in production of sulfated mucins. We also find evidence of associations in or near several genes associated with allergies in the 23andMe cohort, including *ID2*, *LPP*, *TLR1*, and *PTGER4*.

Most genetic associations with asthma and allergy are shared, with similar effect sizes, and identifying risk factors that distinguish between different allergic disease types seems more challenging. We further explore this issue with analyses of the susceptibility loci stratified by disease subgroups, determined by age of onset, triggers, and severity, and with patterns of allergy symptoms.

922T

A genome-wide association study highlights multiple variants associated with Epstein-Barr virus load in the 1000 Genomes and HapMap lymphoblastoid cell lines. *C.J. Houldcroft[1], J.Z. Liu[1], A. Gall[1], D. Frampton[2], C.A. Anderson[1], P. Kellam[1,2].* 1) Wellcome Trust Sanger Institute, Cambridge, Cambs, United Kingdom; 2) UCL/MRC Centre for Medical Molecular Virology, Department of Infection, University College London, Cleveland Street, London W1T 4JF, UK.

Epstein-Barr virus (EBV) infects 95-98% of adults world-wide, persists for life and causes benign infectious mononucleosis in ~50% of young adults upon primary infection. EBV is also linked to more severe conditions such as Burkitt's lymphoma, nasopharyngeal carcinoma and post-transplant lymphoproliferative disorder. Persistent life-long infection can be modelled using lymphoblastoid cell lines (LCLs), human B cells immortalised with EBV. It is unknown to what extent host genetic variants influence EBV load within LCLs and whether variation in EBV load affects gene expression studies of these cell lines.

We measured relative EBV genome load in LCLs from 1000 Genomes and HapMap projects and performed a heritability analysis and genome-wide association study (GWAS) on 899 of these individuals. This combined sequencing data from 1000 Genomes Project Phase 1 with genotype data from Illumina Omni 2.5M and HapMap Phase III. We controlled for population structure using FaST-LMM-Select. After filtering, there were 30M SNPs included in the final analysis.

Estimates of heritability from genome-wide covariance and trio methods were consistently between 30-50%. The top SNP identified in the GWAS was rs182621445 on chromosome 6 (p=5.55E-10), downstream of DAAM2, a gene over-expressed in nasopharyngeal carcinomas. Genome-wide significant results were also seen for SNPs close to or located within genes CD109 (a negative regulator of TGFβ, which has been implicated in immunity to infectious mononucleosis), ZNF839 and WAC. Gene-based association analysis (VEGAS) identified TBXAS1 as most strongly associated with EBV load (p=8.39E-08); followed by a cluster of zinc-finger proteins (ZKSCAN4, ZSCAN9, ZSCAN31) on chromosome 6 (p=1.55E-07). Future work will characterise the frequency of these mutations in a population of UK university students with infectious mononucleosis, as well as functional validation of the role of these variants through gene expression studies, over-expression in lentiviral vectors and RNAi gene knockdown.

923F

Identification of multiple genetic susceptibility loci in Takayasu's arteritis. *T. Hughes[1], G. Saruhan-Direskeneli[2], P. Coit[1], J.M. Guthridge[3], J.A. James[3], P.A Merkel*[4], H. Direskeneli**[5], A.H. Sawalha[1], *on behalf of Vasculitis Clinical Research Consortium;**on behalf of Turkish Takayasu Study Group.* 1) Division of Rheumatology, University of Michigan, Ann Arbor, MI, USA; 2) Department of Physiology, Istanbul University, Istanbul Faculty of Medicine, Istanbul, Turkey; 3) Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; 4) Division of Rheumatology, University of Pennsylvania, Philadelphia, PA, USA; 5) Department of Rheumatology, Marmara University, Faculty of Medicine, Istanbul, Turkey.

Background: Takayasu's arteritis is a rare inflammatory disease of large arteries. The etiology of Takayasu's arteritis remains poorly understood, but genetic contribution to the disease pathogenesis is supported by the confirmed genetic association with HLAB*52. Genomic studies in Takayasu's arteritis have not been previously performed. Methods: We genotyped ~200,000 genetic variants in two ethnically divergent Takayasu's arteritis cohorts from Turkey (339 patients and 516 controls) and North America (112 patients and 599 controls) using a custom designed genotyping platform (Immunochip). Additional genetic variants and the classical HLA alleles were imputed and analyzed. Results: We identified and confirmed two independent susceptibility loci within the HLA region (r2<0.2): HLA-B/MICA (rs12524487, OR= 3.29, P= 5.57×10^{-16}), and HLADQB1/ HLA-DRB1 (rs113452171, OR= 2.34, P= 3.74×10^{-9}, and rs189754752, OR=2.47, P= 4.22×10^{-9}). In addition, we identified and confirmed a novel genetic association between Takayasu's arteritis and the FCGR2A/FCGR3A locus on chromosome 1 (rs10919543, OR= 1.81, P= 5.89×10^{-12}). The risk allele in this locus results in increased mRNA expression of FCGR2A. We also established the genetic association between IL12B and Takayasu's arteritis (rs56167332, OR= 1.54, P= 2.18×10^{-8}). An association with an additional locus on chromosome 21q22 downstream of PSMG1 did not pass the threshold for genome-wide significance and requires replication (P=4.39×10^{-7}). Conclusion: We established multiple genetic susceptibility loci for Takayasu's arteritis with a genome-wide level of significance including two independent susceptibility loci in the HLA region, and disease susceptibility loci in FCGR2A/FCGR3A and IL12B.

924W

Genome-wide association confirms TCF4 as a major locus for Fuchs Endothelial Corneal Dystrophy and identifies novel loci. *S.K. Iyengar[1,4], N. Afshari[2], R.P. Igo, Jr.[1], Y-J. Li[3], J.L. Lass[4,1], G. Klintworth[5] for the FECD Consortium.* 1) Dept Epid/Biostat, Case Western Reserve Univ, Cleveland, OH; 2) Dept. Ophthalmology, UCSD, San Diego, CA; 3) Department of Biostatistics & Bioinformatics, Duke University Medical Center Center for Human Genetics, Duke University Medical Center, Durham, NC; 4) Department of Ophthalmology and Visual Sciences, Case Western Reserve Univ, Cleveland, OH; 5) Duke Eye Center, Duke University Medical Center Durham, NC.

Background. Fuchs Endothelial Corneal Dystrophy (FECD) is a progressive, age-dependent disease of the corneal endothelium, associated with decreased vision. Women are more commonly affected than men. Histological hallmarks of the disease include decreased endothelial cells, thickened Descemet membrane (DM) due to excessive accumulation of extracellular matrix, and the formation of excresences on DM called corneal guttae. FECD is one of the most common indications for corneal transplants in developed countries, with a prevalence of approximately 4%;, and heritability in the range of 30-40%;. **Methods.** We organized a multi-center consortium, with a large cohort of patients uniformly graded for FECD. We conducted a genome-wide association study in 1452 unrelated cases and 2535 controls using the Illumina Omni 2.5 chip. Primary analyses consisted of association testing via logistic regression, controlling for population structure via six principal components, and covariates age and sex. We also conducted sensitivity analysis for histopathology validated cases, and determined whether female versus male gender showed differences in odds ratios at the best loci. Genetic risk models were built using forward selection and were compared using the area under the receiver operating characteristic (AUC). **Results.** Transcription factor 4 (TCF4) emerged as a major locus for FECD (per-allele odds ratio = 5.78; p= 3.9×10^{-146}, confirming a prior small GWAS (Baratz et al. (2010) NEJM 363, 1016). Besides TCF4, five novel loci on chromosomes 1q and 11p also reached genome-wide significance with minor-allele odds ratios ranging from 0.75 to 1.43; the latter are being validated in independent samples. Sensitivity analysis of histopathology validated cases confirmed all loci. TCF4 shows genome-wide significance in sex-specific analysis. The chr1q and chr11p loci remain strongly significant in the Female subset ($P < 0.001$ with min P= 2.8×10^{-8}), but two loci show promising results in the Male subset ($P< 0.01$, min P=2.6×10^{-6}). Markers at all six loci yielded an AUC of 0.77, but the majority of the predictive value came from TCF4 (AUC=0.75). **Conclusions.** The FECD Genetics Consortium has conducted the largest genome-wide association study to date on FECD, leading to confirmation of the role of TCF4 on FECD. Five novel loci on chromosomes 1 and 11 were identified. Further analyses of these genes should help elucidate different pathways in FECD pathogenesis.

925T

Sub-phenotype Mapping in Systemic Lupus Erythematosus Identifies Multiple Novel Loci Associated with Circulating Interferon Alpha. *S. Kariuki[1], Y. Ghodke-Puranik[2], J. Dorschner[2], B. Chrabot[1], J. Kelly[3], B. Tsao[4], R. Kimberly[5], M. Alarcón-Riquelme[3,6], C. Jacob[7], L. Criswell[8], K. Sivils[3], C. Langefeld[9], J. Harley[10], A. Skol[1], T. Niewold[2].* 1) University of Chicago, Chicago, IL; 2) Mayo Clinic, Rochester, MN; 3) Oklahoma Medical Research Foundation, Oklahoma City, OK; 4) University of California, Los Angeles, CA; 5) University of Alabama, Birmingham, AL; 6) Pfizer - Universidad de Granada - Junta de Andalucia Centre for Genomics and Oncological Research (GENYO), Granada, Spain; 7) University of Southern California, Los Angeles, CA; 8) University of California, San Francisco, CA; 9) Wake Forest University, Winston Salem, NC; 10) Cincinnati Children's Hospital Medical Research Center, Cincinnati, OH.

Systemic Lupus Erythematosus (SLE) is a phenotypically heterogeneous complex disease. Our previous work has documented significant genetic heterogeneity, with some well-validated risk factors demonstrating strong sub-group effects. Approximately 50% of patients have high circulating levels of interferon alpha (IFN-α), and many lines of investigation support IFN-α as a heritable and primary causal factor in human SLE. This study aims to genetically map the serum IFN-α trait in SLE patients, allowing for novel genetic discovery in this heterogeneous disease. GWAS data were obtained from 450 European ancestry SLE cases who were genotyped as part of the Systemic Lupus Erythematosus Genetics (SLEGEN) study. Genotypes were generated on the Illumina Infinium HumanHap300 genotyping platform, and principal component analysis was used to correct for population stratification. Sera were obtained from each of these subjects, and IFN-α activity was measured using a sensitive and specific reporter cell assay. Associations between genome-wide SNP markers and serum IFN-α were detected using logistic regression conditioned on the principal components to control for structure. IFN-α activity was studied as a categorical trait. Patients with IFN-α levels two standard deviations above the mean of healthy controls were designated as high IFN-α, and the remainder as low IFN-α. Top novel associated loci in the GWAS screen include multiple SNPs in the C7orf57, PRKG1, ANKRD44, and PNP loci. Interestingly, three of the five top SNPs are missense SNPs. Strong association signals were also detected in chromosomes 12 and 14. Genome-wide imputation using SNPs from the 1000 Genomes Project did not yield additional significant association signals beyond those identified by the directly genotyped SNPs. These novel loci have not been previously associated with SLE in case-control analyses. This supports the concept that studying pathogenic subgroups within the complex disease SLE will be important in our efforts to fully map disease susceptibility. These loci could provide novel therapeutic targets in the IFN-α pathway and assist in personalizing therapy in this disease.

926F

Presence of genome-wide associations for Parkinson's disease on the X chromosome. *M.F. Keller[1, 2], M.A. Nalls[1], A.B. Singleton[1], International Parkinson's Disease Genomics Consortium (IPDGC).* 1) Laboratory of Neurogenetics, NIH NIA, Bethesda, MD; 2) Department of Biological Anthropology, Temple University, Philadelphia, PA.

GWA studies have identified over 1,200 associations with p-values of < 5x 10-8 for over 200 traits, yet only 7 such associations have been reported on the X chromosome. While a plethora of GWA data is available on the X chromosome, it is currently underutilized. This is largely due to exclusion of X chromosome variants from analyses even though these regions are assayed on many current microarray platforms. Employing genotype and phenotype data from 6 European ancestry cohorts, we imputed the genotypes of 9,511 controls and 8,497 cases to test for genome wide associations to Parkinson's disease status, and to the quantitative trait age of Parkinson's disease onset. Our analyses identified a number of statistically significant SNPs clustered in three regions of the X chromosome, which suggest risk to Parkinson's disease development. Two of these regions contain protein-coding transcripts expressed in the testis, and two contain non-coding micro RNA. Replication is underway in a supplementary cohort containing over 14,000 individuals. In addition to identifying novel genetic variants associated with Parkinson's disease, our results indicate that utilization of X-chromosome GWA data can facilitate a more comprehensive understanding of complex disease architecture in a previously unexplored region of the genome. Continued development of study designs examining the sex chromosomes and disease status are necessary, and our work provides a first step towards documenting the applicability of GWA analyses to complex diseases in the X chromosome.

927W

Feasibility of routine bi- & tri-variate exhaustive analysis of Case-Control GWAS. *A. Kowalczyk[1,2], Q. Wang[1,2], B. Goudey[1,2], H. Ferra[1], D. Rawlinson[1], J. Bedo[1,2], A.C. Kowalczyk[2], G. Abraham[3], F. Shi[1,2], L. Gor[1], R.M. Campbell[1,7], C. Iglesias[1], A. Zarnegar[1,5], E. Kikianty[1,6], G.J. Macintyre[1,2], I. Haviv[1,4], C.S. Ong[1,2], M. Inouye[3].* 1) NICTA, National ICT Australia, Parkville, Victoria, Australia; 2) Computing and Infromation Systems, The University of Melbourne, Australia; 3) Pathology, The University of Melbourne, Australia; 4) Bar Ilan University, Israel; 5) Victorian AgriBioscience Centre, Latrobe University, Australia; 6) University of The Witwatersrand, South Africa; 7) Electrical and Electronic Engineering, The University of Melbourne.

Genome-wide association studies (GWAS) genotype a large number of single nucleotide polymorphisms (SNPs) and attempt to determine their association with a given phenotype. Typically, the follow up data analysis examines the association of each SNP individually. Yet many common diseases have complex aetiologies that involve combinations of SNPs from different genes which may vary within the population of affected individuals. The search for such combinations is hindered by such extreme computational and statistical difficulties, that many researchers have declared that exhaustive search for 3-way interactions in modern human GWAS will not be practical. We show that such pessimism is unfounded. We recently demonstrated a model-free, ROC-based strategy called GWIS (Goudey et. al, BMC Genomics, 2013) that exhaustively searches medium-size GWAS studies for pairwise interactions within minutes. For all seven WTCCC datasets, we detected nearly all top hits reported previously in the literature within a short list of a few thousand candidate pairs. GWIS is freely available as a web service at: bioinformatics.research.nicta.com.au/software/gwis/. We have now extended GWIS to exhaustive search for 3-way interactions. For illustration, an exhaustive 3-way analysis of Celiac disease GWAS from UK containing ~310K SNPs and 2200 samples using a cluster of 200 GPUs requires 7 days of computing time. To our knowledge this is the first time such an analysis has been shown to be practical. The runtime reduces significantly for more targeted analysis, of a specific DNA region or of a preselected set of SNPs. In particular, exhaustive filtering through all triplets in ~2500 SNPs, including the extended MHC region, required <3 minutes on a standard PC with a single GTX470 NVIDIA GPU. Using the above discovery study, we have found hundreds of triplets having statistically significant association with disease, replicating across four independent Celiac cohorts (a second UK, Dutch, Finnish and Italian cohorts of 5603, 1649, 2476 and 1040 samples, respectively). In particular, within the top ten triplets we observed an area under the ROC curve of high 80%'s and variance explained for balanced data up to 50%. Notably, the ranking of genotypes according to relative risk for many of the top triplets was preserved across cohorts. In conclusion, we demonstrate that exhaustive tri-variate search for epistatic interactions is possible with existing computing hardware.

928T

Genetic variants associated with total lung capacity in chronic obstructive pulmonary disease. *J.H. Lee[1,2], G.M. Hunninghake[3], M.N. McDonald[1], M.H. Cho[1,3], M. Hardin[1,3], E.S. Wan[1,3], P.J. Castaldi[1], A. Gulsvik[4], P. Bakke[4], E.K. Silverman[1,3], C.P. Hersh[1,3], The COPDGene and ECLIPSE investigators.* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Department of Internal Medicine, School of Medicine, Ewha Womans University, Seoul, Korea; 3) Pulmonary and Critical Care Division, Brigham and Women's Hospital, Boston, MA; 4) Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway.

Background: Chronic obstructive pulmonary disease (COPD) is characterized by expiratory flow limitation, resulting in air trapping and lung hyperinflation. Hyperinflation is closely related to exercise tolerance and quality of life in COPD patients. It could also be used as a distinguishing feature to explain COPD heterogeneity. The aim of this study was to identify genetic variants associated with hyperinflation in COPD. Methods: We performed genome-wide association studies in three cohorts: COPDGene; the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE); and GenKOLS (Bergen, Norway). All subjects were Caucasians and had at least moderate COPD, defined by a ratio of forced expiratory volume in 1 second to forced vital capacity (FEV_1/FVC) <0.7 and FEV_1 <80% predicted in post-bronchodilator spirometry. Total lung capacity (TLC) was used as a marker of hyperinflation and calculated by using volumetric computed tomography scans of the chest. Genotyping in each cohort was performed on the Illumina platforms, with additional markers imputed using 1000 Genomes CEU or EUR reference panels. To find genetic variants associated with TLC, we used linear regression models adjusting for age, sex, pack-years of smoking, height, and principal components for genetic ancestry in PLINK version 1.07, with meta-analysis of the results using METAL. Results: The most significant loci were on 11q24.2 (β= -0.19, p=3.55×10^{-7}), 8p23.2 (β= -0.16, p=4.22×10^{-7}) and 5q23.1 (β= -0.17, p=6.14×10^{-7}) in the COPDGene cohort; 19p13.2 (β= -0.50, p=6.95×10^{-7}), 13q34 (β= -0.95, p=9.34×10^{-7}), and 12q24.33 (β= -1.22, p=9.64×10^{-7}) in the ECLIPSE cohort; and 8p11.21 (β= -2.58, p=2.51×10^{-7}), 3q24 (β= -1.87, p=4.88×10^{-7}) and 9q34.13 (β= -1.97, p=5.05×10^{-7}) in GenKOLS cohort. In a meta-analysis of these three cohorts, the most significant loci were on 17p13.2 (β= -0.17, p=1.82×10^{-7}) and 19p13.12 (β= 0.13, p=2.36×10^{-7}). Conclusions: Although no single genetic variant reached genome-wide significance, there were several suggestive associations with TLC in three COPD cohorts. Expanding to another cohort could improve statistical power to elucidate associations between genetic variants and hyperinflation-related phenotypes. Funding: NIH R01 HL089856 and R01 HL089897.

929F

The genome-wide association analysis and meta-analysis of sciatica in two Finnish populations. *S. Lemmelä[1], S. Solovieva[1], R. Shiri[1], I. Seppälä[2,3], M. Heliövaara[4], M. Kähönen[2,3], M. Juonala[5,6], J. Viikari[5,6], O. Raitakari[5,6], T. Lehtimäki[2,7], E. Viikari-Juntura[1], K. Husgafvel-Pursiainen[1].* 1) Finnish Institute of Occupational Health, Helsinki, Finland; 2) Tampere University Hospital, Tampere, Finland; 3) University of Tampere, Tampere, Finland; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) University of Turku, Turku, Finland; 6) Turku University Hospital, Turku, Finland; 7) Fimlab Laboratories, Tampere, Finland.

Sciatic pain is one of the leading causes of disability in the working age population. A strong genetic component in family and twin studies as well as a multifactorial etiology as documented in epidemiological studies have been shown for sciatica. A genome-wide association study was conducted in two large population-based Finnish cohorts informative for sciatica. In our current study, data containing more than 500 000 genotyped SNPs and several million imputed SNPs as based on 1000 Genomes imputation reference constructed by the IMPUTE program were analysed using PLINK and SNPTEST, respectively. Results were adjusted for age and gender. Finally, meta-analysis using result data of both cohorts was conducted using GWAMA. In summary, our preliminary results revealed novel candidate regions for sciatica. These results, with several highly interesting regions warranting further investigation, will be discussed. The study received financial support from the Academy of Finland (project no 129364, MSDs@Lifecourse, The SALVE Program).

930W

Replication of GWAS results for dental caries in the permanent dentition. *D. Lewis[1], J. Shaffer[1], E. Feingold[1,2], M. Cooper[3], R. Weyant[4], D. McNeil[5], R. Crout[5], S. Reis[8], A. Vieira[4], M. Vanyukov[7], M. Marazita[1,3,6].* 1) Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 2) Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 3) Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, Pittsburgh, PA; 4) School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 5) Dental Practice and Rural Health, West Virginia University, Morgantown, WV; 6) Clinical and Translational Science, School of Medicine, University of Pittsburgh, Pittsburgh, PA; 7) University of Pittsburgh School of Pharmacy, Pittsburgh PA; 8) Health Policy Institute, University of Pittsburgh, PA.

There is increasing evidence that variation in the risk for dental caries (i.e., tooth decay) includes a genetic component. We previously published several GWAS studies for dental caries in the permanent dentition (adult teeth), primarily based on the dataset from the GENEVA dental caries study. Those publications considered several different phenotypes, including overall number of carious teeth in the mouth(DMFT)and numbers of caries lesions in specific types of tooth surfaces. Interesting nominated genes included BMP4, BCOR, LYZL2, and ABCG2. BMP4 is known to initiate and regulate the repair of carious tissue. Mutations in BCOR cause oculofaciocardiodental syndrome. In the current work, we did follow-up genotyping in suggestive regions in four study populations totaling approximately 2000 individuals in order to replicate and fine-map our original results.

Grants
R01-DE014899
U01-DE018903.

931T

Negative-regulation-of-apoptotic-process identified to be associated with appendicular lean mass through meta-analysis of pathway-based genome-wide association analysis. *J. Li[1], C. Xu[1], YF. Pei[1,2], Q. Tian[1], H-W. Deng[1,2].* 1) Center for Bioinformatics and Genomics, Biostatistics and Bioinformatics, Tulane University, New Orleans, LA, 70112 USA; 2) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

Background: Sarcopenia, characterized by degenerative loss of skeletal muscle mass and strength with aging, is a major public health problem. It is a heritable disease with the heritability of muscle mass and strength being estimated as high as ~60%. Body lean mass (LM) and appendicular lean mass (aLM, the total muscle mass in limbs) are often used as measures for sarcopenia. Recent studies, especially, genome-wide association studies (GWAS), had linked a few genes to LM or aLM. However, they only account for a small proportion of the total LM/aLM variation. Two potential approaches for better identification of genes, especially those with relatively small effects, related to LM/aLM are meta-analyses and pathway-based analysis. In this study, we had conducted combined meta-analysis and pathway-based association analysis for LM and aLM. Methods: Using seven population samples with over 10,000 subjects and leveraging on the genomic variant information from 1000 Genomes Project, we performed GWAS on the imputed genotypes within each sample. We then conducted pathway-based analyses on the GWAS results through GenGen, before conducting meta-analysis on the pathway-based analysis results. Results: The negative regulation of apoptotic process (NROAP) pathway was identified to be the most significant pathway associated with aLM (raw p-value 1.06*10-8 for 885 pathways after QC). The pathway had q-value less than 0.05 in the majority of the samples in which aLM measurements were available. Previous studies had indicated that NROAP included genes for negative regulation of muscle cell and mesenchymal cell apoptotic process, potentially affecting the muscle mass changes in the human body. Conclusion: In this study, we identified the potential role of NROAP pathway in human aLM variation and sarcopenia. These results provided new insights into the genetic basis of osteoporosis, and potential targets for further functional analysis.

932F

Cannabis Dependence Is Associated with Genetic Variants in Genes Linked to Schizophrenia. *M. Lohsen[1, 2, 3], R. Sherva[1, 2], R. Koesterer[1], H. Kranzler[4], H. Zhao[5], L. Almasy[6], L. Farrer[1, 2, 3], J. Gelernter[7].* 1) Boston University School of Medicine, Department of Medicine, Section of Biomedical Genetics, Boston, MA; 2) Boston University School of Medicine, Department of Medicine, Graduate Program in Molecular Medicine, Boston, MA; 3) Boston University, Transformative Training Program in Addiction Sciences, Boston, MA; 4) University of Pennsylvania School of Medicine, Department of Psychiatry and VISN4 MIRECC, Philadelphia VAMC, Philadelphia, PA; 5) Yale School of Public Health, New Haven, CT; 6) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 7) Yale University School of Medicine, Departments of Psychiatry, Genetics and Neurobiology, New Haven, CT.

Background: Cannabis dependence (CaD) is a major public health problem, but little is known about the genetic risk factors. A case-control GWAS on a sample of 708 cases and 2346 exposed controls from the Study of Addiction: Genetics and Environment (SAGE) revealed suggestive evidence of association with the *ANKFN1* gene (Agrawal *et al.*, 2011), which were not genome-wide significant (GWS). Our study provides data from a larger sample with a discovery data set and meta-analysis using results from SAGE. **Methods:** To determine potential links between genetic variants and addiction, users of various substances (alcohol, cocaine and opioids) were recruited in multiple U.S. cities. Many had other comorbid substance dependencies as well, including CaD. The Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) was administered for psychiatric diagnosis. 988,306 autosomal SNPs (889,659 usable after QC) were genotyped on the IlluminaOmni1-Quad microarray. IMPUTE2 was used to impute the subjects' genomes onto the 1K Genomes reference panel and yielded ~30 million SNPs for analysis. A GWAS was performed using the summation of the 7 DSM-IV CaD diagnostic criteria in the discovery and replication samples and combined by meta-analysis. Samples consisted of cannabis-exposed controls, cases meeting 3 or more of the 7 DSM-IV diagnostic criteria and others (unexposed or meeting fewer than 3 criteria). African Americans ((AAs) discovery: N=3318; 895 cases, 1470 controls, 953 others; SAGE: N=1311; 289 cases, 796 controls, 226 others) and European Americans ((EAs) discovery: N=2379, 781 cases, 938 controls, 660 others; SAGE: N=2752; 458 cases, 1589 controls, 705 others) were analyzed separately in each subsample and then combined by meta-analysis. **Results and Discussion:** We identified several GWS and near GWS associations. The most robust finding was with relatively common variants (MAF~10%) in the *PI4K2B* gene for both AAs and EAs (rs313544; p_{meta}=2.93E-08). The direction of effect was the same for all 4 groups (AAs and EAs, both in discovery and SAGE). This gene was associated with schizophrenia (Houlihan *et al.*, 2009). After *NRG1* (Han *et al.*, 2012), this is the second schizophrenia-associated gene we found to be associated with CaD as well. Several other common SNPs approached, but did not reach, GWS, in one or more of the samples. This study presents additional evidence for biological convergence between schizophrenia and substance dependence traits.

933W

Examining genetic risk factors for progression of chronic kidney diseases: genome-wide pathway analysis of PediGFR. *L. Luo[1], P.A. Kanetsky[2], J. Gupta[2], N. Mitra[2], E. Wühl[3], A. Köttgen[4], S.L. Furth[2,5], B.A. Warady[6], F. Schaefer[3], C. Wong[7].* 1) Dept. of Internal Med., Univ. of New Mexico (UNM), Albuquerque, NM, US; 2) Dept. of Biost. and Epi., The Perelman School of Med. at the Univ. of Pennsylvania, Philadelphia, PA, US; 3) Ped. Nephrology Div., University of Heidelberg, Heidelberg, DE; 4) Renal Div., Univ. Med. Center Freiburg, Freiburg, DE; 5) Dept of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, US; 6) Ped. Nephrology Division, Children's Mercy Hospital, Kansas City, MO, US; 7) Ped. Nephrology Div., UNM Children's Hospital, Albuquerque, NM, US.

Background: Chronic kidney disease (CKD) is a worldwide public health issue. The PediGFR study is an international effort to determine genetic factors associated with kidney function decline in children with CKD via the prospective cohorts of the CKiD, ESCAPE and 4C studies. In addition to the classical GWA single SNP analysis, we are performing a complementary pathway analysis to investigate the role of pre-specified biological pathways in pediatric CKD progression. **Methods**: Longitudinal assessments of either measured or estimated Glomerular Filtration Rate (GFR), a measure of kidney function, were determined in the pediatric CKD cohorts. Among those providing consent for genetic testing, DNA was genotyped using the Illumina® Omni 2.5 BeadChip. Principal component analysis was used to assess subjects' genomic ancestry. Linear regression models adjusting for age, sex, baseline GFR, and population stratification were performed to determine SNP associations with GFR slope. We evaluated the enrichment of GFR slope associations for pathways as grouped in multiple databases (GO Term, KEGG, PANTHER, BIOCARTA, INGENUITY and REACTOME) using Meta-Analysis Gene-set Enrichment of variaNT Associations(MAGENTA) method. **Results**: Among the 444 subjects genotyped in the CKiD study and 1006 subjects genotyped in the ESCAPE/4C study, 267 subjects in the CKiD study and 885 subjects in the ESCAPE/4C study with European ancestry were available after QC. Based on the 267 CKiD subjects, we identified 87 pathways that were nominally significantly enriched ($P<0.05$) for genes associated with GFR slope. The enrichment for four pathways (Autophagy, Translational Regulation, Receptor Mediated Endocytosis, Toll Receptor) was robust to different pathway annotations across databases. However, none of the pathways were significant after correcting for multiple comparisons. **Conclusions**: We identified pathways that were enriched for genes associated with CKD progression in the CKiD cohort with modest significance. We are in the process of completing single SNP and pathway analysis based on the GWAS meta-analysis of the CKiD, ESCAPE and 4C studies. The top enriched pathway findings will be tested for replication in the adult Chronic Renal Insufficiency Cohort study. Pathway analysis of PediGFR has the potential to provide insight into the genetic architecture of CKD progression in children by identifying functionally-related biological pathways enriched for multiple genetic risk factors.

934T

Systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *J.E. Martin[1], S. Assassi[2], L.M. Diaz-Gallo[1], J.C. Broen[3], X. Zhou[2], J.D. Reveille[2], O. Gorlova[4], B.P.C. Koeleman[5], T.R.D.J. Radstake[3], T. Vyse[6], M.D. Mayes[2], M.E. Alarcon-Riquelme[7], J. Martin[1], SLEGEN.* 1) Instituto de Parasitologia y Biomedicina Lopez-Neyra, CSIC, Granada, Granada, Spain; 2) The University of Texas Health Science Center-Houston, Houston, Texas, USA; 3) Department of Rheumatology, Clinical Immunology and Translational Immunology, University Utrecht Medical Center, Utrecht, The Netherlands; 4) Department of Epidemiology, M. D. Anderson Cancer Center, Houston, Texas, USA; 5) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Divisions of Genetics and Molecular Medicine and Division of Immunology, Infection and Inflammatory Disease, King's College London, Guy's Hospital, London, UK; 7) Centro de Genómica e Investigación Oncológica (GENYO) Pfizer-Universidad de Granada-Junta de Andalucía, Granada, Spain.

Systemic sclerosis (SSc) and systemic lupus erythematosus (SLE) are two archetypal systemic autoimmune diseases which have been shown to share multiple genetic susceptibility loci. In order to gain insight into the genetic basis of these diseases we performed a pan-meta-analysis of two genome-wide association studies (GWAS) together with a replication stage including additional SSc and SLE cohorts. This increased the sample size to a total of 21,109 (6,835 cases and 14,274 controls). We selected for replication 19 SNPs from the GWAS data. We were able to validate *KIAA0319L* (P = 3.31×10-11, OR = 1.49) as novel susceptibility loci for SSc and SLE. Furthermore, we also determined that the previously described SLE susceptibility loci *PXK* (P = 3.27×10-11, OR = 1.20) and *JAZF1* (P = 1.11×10-8, OR = 1.13) are shared with SSc. Supporting these new discoveries, we observed that *KIAA0319L* was overexpressed in peripheral blood cells of SSc and SLE patients compared to healthy controls. With these, we add three (*KIAA0319L*, *PXK* and *JAZF1*) and one (*KIAA0319L*) new susceptibility loci for SSc and SLE, respectively, increasing significantly the knowledge of the genetic basis of autoimmunity.

935F

Replication Study of Age-Related Macular Degeneration Susceptible Gene Using Large Genome-Wide Association Study of Japanese.
M. Miyake[1,2], K. Yamashiro[1], H. Nakanishi[1,2], I. Nakata[1,2], Y. Akagi-Kurashige[1,2], K. Kumagai[1], M. Oishi[1], A. Oishi[1], N. Gotoh[1], A. Tsujikawa[1], M. Saito[3], Y. Kurimoto[4], T. Kawaguchi[2], CC. Khor[5], CY. Cheng[5], TY. Wong[5], R. Yamada[2], F. Matsuda[2], N. Yoshimura[1], the Nagahama Study Group. 1) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Center for Genomic Medicine/Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Fukushima Medical School. Fukushima, Japan; 4) Kobe City General Hospital, Kobe, Japan; 5) Genome Institute of Singapore, Singapore, Singapore.

Age-related macular degeneration (AMD) is a major cause of progressive, irreversible visual impairment among elderly population in developed countries. To date, more than 10 genome-wide association studies (GWASs) have been conducted on AMD. In 2013, a large international GWAS (≥ 70,000 subjects) have identified 7 novel AMD susceptible genes (The AMD Gene Consortium, Nature Genetics). However, since this study mainly consists of Caucasians (≥ 94%), this result needs to be replicated in other ethnicity. In the current study, we performed GWAS using large cohort of Japanese to look up the results. Case group consists of 1576 wet AMD patients recruited from all over Japan. These samples were genotyped using Illumina HumanOmni2.5M or OmniExpress. For the control group, we used 3248 healthy Japanese individuals recruited from the Nagahama Prospective Genome Cohort for the Comprehensive Human Bioscience (The Nagahama Study) dataset. After our standard quality control, 558,850 SNPs were included. GWAS was conducted with an adjustment for age, sex, and first 7 principal components. Inflation factor lambda of 1.07 suggested that population stratification was well adjusted. ARMS2 and CFH, that are the established disease susceptible genes of AMD, reached genome-wide significance ($P = 4.7\times10^{-22}$, 1.8×10^{-8}, respectively). We screened ±25kb region of the 7 single nucleotide polymorphisms (SNPs) newly reported by The AMD Gene Consortium. Though p-value of ≤0.05 was observed in ADAMTS9 (lowest P-value = 3.5×10^{-2}), COL8A1 (lowest P-value = 2.9×10^{-3}), and SLC16A8 (lowest P-value = 1.8×10^{-2}), only COL8A1 showed significant association after Bon-ferroni correction. Since no SNP was genotyped within ±25kb around rs3130783 which is located between IER3 and DDR1, we screened ±50kb region for this SNP. This analysis revealed significant association of this region (lowest P-value = 3.0×10^{-3}) before multiple comparison correction. However, it was no more significant after Bon-ferroni correction. In conclusion, 4 regions out of 7 regions newly reported by the AMD Gene Consortium might be associated with AMD in Japanese.

936W

Imputation-Based Genomic Coverage Assessments of Current Human Genotyping Arrays.
S.C. Nelson[1], K.F. Doheny[2], E.W. Pugh[2], J.M. Romm[2], H. Ling[2], S.R. Browning[1], B.S. Weir[1], C.C. Laurie[1]. 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Center for Inherited Disease Research, Johns Hopkins University School of Medicine, Baltimore, Maryland.

Microarray SNP genotyping, combined with imputation of untyped variants, has been widely adopted as an efficient means to interrogate variation across the human genome. The total proportion of genomic variation captured by an array, either by direct observation or imputation, is referred to as 'genomic coverage.' We have performed imputation-based genomic coverage assessments of eight current genotyping arrays that assay ~0.3 to ~5 million variants. Coverage was estimated using the 1000 Genomes Project phase 1 release, with 1,092 samples representing 14 populations in four continental ancestry groups. Samples were divided into ten batches, balancing across populations. In each batch, array variants were used to impute the remaining 1000 Genomes variants, with the rest of the samples serving as the imputation reference. The squared correlation between the observed and imputed allelic dosage (imputation r^2) was calculated at each imputed variant over all samples within an ancestry group. The percentage of variants with minor allele frequency (MAF)>0.05 and r^2≥0.8 is >75% for all arrays and ancestry groups except for African ancestry, and up to ~90% in all ancestries for the highest density arrays. The percentage of variants with 0.01<MAF<0.05 and r^2≥0.8 is substantially lower: <40% for low density arrays in all ancestries and 50-80% in high density arrays, depending on ancestry. We calculated genome-wide power to detect variant-trait association in a case-control design, as an average over a grid of MAF by r^2 bins weighted by the number of variants per bin, assuming an additive genetic model, prevalence of 5%, significance level of 5×10^{-8} and with variable sample size (N) and genotype relative risk (GRR). Genome-wide power for array genotyping with imputation was compared with a hypothetical array that would type all variants in 1000 Genomes. These results can be used to estimate cost efficiency of different array choices. For example, for variants with MAF>0.05 and GRR= 1.3, achieving 80% power requires N=8,700 (1000 Genomes), 9,700 (Omni2.5M) or 11,700 (HumanCore) in European ancestry and N=10,100 (1000 Genomes), 11,700 (Omni2.5M) or 19,600 (HumanCore) in African ancestry, where N is the total sample size (including equal numbers of cases and controls). These imputation-based genomic coverage and power analyses are intended as a practical guide to researchers weighing the costs and benefits of current array choices in different ancestry groups.

937T

Understanding of IL28B gene associated with treatment response for HCV patients.
N. Nishida[1,2], Y. Tanaka[3], M. Sugiyama[1], Y. Mawatari[1], M. Ishii[1], C. Haga[1], K. Tokunaga[2], M. Mizokami[1]. 1) Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Chiba, Japan; 2) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan; 3) Department of Virology & Liver unit, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

Genome-wide association studies (GWAS) identified single nucleotide polymorphisms (SNPs) near interleukin-28B (IL28B) gene to be associated with response to PEG-IFN-α/RBV therapy for HCV patients (Tanaka et al. Nat Genet 2009, Ge et al. Nature 2009). The SNPs (rs8099917, rs1297860) could predict treatment responses at approximately 80%, however, the remaining 20% of HCV patients would have unrevealed host genetic factors. A recent report showed that a frame-shift mutation (ss469415590), which located upstream of IFNL3 (IL28B), generated a novel gene, designated IFNL4 (Prokunina-Olsson et al. Nat Genet 2013).

To identify novel host genetic factors except for IL28B (rs8099917), we performed a GWAS with increased number of HCV individuals (118 with null virological response (NVR) and 140 with virologic response (VR)). Candidate genetic regions for replication analysis were selected from the result of GWAS by combining the information of gene-gene interactions and genetic pathways. We performed the replication analysis using a total of 895 HCV patients including (321 NVR and 574 VR). A GWAS using Japanese 118 NVR and 140 VR identified the strongest associations of rs8099917 with NVR ($P = 9.6\times10^{-18}$, OR=7.62), and also detected 16 novel candidate genetic regions with P values below 10-4. Moreover, there were 38 additional candidate genetic regions to be selected by gene-gene interactions and genetic pathways.

We found no significant association of candidate genes except for IL28B gene, in a total of 895 samples. The combined p-value for rs8099917 reached $P = 2.5\times10^{-42}$ (OR=6.00). We also determined the LD structure in a genetic region including IL28B in the Japanese population. Two known associated SNPs (rs1297860 and rs8099917) with HCV drug response had a strong LD with the frame-shift mutation (ss469415590).

Further studies following our report of these robust genetic associations to NVR may make it possible to develop a pre-treatment predictor of which individuals are likely to respond to PEG-IFN-α/RBV treatment. This would remove the need for the initial 12-24 weeks of treatment that is currently used as a basis for a clinical decision about whether treatment should be continued.

938F

Association of body mass-associated polymorphisms in *FTO* intron 1 with determination of Sasang constitutional types in Koreans. *A. Park, S. Cha, H. Yu, J. Kim.* Korea Institute of Oriental Medicine, Deajon, South Korea.

Sasang constitutional (SC) medicine, a part of traditional Korean medicine, classifies humans into 4 constitutional types~Tae-Yang, Tae-Eum (TE), So-Yang, and So-Eum (SE)~based on responses to herbal medicine and individual psycho-physiologic characteristics. Interestingly, the constitutional types are inheritable as revealed by twin and family studies. Recently, several genetic polymorphisms have been associated with SC types from 1,222 subjects via genome-wide association (GWA) analysis. However, the association signal appears to be weak due to the small size of the population used, and the associations are not replicated in the following study. Here, we performed a GWA analysis and replication analysis for SC types determined using an SC analytic tool on the basis of top tertile of the probability values for each type, by scaling up the population size into 5,478 subjects (3,798 subjects in the discovery stage and 1,680 subjects in the replication stage). We found that the minor alleles of polymorphisms in intron 1 of fat mass and obesity associated (*FTO*) gene were replicably associated with SE type in 2 populations (a peak polymorphism in combined: odds ratio (OR) = 0.729, p = 1.45E-07). For the other SC types, there were additional associations with the TE type (a peak polymorphism in combined: OR = 1.27, p = 1.45E-05) despite weak association signals in the discovery population, but not with the So-Yang type. Since the minor alleles of the polymorphisms in *FTO* intron 1 have been associated with increased body mass index (BMI), we performed association analyses by controlling BMI. The association signals in SE type remained significant (a peak polymorphism in combined: OR = 0.688, p = 1.28E-04) but not in the TE type. These results corresponded well with bodily characteristics of SE and TE types. That is, the subjects with SE type are known to be the slimmest among SC types, whereas the subjects with TE type tend to gain more body mass than other SC types. In conclusion, the BMI-associated polymorphisms in *FTO* intron 1 were significantly associated with SE type, independent of BMI as per the GWA and replication analyses. Therefore, the minor alleles of the *FTO* polymorphisms appear to be involved in not only body mass increase but also SC type determination.

939W

A genome-wide association study to identify susceptibility loci for bronchiolitis. *A. Pasanen[1], M.K. Karjalainen[1], M. Ruotsalainen[2], E. Piippo-Savolainen[2], E. Goksör[3], G. Wennergren[3], M. Hallman[1], M. Rämet[4], M. Korppi[5].* 1) Department of Pediatrics, Institute of Clinical Medicine, Clinical Research Center, University of Oulu, Oulu, Finland; 2) Kuopio University Hospital, Pediatrics, University of Eastern Finland, Kuopio, Finland; 3) Queen Silvia Children's Hospital, Pediatrics, University of Gothenburg, Gothenburg, Sweden; 4) Institute of Biomedical Technology and BioMediTech, University of Tampere and Department of Pediatrics, Tampere University Hospital, Tampere, Finland; 5) Pediatric Research Center, Tampere University and Tampere University Hospital, Tampere, Finland.

Background: Bronchiolitis is a severe respiratory disease primarily affecting infants less than one year of age. The most common causative agent of bronchiolitis is respiratory syncytial virus. The severity of bronchiolitis varies among healthy infants; therefore genetic factors likely affect the disease outcome. So far, a number of polymorphisms have been associated with severe disease in candidate gene studies conducted mainly on immunity genes. To our knowledge, genome-wide analyses have not been previously published. **Objective:** To identify genetic susceptibility loci predisposing to severe bronchiolitis. **Methods:** Individuals from Finnish and Swedish populations were genotyped with Illumina HumanOmniExpress BeadChip containing approximately 700,000 SNPs. Finnish population consisting of two birth cohorts with a total of 64 cases and 64 controls was collected in Kuopio during 1979-1982 and 1990-1993. Swedish population with 31 cases and 31 controls was collected in Gothenburg during 1981-1982. Statistical analyses were performed using plink v.1.07 and imputation was carried out to boost resolution. Cases and controls were matched based on birth date (± two months). Only individuals of European origin were included in the analyses. The most significant SNPs will be further analyzed in Finnish populations from Tampere (cases n = 160, controls n = 320), Oulu (n = 200) and Seinäjoki (n = 60) regions. **Results:** Several signals suggestive for association with bronchiolitis (p-value $\le 10^{-5}$) were found. Statistically significant signals were not expected with the present small sample sizes. In the next step we will genotype the best associating SNPs in larger, independent populations to identify significant associations. One of the most promising genes showing association was *KLRK1* (killer cell lectin-like receptor subfamily K, member 1), which is centrally involved in natural killer (NK) cell function. **Conclusions:** We found suggestive association signals for several SNPs. Because one of the signals was within the *KLRK1* gene, we hypothesize that an altered NK cell function could affect the susceptibility to bronchiolitis. The most promising associating SNPs will be further analyzed in independent Finnish populations to identify significant association signals. Later, functional studies of relevant candidate genes will be performed.

940T

The contribution of low frequency coding variants to normal variation in adult height and BMI. *D. Pasko[1], K. Stirrups[2], N. Masca[3], L. Southam[4], T.V. Varga[5], H. Zhang[6], R. Mägi[4, 7], D. Thompson[8], W. Zhang[9], A. Mahajan[4], R.A. Scott[10] on behalf of UK ExomeChip Consortium.* 1) Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom; 2) Wellcome Trust Sanger Institute, Hixton, Cambridge, United Kingdom; 3) Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, United Kingdom; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 5) Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centerm Malmo, Sweden; 6) Division of Cardiology, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, United States of America; 7) Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, United Kingdom; 9) School of Public Health, Imperial College London, London, United Kingdom; 10) MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom.

The role of rare (MAF<1%) and low frequency (1-5% MAF) protein coding variants in common traits is unknown. We used the HumanExome Beadchip to test the role of low frequency and rare coding variants in 47,224 and 32,733 European-ancestry individuals from up to 20 studies for adult height and BMI respectively. We aimed to identify low frequency large effect (OR>1.5) exonic variants, and test the distribution of the effect sizes that we could exclude and how these compared to the known common signals. We combined summary statistics (single-variant) from up to 190,918 (height) and 187,603 (BMI) autosomal variants across studies by meta-analysis. Fifteen percent of variants were low frequency and 78% were rare. We performed analyses in a discovery sample that left out studies representing 15-20% of the data and looked for directional consistency of the most strongly associated 100 variants in the left out studies across 3 categories of variants: rare and low frequency variants i) within 1Mb of known GWAS signals, ii) further than 1Mb from known GWAS signals and iii) within 1Mb of known monogenic genes. Finally we compared the effect sizes of the 20 most strongly associated variants, regardless of their significance, to the leading known common signals. No rare and low frequency coding variants achieved study wide significance (2.8×10-7) for height or BMI. In contrast, for height, for each of the 100 most strongly associated low frequency or rare variants (all MAF<5%) in the discovery samples that fell within 1Mb of a GWAS signal, further than 1Mb from a GWAS signal, and within 1Mb of a monogenic genes, 63%, 65% and 58% respectively were directionally consistent in the left out studies. For BMI the equivalent figures were 60%, 52% and 53% (p=0.02 for 60% consistency, p=0.003 for 65% consistency). In the left out studies we observed that, of the 100 leading low frequency variants from the discovery data, 28 and 43 were consistent with effect sizes greater than the strongest known common signals for height (HMGA1 0.09 SDs) and BMI (FTO 0.08 SDs) respectively. Our results are consistent with the presence of low frequency coding variants contributing to human anthropometric phenotypes but demonstrate that larger sample sizes will be needed to identify specific variants that contribute effects of similar or greater magnitude to the strongest known common variants such as those in FTO with BMI.

941F

On individual genome-wide association studies and their meta-analysis. *Y. Pei[1,2], L. Zhang[1,2], C. Papasian[3], Y. Wang[2], H. Deng[1,2].* 1) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China; 2) Center for Bioinformatics and Genomics, Department of Biostatistics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA; 3) Department of Basic Medical Science, University of Missouri-Kansas City, Kansas City, MO 64108, USA.

Individual genome-wide association (GWA) studies and their meta-analyses represent two approaches for identifying genetic loci associated with complex diseases/traits. Inconsistent findings and non-replication among/between individual GWA studies and meta-analyses are commonly observed, hence posing the critical question as to how to interpret their respective results properly. In this study, aiming to investigate and compare the statistical properties of the two approaches, we performed simulation studies. Our results show that: 1) As expected, meta-analysis is more powerful than individual GWA studies under the ideal setting of population homogeneity among the individual GWA studies because of enlarged sample sizes; 2) Under the realistic setting of heterogeneity between individual GWA studies included in meta-analyses, detection of heterogeneity is usually difficult and meta-analysis (even with the random-effects model) may introduce elevated false positive and/or negative rates; 3) Even with relatively small sample sizes, well-designed individual GWA studies still have the capacity to identify novel loci for complex traits; 4) Replicability between meta-analysis and independent individual GWA studies or even between different independent meta-analyses is limited, thus the inconsistent findings (even for true positive results) among studies should not be surprising.

942W

Genome-wide association study and admixture mapping reveals new loci associated with total IgE levels in Latinos. *M. Pino-Yanes[1], C.R. Gignoux[2], J.M. Galanter[1,2], A.M. Levin[3], R. Mathias[4], C. Eng[1], E.A. Nguyen[1], L.A. Roth[1], S. Huntsman[1], K. Sandoval[5], A. Moreno[5], C.A. Winkler[6], L.N. Borrell[7], B.A. Raby[8], S.T. Weiss[8], D.L. Nicolae[9], C. Ober[9], D.A. Meyers[10], E.R. Bleecker[10], F.D. Martinez[11,12], S. Sen[13], R. Kumar[14], C. Bustamante[5], K.C. Barnes[4], L.K. Williams[15,16], D.G. Torgerson[1], E.G. Burchard[1,2] On behalf of the GALA II investigators.* 1) Department of Medicine, University of California, San Francisco, CA, USA; 2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA; 3) Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan; 4) Division of Allergy & Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA; 5) Department of Genetics, Stanford University, Stanford, CA, USA; 6) Basic Research Laboratory, SAIC-Frederick, Inc., Center for Cancer Research, National Cancer Institute, Frederick, MD, USA; 7) Department of Health Sciences, Graduate Program in Public Health, City University of New York, Bronx, NY, USA; 8) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 9) Department of Human Genetics, University of Chicago, Chicago, Illinois, USA; 10) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA; 11) Arizona Respiratory Center, University of Arizona, Tucson, Arizona, USA; 12) Center for Health Policy and Health Services BIO5 Institute, University of Arizona, Tucson, Arizona, USA; 13) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA; 14) Children's Memorial Hospital and the Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; 15) Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA; 16) Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan.

Immunoglobulin E (IgE) is key mediator of allergic inflammation and is frequently elevated in allergic disorders, such as asthma. Although IgE levels are influenced by the environment, there is an important genetic contribution with estimates of heritability as high as 80%. IgE varies by race-ethnicity in the United States, with higher levels reported in both Latinos and African Americans. While genome-wide association studies (GWAS) have identified ten genes associated with total serum IgE levels, most have been performed in populations of European descent.

We performed a GWAS and admixture mapping study of total IgE levels in 3,334 Latinos participating in the Genes-environments & Admixture in Latino Americans (GALA II) Study. Our GWAS identified a genome-wide significant association of a common intronic indel in *ZNF365* with total IgE (rs200076616, $p=2.3\times10^{-8}$), which replicated in an independent sample of Latinos (n=216, $p=0.03$). We identified four admixture mapping peaks where significant correlations with IgE were observed with African, European, or Native American ancestry at 6p21-22, 13p22-31, 14q23.2, and 22q13.1; all but 22q13.1 were replicated in an independent sample of Latinos, and two of the peaks replicated in African Americans (6p21-22 and 14q23.2) ($p\leq0.05$). The most significant peak in GALA II was 6p21-22, where Native American ancestry was associated with lower levels of IgE ($\beta=-0.10$, $p=4.95\times10^{-8}$), and African ancestry was associated with higher levels of IgE ($\beta=0.07$, $p=1.6\times10^{-4}$). Fine mapping using imputation and allelic association testing within 6p21-22 identified a genome-wide significant SNP approximately 12 kb 5' of *HLA-DQA1* (rs1846190, $p=3.5\times10^{-8}$) that replicated in Latinos ($p=0.023$). Locus-wide significant SNPs were also identified at 13p22.3 near *MYCBP2* ($p<1\times10^{-5}$). We then tested the generality of these findings across African Americans (n=3,187) and European Americans (n=1,564). Overall our results support the importance of studying diverse, multiethnic populations and the use of admixture mapping to uncover new loci associated with IgE. Future studies are required to identify the causal variation behind these associations.

943T

Sequencing-based GWAS contributes to the refinement of association signals for inflammatory markers in Sardinians. *E. Porcu[1,2,3], C. Sidore[3], M. Steri[3], A. Mulas[3], M. Zoledziewska[3], F. Busonero[3], A. Maschio[1,3], F. Danjou[3], G. Pistis[1,2,3], M.G. Piras[3], M. Dei[3], S. Lai[3], H.M. Kang[1], C. Brennan[4], R. Berutti[2,5], M.F. Urru[5], M. Oppo[5], A. Angius[3,5], D. Schlessinger[6], F. Cucca[2,3], G. Abecasis[1], S. Naitza[3], S. Sanna[3].* 1) Center for Statistical Genetics, Biostatistics Department, University of Michigan, Ann Arbor, MI; 2) Università degli Studi di Sassari, Dip Scienze Biomediche, Sassari, 07100, Italy; 3) Istituto di Ricerca Genetica e Biomedica-CNR, Monserrato (CA), 09044, Italy; 4) DNA Sequencing Core, University of Michigan, Ann Arbor, MI, USA; 5) Center for Advanced Studies, Research and Development in Sardinia - CRS4, Pula, Italy; 6) Laboratory of Genetics, NIA, Baltimore, MD, 21224.

Inflammation is a response of a tissue to injury often caused by invading pathogens. It has also been implicated in the pathogenesis of numerous chronic diseases but we are still far from a full understanding of this complex phenomenon. Identifying the genes that influence levels of pro-inflammatory molecules can help to elucidate the factors and mechanisms underlying inflammation and their consequences on health. Genome-wide association studies have revealed several associated variants, but those are insufficient to explain the estimated heritability. Here, we measured the levels of the key inflammatory biomarkers Adiponectin, the high-sensitivity C-reactive protein (hsCRP), erythrocyte sedimentation rate (ESR), monocyte chemotactic protein-1 (MCP-1) and interlukin-6 (IL-6) in ~6,000 individuals enrolled in the SardiNIA project. Samples were genotyped with four different Illumina genotyping arrays: OmniExpress, ExomeChip, Cardio-MetaboChip and ImmunoChip and in order to increase the genomic resolution, we imputed ~17 million variants using a reference panel derived from whole-genome low-pass sequencing of 2,120 Sardinians. For each trait, we carried out a sequencing-based genome-wide association scan that allowed us to fine map known signals and detect novel loci. The impact of using sequencing data was clear for all traits, with the exception of IL-6, for which no major differences were seen compared to our previous HapMap GWAS. Interestingly, the same genomic resolution was not reached when using the 1000 Genomes reference panel for imputation. For example, with the 1000 Genomes imputation we were unable to detect the association at the HBB gene for erythrocyte sedimentation rate, with the stop codon variant Q40X. This mutation is common in Sardinia (MAF=5%) but very rare elsewhere. It was present in only one haplotype among 1000 Genomes samples, and thus poorly imputed as rare variant ($R^2=0.31$, MAF=0.008%). Association was missed in the 1000 Genomes imputed data set for three other variants. Consequently, when considering the top variants detected with imputation of the Sardinian haplotypes, the explained heritability was higher for four of five traits than when estimated with the top variants observed after 1000 Genomes imputation. Our study proves the importance of population specific reference panels and their value over the base 1000 Genomes reference set.

944F

Genetic variants underlying endometriosis risk: Meta-analysis on eight genome-wide association and replication studies. *N. Rahmioglu[1], D. Nyholt[2], S.A. Missmer[3], G.W. Montgomery[2], K.T. Zondervan[1].* 1) Wellcome Trust Centre for HG, University of Oxford, Oxford, United Kingdom; 2) Queensland Institute of Medical Research, Brisbane, QLD 4029, Australia; 3) Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA.

Endometriosis is a complex women's disease with heritability estimated at 52%. A meta-analysis of the published Japanese, Australian and UK endometriosis GWASs confirmed three published loci including, CDKNA2-BAS (rs1537377; P=2.4×10^{-9}), intergenic chromosome 7 (rs12700667; P=3.6×10^{-9}), WNT4 (rs7521902; P=3.2×10^{-11}) and provided evidence for a further four (Nyholt 2012). However, since the publication of the meta-analysis, a fourth GWAS in women of European ancestry from the US was published (Albertsen 2013), involving 2019 cases and 14471 controls. In addition, two replication studies in women of European ancestry were published involving: (1) 1129 cases and 831 controls from Belgium (Sundqvist 2013), (2) 305 cases and 2710 controls from Italy (Pagliardini 2013). Some of the replication evidence for locus-specific associations in these recent papers was non-significant and has generated hypotheses of heterogeneity in the genetic loci underlying endometriosis in different clinical studies even within populations of similar-European-ancestry. Here, our aim is to investigate the heterogeneity and consistency of results across all published eight GWAS and replication studies from Australia, Belgium, Italy, Japan, the UK, and the US through fixed and random effect meta-analyses. The results show robust evidence for nine genetic loci associated with endometriosis, including rs12700667 (P=9.8×10^{-11})-of which association with endometriosis was contested in the two published replication studies-and rs7521902 in WNT4 (P=1.4×10^{-12}), with consistent directions of effect across studies and populations and showing no evidence for heterogeneity. Furthermore, six of nine loci are associated with stronger effect sizes among more severe endometriosis cases(rAFS 3/4), implying that they are likely to be implicated only in severe disease. Whilst three variants are in intergenic regions and have unknown functionality, the remainder are in/near genes with known functions with biological relevance to endometriosis. Furthermore, although the results are reassuring in terms of their consistency, they also show that the phenotypic definitions used in the GWAS to date are crude. This emphasises the importance for future studies to include more detailed phenotype data, an objective that forms the basis of the global EPHect initiative aimed at harmonisation of clinical phenotype and biological sample collection in endometriosis research (http://endometriosisfoundation.org/ephect/).

945W

Searching genes for body shape: a genome-wide meta-analysis in >170,000 individuals of the GIANT consortium. *J.S. Ried[1], J.M. Jeff[2], J. van Dongen[3,4], J.E. Huffman[5], J. Bragg-Gresham[6], A.Y. Chu[7], M. Müller-Nurasyid[1,8,9], R.J.F. Loos[2,10,11] on behalf of the GIANT consortium.* 1) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 2) The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 3) Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; 4) EMGO+ Institute for Health and Care Research, VU Medical Centre, Amsterdam, The Netherlands; 5) MRC Human Genetics, MRC IGMM, University of Edinburgh, Edinburgh, Scotland, United Kingdom, EH4 2XU; 6) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA; 7) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA; 8) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 9) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany; 10) The Mindich Child Health and Development Institute,The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 11) MRC Epidemiology Unit, University of Cambridge, Cambridge, UK.

Analyses within the GIANT and other consortia have revealed hundreds of anthropometric trait loci. Although these studies have been successful in identifying loci for one trait at a time, the question remains open if there are genetic variants that affect 'body shape' as a more comprehensive composite phenotype that is represented by multiple anthropometric traits. Therefore, we calculated four average principal components (avPCs) based on six anthropometric traits (body mass index (BMI), height, hip circumference, waist circumference, and waist hip ratio (WHR)) in a standardized manner across studies. The association between ~2.5 million SNPs and each of the four avPCs was meta-analyzed over 43 studies (N>132,000). SNPs in 606 independent loci reached promising associations ($P<5\times10^{-6}$); 93 for avPC1, 328 for avPC2, 133 for avPC3, 52 for avPC4. These loci were followed up in a second stage of 22 studies (N>39,000), revealing 199 loci that reached genome-wide significance. We confirmed many loci that were found previously in GIANT analyses on height, WHR and BMI. The number and type of confirmed loci represented the main body size phenotypes captured by each avPC. Most confirmed loci for the avPC1 were found with BMI and analysis of avPC2 identified many loci known from height. The analyses of avPC3 and avPC4 mainly confirmed loci known from WHR or height analyses. Nevertheless, eight loci had not been previously identified, 2 for avPC1 (nearest gene *FANCL* and *CD47*), 1 for avPC2 (nearest gene *IL17RC*), 2 for avPC3 (*GANAB*, nearest gene *RPS6KA5*) and 3 for avPC4 (*ARL15, PLCE1, CCDC91*). We followed these eight loci in previously published meta-analyses on various traits to unravel potential pleiotropic effects. For example the locus *CD47* was also associated with HDL levels (GLGC, p=3×10^{-3}) as well as fasting insulin (MAGIC, p=1×10^{-3}), HOMA-IR (MAGIC, p=3×10^{-3}) and weakly with HOMA-B (MAGIC, p=0.02) and diabetes (DIAGRAM, p=0.01). Using this alternative approach, we successfully confirmed previously established associations with anthropometric traits while we also identify novel loci that associate with our comprehensive body shape dimensions. This suggested that the avPCs represent body shape phenotypes that are not solely represented by height, WHR and BMI. This work highlights the advantage of using multiple traits to define complex phenotypes for discovery, which are not otherwise captured by single-trait analyses and may shed light onto new pathways.

946T

Metabolomics in Genome-Wide Association Studies: Revealing Gene-Metabolite-Disease Links. *R. Rueedi[1,2], M. Ledda[3], S. Bergmann[1,2], U.K. Genick[3], Z. Kutalik[1,2].* 1) CBG/DGM, Université de Lausanne, Lausanne, VD, Switzerland; 2) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 3) Department of Food-Consumer Interaction, Nestlé Research Center, Lausanne, Switzerland.

While still in its infancy, the use of metabolic profiles in genome-wide association studies has improved the detection of, and provided biological context to, the sometimes poorly understood effects of genetic variants on clinical phenotypes. We ran a metabolome- and genome-wide association study on 1H-NMR urine profiles from 835 individuals of the Cohorte Lausannoise. We used an untargeted approach, with the NMR signal intensities as the study's phenotypes, thereby carrying all available metabolomic information into the study. We then identified the metabolites underlying significant association using, in addition to manual identification, a newly developed automated identification method. The two novel gene-metabolite associations resulting from this study are of particular interest because both involve metabolites which we also found to associate with clinical phenotypes. These associations thus provide specific, statistically-determined, evidence of the propagation of genetic effect from gene to disease via the metabolome.

947F

A Genome Wide Association Study To Determine the Susceptibility to *Staphylococcus aureus* Infection. *S.K. Shukla, D. Vasco, T. Carter, M.H. Brilliant, S.J. Schrodi.* Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI., USA.

Staphylococcus aureus is a significant opportunistic pathogen capable of causing a wide range of diseases from boil and carbuncle to bacteremia, endocarditis, and osteomyelitis. However, little is known about the role of host genetics to determine the susceptibility to S. aureus infection. The goal of this study was to conduct the first genome-wide association study (GWAS) to identify specific genetic variants that underlie susceptibility to infection from this pathogen. We conducted a case/control GWAS using samples from the Personalized Medicine Research Project (PMRP) at Marshfield Clinic. All 3616 subjects (408 cases/3208 controls) were genotyped at 522,204 SNPs on the 660W-QUAD Illumina platform. All 408 cases were defined as individuals who have at least one laboratory confirmed evidence of S. aureus infection. The remaining individuals who did not have evidence of the infection in their electronic medical record were considered as controls. PLINK was used for the association statistics. Gene-based and pathway analyses were done using VEGAS and GOrilla and DAVID, respectively. Extensive permutation tests were conducted to obtain appropriate P-values for gene-based and pathway-based results. Additionally, we compared results to those generated under the null by randomly sampling from the entire dataset to minimize the likelihood of false positive results. Males were significantly more likely to be cases than females (P=1.50E-06). The avg age of cases mildly exceeded that in controls (Case Avg=75.6 y/Control Avg=73.6 y; P=5.70E-04). Subjects with higher BMI and/or type 2 diabetes were enriched in cases (BMI: Case Avg=31.9/Control Avg=29.5; P=5.44E-09) (T2D: Case freq=22.0%/Control freq=13.0%; P=2.50E-06). Q-Q plots showed no evidence of confounding by population stratification. No individual SNP exceeded genome-wide significance but eight SNPs exceeded the P<10-6 warranting further investigation. The top results in the gene-based analysis highlighted several biologically interesting candidates involved in inflammation including interleukin subunits and an integrin binding molecule gene. Pathway analysis identified potentially critical molecular networks including the protocadherin-gamma pathway and MAPK signalling pathway. We have identified several SNPs that may influence susceptibility to S. aureus infection. Pathway analysis suggests the potential role of adhesion and inflammation pathways.

948W

Uncovering Genetic Modifiers of Sickle Cell Anemia in a Sub-Saharan African Population - A Genome-Wide Association Study of Sickle-Cell Severity. *T. Singh[1], S. Nkya[2], S.E. Cox[2,3], J.C. Barrett[1], J. Makani[2,4].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK; 2) Muhimbili Wellcome Programme, Muhimbili University of Health and Allied Sciences, Dar-es-Salaam, Tanzania; 3) MRC International Nutrition Group, London School of Hygiene & Tropical Medicine, London, UK; 4) Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK.

BACKGROUND: Sickle cell anemia (SCA) is the most common inherited hemoglobinopathy caused by the homozygous inheritance of sickle haemoglobin (HbS) and mostly affects populations of African descent. While the causal variant of the disorder has long been known, SCA results in a heterogeneous set of clinical outcomes affecting all organ systems, and usually includes varying degrees of chronic anemia and episodic acute pain. In our association study, we hope to find common variants that explain the heterogeneity of SCA severity, and in particular, to find further variants that contribute to the heritability of fetal haemoglobin (HbF) levels, which correlate with SCA outcome. This is the largest genome-wide association study analyzing SCA in a population based in Africa, with the widest range of hematological traits and parameters. METHODS: 1167 patients with a phenotype of HbSS or HbSF, based on quantification of hemoglobin fractions using HPLC (Variant I, Beta-thalassaemia short programme, BioRad, Hercules, CA, USA) from the Muhimbili Sickle Cohort in Dar-es-Salaam, Tanzania were genotyped. Following QC and imputation, associations tests were performed on fetal and haemoglobin (HbF) using ~2.4 million SNPs typed on the Illumina Omnichip platform. RESULTS: Despite including only Tanzanian SCA patients in our cohort, substantial population stratification was still observed in our initial association tests. A linear mixed model framework successfully controlled for population stratification, genetic and cryptic relatedness. Our study successfully replicated known associations for HbF in BCL11A (rs1427407, p=4.9e-42) and the HBS1L-MYB intergenic region (rs9494145, p=4.3e-10), while a few novel loci that are genome-wide suggestive were also identified. We will conduct similar regression analyses on additional clinical phenotypes relevant to SCA outcomes with the same dataset, and are in the process of initiating a replication study to verify our results in an independent population with SCA.

949T

Identification of 3 Susceptibility Loci for Crohn's Disease in a Genome-Wide Association Study of a Korean Population. *K. Song[1], S. Yang[2], M. Hong[3], W. Zhao[1], J. Baek[1], I. Liu[3], Y. Jung[1].* 1) Biochem and Molecular Biol, Univ of Ulsan College of Medicin, Seoul, South Korea; 2) Internal Medicine, Univ Ulsan Col Medicine, Seoul, Korea; 3) Genome Institute of Singapore, Singapore.

We have performed a three-stage genome-wide association study in 2,311 Korean patients with Crohn's disease and 2,442 controls. Three new susceptibility loci were identified at genome-wide significance: SNP1 at 4p14 (OR= 1.43, Pcombined = 3.60 × 10-14), SNP2 at 10q25 (OR=1.42, Pcombined = 1.55 × 10-10), and SNP3 at 11q13 (OR=1.46, Pcombined = 7.15 × 10-9). Using data from the International IBD Genetics Consortium, disease associations of SNP1 (P= 0.00024) and SNP2 (P= 5.32 × 10-5) were replicated in Caucasians. We also replicated four previously reported loci of Caucasian population: TNFSF15, IL23R, the major histocompatibility complex (MHC) region, and the RNASET2-FGFR1OP-CCR6 region. Together, the novel and replicated loci accounted for 5.31 % of the total genetic variance for CD risk in Koreans. Our study provides new biological insight to CD and supports the complementary value of genetic studies in different populations.

950F

Genetic burden of common variants in progressive and bout onset multiple sclerosis. *M. Sorosina[1], P. Brambilla[1], F. Clarelli[1], N. Barizzone[2,3], S. Lupoli[4], C. Guaschino[1,5], G. Liberatore[1,5], A.M. Osiceanu[1], V. Martinelli[5], D. Cusi[4], M. Leone[2], G. Comi[5], S. D'Alfonso[2,3], F. Martinelli Boneschi[1,5], International Multiple Sclerosis Genetics Consortium, PROGEMUS, PROGRESSO.* 1) INSPE, San Raffaele Scientific Institute, Milan, MI, Italy; 2) Interdisciplinary Research Center of Autoimmune Disease IRCAD, University of Eastern Piedmont, Novara, Italy; 3) Department of Health Sciences, University of Eastern Piedmont, Novara, Italy; 4) Department of Health Sciences, University of Milan and Genomics & Bioinformatics Unit, c/o Fondazione Filarete, Milan, Italy; 5) Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy.

Multiple sclerosis (MS) is a neurological autoimmune disease in which genetic and environmental risk factors influence the risk to develop the disease. It could be divided in two main clinical courses based on the characteristic of the onset: progressive onset (PrMS) and bout onset (BOMS) MS. The genetic component of MS has been extensively studied and several loci was found to be associated with the susceptibility to the disease, while the contribution of common genetic variants in the risk of developing the different clinical courses is still unclear. We estimated the proportion of variance in disease liability explained by 296,391 autosomal SNPs in a cohort of Italian PrMS and BOMS patients using the genome-wide complex trait analysis (GCTA) tool and we calculated a weighted log-additive genetic risk score (wGRS) based on the known MS genetic risk variants to evaluate and compare the role of common variants in the disease courses of MS. Our results identified that common SNPs explain more phenotypic variance in BOMS (36.5%±10.1%) than in PrMS (20.8%±6.0%), and this difference further increased (14.7%±6.9%) when considering pure primary progressive (PPMS) cases. Similarly, genetic burden measured using wGRS and explained variance by MS-associated SNPs were higher in BOMS than PPMS in males (wGRS: 6.63 vs 6.51, p=0.04; explained variance: 4.8%±1.5% vs 1.7%±0.6%; p=0.05). Our results suggest that the liability of disease is better explained by common genetic variants in BOMS than PrMS cases. Moreover, the absence of superimposed relapses and male gender further increase the difference between clinical courses.

951W

Dissecting genomic architecture through advanced SNP-based heritability analysis. *D. Speed[1], V. Plagnol[1], M. Johnson[2], D. Balding[1].* 1) UCL Genetics Institute, University College London, London, United Kingdom; 2) Division of Brain Sciences, Imperial College London, United Kington.

SNP-based heritability analysis has had a large impact on the study of complex traits in recent years, but we show that extensions of the method can be used to answer many more questions than have been appreciated to date. We develop several such extensions and apply them to make detailed inferences about the genetic architecture of complex traits. We develop a likelihood ratio test to compare the heritability of different sets of SNPs relative to their sizes (measured in terms of genetic variance). Using this approach, we find that for rheumatoid arthritis and type 1 diabetes, the eQTLs have 50 to 100 times higher intensity of heritability than non-eQTL SNPs, but for other traits eQTLs carry little more heritability than expected by chance. We further find that exonic SNPs explain on average 4.5 times the heritability of non-genic SNPs, and that regions flanking exons have significantly inflated contribution up to 30Kb (after allowing for LD). We compare the heritability of SNPs that are, and are not, associated with a second trait, and use this approach to replicate the known concordance between Schizophrenia and bipolar disorder, and to obtain novel evidence of Schizophrenia and type 1 diabetes having a shared genetic architecture. By likelihood-based comparison of different heritability models, we find that although there is a tendency for variants of lower frequency to have greater effect size, this relationship is weaker than commonly assumed. Finally, we find that SNP-based heritability analysis provides a robust method for gene-based analysis, one which is better able to detect causal genes than single-SNP tests of association and that lends itself to subsequent testing of biological pathways. All analyses are facilitated by our software LDAK.

952T

GWAS Identifies Classical HLA Alleles Associated with Susceptibility to Infectious Diseases. *C. Tian, J. Mountain, N. Eriksson, J. Tung, A. Kiefer, D. Hinds.* Research, 23andme, Mountain View, CA.

The human leukocyte antigen (HLA) system is a highly polymorphic region on chromosome 6 that encodes several hundred genes critical for the immune system. HLA has long been known to play an important role in susceptibility and resistance to many infectious diseases and responsiveness to pathogens or vaccines. Many studies have shown significant associations between HLA loci and major infectious diseases, such as HIV/AIDS, hepatitis and malaria etc. However most studies have tested only candidate loci in small samples, and the associations between HLA alleles and many other infectious diseases are not well studied. We conducted genome wide association studies with over 61,000 individuals who were genotyped on the 23andMe platform and asked to report on multiple infectious diseases. We identified single nucleotide polymorphisms in the HLA region that are associated with cold scores, plantar warts, mumps, positive TB test, scarlet_ fever and shingles at a genome-wide-significant level. We did not detect associations between HLA loci and bladder infection, urinary tract infection, measles, intestinal parasite, myringotomy, mononucleosis or rheumatic fever. To explore the relationship between individual HLA alleles and susceptibility to infectious diseases we used HIBAG, a statistical method combining a large database of individuals with known HLA alleles and SNP variation within the MHC region, to impute HLA alleles at key class I and class II loci over all 23andMe customers of European ancestry. We found that HLA-B*18:01 is the HLA allele most significantly (p=5.5e-07) associated, in a protective manner, with cold sores. Individuals with the HLA-DQA1*03:01 allele were less likely to have plantar warts (p=1.7e-08). HLA-A*02:05 is the HLA allele most significantly associated with mumps (p=1.6e-12). Individuals with the HLA-DRB1*01:03 allele are more susceptible to tuberculosis (p= 4.5e-12). Individuals with the HLA-DQB1*03:01 allele are less likely to have scarlet_fever (p=9.0e-09). HLA-B*44:02 allele is an important factor in the resistance to shingles (p=3.4e-11). Although additional studies will be required to separate and validate the association signals in this complex region, combining accurate HLA typing with our knowledge of the function of specific HLA molecules in immune response and with GWAS can be a powerful tool for dissecting infectious disease etiology.

953F

Improving Genome-Wide Association Studies via Markers Decorrelation. *O. Weissbrod[1], D. Geiger[1], N. Zaitlen[2], D. Heckerman[3].* 1) Computer Science Department, Technion - Israel Institute of Technology, Haifa, Israel; 2) Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco; 3) eScience Group, Microsoft Research, Los Angeles.

Contemporary state of the art genome-wide association studies typically use linear mixed models (LMMs), due to their robustness to false positive results in the presence of population or family structure. It has recently been widely recognized that case-control studies which use covariates are prone to power loss due to ascertainment bias, where the frequency of cases in the study is larger than in the population. Such bias appears in most case-control studies. As we show in this work, LMMs are especially prone to power loss in such cases, because they implicitly treat every single nucleotide polymorphism (SNP) as a covariate. The result is that many risk variants elude detection. We devised a statistical test that bears many similarities to an LMM regarding its ability to properly control the false positive rate, but does not suffer power loss due to ascertainment, thus solving both fundamental problems of association studies. The principal idea behind our method is that true risk variants can be teased apart from false ones via a simple linear (Mahalanobis) transformation for variables decorrelation. After applying the transformation, SNPs that were correlated with the disease due to confounding are no longer correlated. A simple statistical test can then be carried out on the transformed data. We demonstrate that our method outperforms LMMs in ascertained studies of stratified samples via analyses of synthetic and real data. Simulations suggest that our method can have over 40% power gain over LMMs in realistic scenarios, and over 90% power gain under extreme ascertainment, corresponding to diseases with 0.1% prevalence. We used our method for analyses of Crohn's disease, type I diabetes, and Ulcerative colitis from the Wellcome Trust Case Control Consortium, and a highly stratified study of smoking habits from the Genome Analysis Workshop 14 data set. In all cases, our method managed to identify as many or more risk variants than an LMM, with a properly controlled false positive rate. We also verified that our method consistently assigned a statistically significant lower p-value to known risk variants. These results suggest that existing GWAS datasets may be sufficiently powered to reveal dozens of yet undiscovered disease loci.

954W

A group sparse additive model for genome-wide association studies of dynamic complex traits. *J. Yin[1], M. Marchetti-Bowick[2], J. Howrylak[3], E. Xing[1, 2].* 1) Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 2) Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA; 3) Division of Pulmonary and Critical Care Medicine, Penn State, Milton S. Hershey Medical Center, Hershey, PA.

Despite the widespread popularity of using genome-wide association studies (GWAS) to perform genetic mapping of complex traits, most existing GWAS methodologies are still limited to the use of static phenotypes measured at a single time point. In this work, we propose a new method for performing association mapping that considers dynamic phenotypes measured at a sequence of time points. Our approach relies on Group Sparse Additive Models (GroupSpAM) for nonparametric regression. This new technique detects a sparse set of genomic loci that are associated with trait dynamics, while simultaneously learning an explicit representation of the dynamic effects of genetic variants at each such locus. We perform a proof-of-concept analysis for detecting single nucleotide polymorphisms (SNPs) associated with the FEV1 score, a sensitive measure of airway obstruction used to assess asthma severity. We evaluate our method using a set of 1500 SNPs from the TGFB1 gene, which has an established role in asthma pathogenesis and thus serves as a positive control.

955T

Characterizing and redefining clinical subtypes of inflammatory bowel disease (IBD) using genotypes and phenotypes from 47,000 patients.
G. Boucher[1,2] on behalf of the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). 1) Research Center, Montreal Heart Institute, Montreal, Quebec, Canada; 2) Department of Medicine, Université de Montréal, Montreal, Quebec, Canada.

To date, large-scale genetic studies in IBD have concentrated on the broad clinical diagnoses of Crohn's disease (CD) and ulcerative colitis (UC). However, it is clear that these conditions are clinically heterogenous and encompass a wide range of subtypes, each with its own pattern of disease behaviour, location and outcome. Whether these subtypes of IBD are distinct diseases with different aetiologies and/or parts of a continuum remains open to debate. Recent progress in characterizing the genetic architecture of IBD, including the identification of 163 IBD-associated loci, affords a unique opportunity to address important clinical questions through the application of genetic risk algorithms. The IIBDGC (a multi-site international consortium) cohort is the largest IBD bioresource generated to date comprising data from over 47,000 patients. Clinical and demographic information on over 26,700 CD and 21,000 UC cases including gender, age at diagnosis, disease location and behaviour, duration of follow-up, surgical history and smoking status were collected according to agreed criteria. We conducted genotype-phenotype analyses across >150k variants (Immunochip), using both established and novel statistical methods for analyses of multinomial phenotypes. We confirmed genome-wide significant associations between the NOD2 and MHC loci and multiple CD phenotypes. We also confirmed an association of the MHC region with extensive UC. Moreover, our analyses implicated multiple additional variants in clinically relevant sub-types. Using the 163 validated IBD loci we showed a significant association between both CD and UC risk scores and CD disease location and behaviour. This association remained highly significant (p-value$<10^{-20}$) after removing the HLA and NOD2 signals, confirming a role for many IBD risk loci in CD sub-phenotype. Furthermore, we used risk scores to show that colonic CD (a subtype phenotypically similar to UC) is genetically intermediate between ileal CD and UC. In conclusion, our data demonstrate significant genetic influence on IBD clinical heterogeneity. They support the hypothesis that UC, colonic CD and ileal CD could be considered as part of a continuous spectrum. These findings will enhance our understanding of the underlying pathogenesis of the subtypes within IBD, and will likely facilitate a molecular classification of IBD subtypes - a major step on the pathway to personalized medicine.

956F

Genome wide association in 5 isolated populations give new insight on the genetic bases of food preferences.
N. Pirastu[1], M. Kooyman[2], A. Robino[1], C. van Duijn[2,3], D. Toniolo[4,5], P. Gasparini[1]. 1) Medical Genetics, Institute for Maternal and Child Health - IRCCS 'Burlo Garofolo' - Trieste, Univ, Trieste, Trieste, Italy; 2) Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, the Netherlands; 3) Centre for Medical Systems Biology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; 4) DGCB, San Raffaele Research Institute, Milano; 5) 5IGM-CNR, Pavia.

Food preferences are the first factor driving food choice and thus nutrition. It involves numerous different senses such as taste and olfaction plus numerous other factors such as personal experiences and hedonistic aspects. It is clear that all of these have a genetic base although up to now very limited studies have been conducted. In this work we aimed at identifying the genes that underlie to food preferences using 5 different isolated populations. We have collected during whole population screening ~5000 samples coming 3 Italian isolated population from the INGI network, 1 Western European isolated population from the Netherlands (Erasmus Rucphen Family (ERF) study)) plus one coming from isolated communities scattered along 8 countries along the silk road (SR). For each participant we collected numerous information regarding life style, health plus, genome wide genotypes using high density SNP arrays and a food preferences questionnaire which included more than 60 common foods such as liver, orange juice and dark chocolate. Since many of the preferences were specific to each population we decided to restrict our analysis to 22 common foods. We run genome wide association on all 5 cohorts using 1000G imputed SNPs. In order to account for relatedness between samples analyses were conducted using GenABEL and MixABEL. Given the reduced sample size analyses were conducted also on non-additive models (dominant, recessive and overdominant) in order to maximize power to detect associations. Metanalysis was then conducted on the European cohorts and SR was used for replication. Overall we found 19 genome wide significant results for liking of Artichokes (4), Eggplant (2), Whole Milk(2), Dark Chocolate(2), Hot Tea(1), Orange Juice(1), Whipped Cream(2), Red Wine(1), White Wine(2), Mushrooms(1) and Ice Cream (1). Since none of the identified significant SNPs was replicated in the SR cohort we decided to replicate with regional association analysis. We defined around each SNP a 10kb region and estimated region wide significance using PCA-LRT as described in Wang and Abbot 2008. If the SNP fell inside a gene the whole gene was also tested. This type of analysis allowed us to replicate 3 loci one for Artichokes, one for Eggplant and one for Whole Milk. This study shows for the first time replicated genome wide loci for liking of common foods and represent a good starting point in understanding the genetic bases of food preferences and consumption in general.

957W

Analysis of Exomic Variation in the Ashkenazi Jewish Population Identifies Novel Associations to Crohn's Disease in LRRK2. *K. Hui[1], W. Zhang[1], T. Haritunians[2], S. Carmi[3], B.M. Bowen[1], S.R. Brant[4], J.D. Rioux[5], M. Silverberg[6], S. Katz[7], A. Chaifetz[8], H. Zhao[9], G. Atzmon[10], L. Ozelius[11], S. Bressman[12], L.N. Clark[13], I. Pe'er[3], T. Lencz[14], D.P. McGovern[15], R.H. Duerr[16], J.H. Cho[1], I. Peter[11].* 1) Internal Medicine and Genetics, Yale University, New Haven, CT; 2) Genetics, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Computer Science, Columbia University, New York, NY; 4) Medicine, Johns Hopkins University, Baltimore, MD; 5) Medicine, University of Montreal, Montreal, QC, Canada; 6) Medicine, University of Toronto, Toronto, ON, Canada; 7) Medicine, North Shore-Long Island Jewish Health System, Manhasset, NY; 8) Medicine, Harvard University, Boston, MA; 9) Biostatistics, Yale University, New Haven, CT; 10) Medicine, Albert Einstein College of Medicine, New York, NY; 11) Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY; 12) Neurology, Albert Einstein College of Medicine, New York, NY; 13) Pathology and Cell Biology, Columbia University, New York, NY; 14) Psychiatry, Hofstra North Shore-LIJ School of Medicine, Manhasset, NY; 15) Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; 16) Medicine, University of Pittsburgh, Pittsburgh, PA.

Background Crohn's disease (CD) is 4.3-7.7 times more prevalent in the Ashkenazi Jewish (AJ) population compared to non-Jewish European-ancestry populations. The unique demographic history of the AJ population, characterized by an extreme bottleneck, subsequent explosive population growth, and endogamy, has resulted in increased sharing of long haplotypes. We showed previously that common polymorphisms in established CD loci did not account for the AJ population's increased disease prevalence. In this study, we hypothesized that uncommon variation, particularly within protein-coding exons poorly assayed by GWAS, contributes uniquely to CD in the AJ population. Methods and results We performed exome sequencing of 50 AJ individuals with CD and identified 4,277 novel coding polymorphisms of interest, which we combined with markers on the Illumina HumanExome beadchip. Using this custom array, we conducted genotyping in 1,463 CD unrelated cases and 2,770 independent healthy controls, which were genetically validated as having full AJ ancestry. We detected conditionally independent novel AJ-enriched association signals in leucine-rich repeat kinase 2 (*LRRK2*) at the low-frequency coding mutations N2081D and N551K (OR=1.8, $P=1.0\times10^{-11}$ and OR=0.66, $P=2\times10^{-5}$, respectively). This elaborates our prior CD finding of genome-wide significant association in the *LRRK2* region in a large study of European-ancestry individuals, as well as our previously detected linkage peak with a LOD score of 2.08, driven by AJ families. Both mutations were previously linked to Parkinson's disease (PD), with N2081D conferring a 24% increased risk, while N551K showed a 12% decreased risk in at least one population. Interestingly, *LRRK2* G2019S, which is very strongly associated with PD in the AJ population and occurs in the same domain as N2081D, showed no evidence of association with CD. Combined analysis of genotype data from Ashkenazi PD and CD cases, based on AJ population-specific imputation using a reference panel of 128 whole genome-sequenced individuals, suggested additional extensive genetic pleiotropy within the extended LRRK2 locus. Discussion These findings demonstrate the critical role that uncommon alleles play in CD heritability, providing a key link between structure and function. Our results also point towards complex relationships between the genetic risks that can be uncovered by performing a focused study in the Ashkenazi Jewish population.

958T

Genome-wide association analysis of diverse immune-related phenotypes highlights complex overlapping pathways of immune response. *J.Y. Tung, N. Eriksson, A.K. Kiefer, D.A. Hinds.* 23andMe, Inc, Mountain View, CA.

The immune system's defense of the body against assault requires a complex interplay of cell-mediated and humoral immunity including a multitude of cell types, antigen processing and presentation systems, cytokines, and inflammatory factors. As a result, teasing out which components underlie each individual's susceptibility to immunological conditions is a daunting task. Genome-wide analyses studying these conditions can shed light on which pathways play the biggest roles for different types of conditions. We compared the top associations for a diverse set of inflammatory phenotypes including poison oak contact dermatitis, mosquito bite size, and tonsillectomy (as a proxy for chronic tonsillitis) with each other and with published associations with autoimmune, allergic, and infectious conditions. For inflammatory phenotypes, significant hits near *IL21* and *GMCSF2*(mosquito bite size) and *OSM*, *LTBR*, *CXCL13*, *IGFBP3*, and *IKZF1* (tonsillectomy) suggest a more prominent role for cytokines and inflammatory pathways for these conditions than in allergy or infectious disease. There are several overlaps as well. For tonsillectomy, the variant near *IKZF1* is in high linkage disequilibrium with a variant strongly associated with systemic lupus erythematosus, an autoimmune condition. The region near *IL2* and *IL21* associated with mosquito bite size is also associated with allergy, though the respective variants are not closely linked. In addition, like for many autoimmune, allergic, and infectious conditions, the HLA region comes up as an important hit for both mosquito bite size and tonsillectomy. In our data, the same variant between *HLA-B* and *MICA* (rs138099588) associated with tonsillectomy is also strongly associated with psoriasis ($p = 6.57e-52$) and shingles ($p = 2.05e-8$). For poison oak contact dermatitis, we see a hit within a cluster of *CD1* genes on chromosome 1, rather than in the HLA region. CD1 proteins mediate presentation of lipids to T cells (instead of proteins like the major histocompatibility complex molecules). This is consistent with the role of urushiol, the oily substance produced by the poison oak plant, in triggering the dermatitis. These results highlight the myriad of mechanisms used by the immune system, as well as the subtle similarities and differences in those mechanisms, to react to toxic substances and invaders in the environment.

959F

Joint association analysis of genome-wide human and HIV-1 variation. *I. Bartha[1,2,3], J. Carlson[4], P.J. McLaren[1], Z. Brumme[5], Ch. Brumme[6], R. Harrigan[6], A. Rauch[7], H. Günthard[8], M. John[9], D. Heckerman[4], T.M. Allen[10], C.L. Galíndez[11], J. Martinez-Picado[12], V. Müller[3], A. Telenti[2], J. Fellay[1], HIV Genome to Genome Study.* 1) School of Life Sciences, École Polytechnique Fédéral de Lausanne, Lausanne, Switzerland; 2) University Hospital and University of Lausanne, Lausanne, Switzerland; 3) Eötvös Loránd University, Budapest, Hungary; 4) eScience Group, Microsoft Research, Los Angeles, CA, USA; 5) Simon Fraser University, Burnaby, BC, Canada; 6) BC Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada; 7) University Hospital Bern & University of Bern Inselspital, Bern, Switzerland; 8) University Hospital and University of Zürich, Zürich, Switzerland; 9) Murdoch University, Murdoch, WA, Australia; 10) Ragon Institute of MGH, MIT and Harvard, Massachusetts General Hospital, Boston, MA, USA; 11) Instituto de Salud Carlos III, Madrid, Spain; 12) Hospital Universitario Germans Trias i Pujol, Badalona, Spain.

Background: Joint analysis of human genetic and HIV sequence variation has been largely limited to HLA alleles and viral mutations in corresponding epitopes. We performed an unbiased genome-wide survey of associations between human SNPs and variation across the HIV proteome. We hypothesized that this non-a priori genome-to-genome analysis would identify and map all host selective pressures on the viral genome.

Methods: Human genome-wide genotyping and HIV-1 full-length sequencing data were available for the study. Binary variables were created for each variable amino-acid positions, for every amino acid that was present in at least 3 HIV genomes. Human SNP imputation was performed using 1000 Genomes data as reference. Associations between all SNPs and HIV-1 amino acids were tested by logistic regression under an additive genetic model. For each amino acid that had genome-wide significant association, we searched for independently associated SNPs by iteratively conditioning on the most significant SNP.

Result: A total of 1071 patients of European ancestry from 7 cohorts and 5 countries were included in the study. After imputation 6 889 656 SNPs were tested for association with more than 3000 different HIV residues. Highly significant associations ($p < 1E-11$) were observed between SNPs in the Major Histocompatibility Complex (MHC) and multiple amino acids in several HIV-1 proteins. SNPs tagging HLA class I alleles strongly associated with viral variation in CTL epitopes targeted by the corresponding alleles. No significant signals were identified outside the MHC.

Conclusion: A non a priori genome-to-genome approach maps associations between host SNPs and HIV genomic variation. We confirm the extensive evolutionary effect that MHC exerts on HIV Gag, Pol and Nef and the lack of common variants with strong effects elsewhere in the host genome on within-host viral evolution.

960W

Shared genetic background for Leprosy an Inflammatory Bowel Disease: chinks in the primary defense against pathogens. *E. Festen[1,2,10], A. Kresentia Irwanto[3,4,10], S. Ripke[5,6], D. Ellinghaus[7], R.K. Weersma[2], A. Franke[7], J.J. Liu[3,4,8,9], IIBDGC.* 1) Department of Genetics, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands; 2) Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands; 3) Human Genetics, Genome Institute of Singapore, Singapore; 4) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; 6) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) Institute of Clinical Molecular Biology, Christian-Albrechts- University, Kiel, Germany; 8) Shandong Provincial Hospital for Skin Diseases, Shandong University, Jinan, China; 9) School of Biological Sciences, Anhui Medical University, Hefei, China; 10) Authors contributed equally to this work.

Introduction: IBD and leprosy are known to have several overlapping genetic risk loci and the histology of the inflammation is similar. The goal of this study is to identify new shared genetic risk variants for inflammatory bowel disease (IBD) and leprosy, which can shed light on shared biology for the diseases. A secondary goal was to discover new risk loci for each of the two diseases.

Methods: We compared 9026 IBD cases and 15256 population controls of European descent to 706 Leprosy cases and 1225 controls of Chinese descent. The cohorts were genotyped on various genome-wide platforms and equalized by imputation with the HapMap 1000-Genomes reference set (feb'12 release) on IMPUTE2. Because the cohorts have different ethnic descent the MANTRA algorithm was used for their meta-analysis. The MANTRA algorithm was especially developed for the meta-analysis of cohorts of different ethnic backgrounds. For replication an IBD cohort of 869 cases and 2023 controls of European descent, and a Leprosy cohort of 545 cases and 700 controls of Chinese descent, are being analyzed.

Results: Eight known shared risk loci for IBD and leprosy were confirmed in this study: *IL23R*, *TNFSF15*, *RIPK2*, *IL12B*, *LACC1*, *HLA-DR/DQ*, *LRRK2*, and *NOD2*. We identified several potentially new shared risk loci that are currently being replicated in independent cohorts: *REL*, *IL3*, *CD6*, *NUCB2*, *NFATC1*, *UBE2L3*, *MST1*, *C11orf30*, *ZNF365*, *IL10* and *JAK2*.

Conclusion: IBD and leprosy share a large number of risk loci. Most shared risk loci point toward an impairment of the innate immune system as the shared biological background of IBD and leprosy.

961T

Novel Genetic Loci Identified for HIV-1 Acquisition: a 1000 Genomes-Imputed Genome-wide Association Study among Injection Drug Users and a Replication. *E.O. Johnson[1], D.B. Hancock[1], N.C. Gaddis[1], J.L. Levy[1], G.P. Page[1], S.P. Novak[1], C. Glasheen[1], N.L. Saccone[2], J.P. Rice[2], Q. Wang[3], M. Moreau[3], K. Doheny[4], J. Romm[4], B. Aouizerat[5], A.I. Brooks[3], L.J. Bierut[2], A.H. Kral[1].* 1) RTI International, Research Triangle Park, NC; 2) School of Medicine, Washington University, St. Louis, MO; 3) Rutgers University Cell and DNA Repository, Piscataway, NJ; 4) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 5) School of Nursing, University of California San Francisco, San Francisco, CA.

Host genetic factors likely influence susceptibility to acquiring HIV-1 upon exposure. However, besides a rare 32-base pair deletion in the CCR5 gene, no replicable genetic associations have been reported. To identify novel genetic factors, we conducted a genome-wide association study (GWAS) of HIV-1 acquisition in 2,004 African Americans and 1,132 European Americans from the Urban Health Study, one of the largest studies of street-recruited injection drug users in North America. Study participants were frequency matched on several important environmental and behavioral risk factors to create an approximate 2:1 ratio of high-risk HIV-negative controls to HIV-positive cases. After genotyping on the Illumina Omni1-Quad BeadChip and applying standard quality control metrics, 1000 Genomes imputation was conducted using IMPUTE2 with reference to the ALL panel, resulting in the availability of ~8 million single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF)>0.5%. Genotyped and imputed SNPs were tested for association with HIV-1 acquisition in the separate ancestral groups using logistic regression models adjusted for age, gender, survey year, behavioral risk class, and 10 eigenvalues. The ancestral-specific GWAS results were then combined via fixed-effects meta-analysis. Eight genetic loci of interest were identified: 1 locus with statistically significant SNP associations ($P<5\times10^{-8}$), 6 loci with nominally significant SNP associations (P between 5×10^{-8} and 1×10^{-6}), and 1 locus with the top genotyped SNP association ($P=1\times10^{-5}$). Across the 8 loci of interest, SNPs with $P<1\times10^{-3}$ were selected for independent replication testing in 1,852 African Americans and 681 European Americans from the Women's Interagency HIV Study. One chromosome 9p13.2 SNP ($P=1.38\times10^{-4}$) exceeded the statistical significance threshold for replication ($P<3.33\times10^{-4}$ based on correction for multiple testing). Several SNPs from 3 other loci on chromosomes 5q31.2, 6p21.32, and 9p24.1 were replicated at nominal significance (minimum P ranging from 1.18×10^{-3} to 3.64×10^{-3}). The GWAS-implicated loci with evidence for replication span genes that have biologically plausible roles in the immune system, some of which have specific functional and regulatory links to protease inhibition and HIV viral DNA integration in host cell DNA. Additional replication testing is underway to further support the statistical evidence for association with HIV-1 acquisition.

962F

Genome-wide discovery of genetic predictors of weight-loss in obese subjects in response to diet and exercise. *C.M. Molony[1], D.F. Reilly[1], H. Zhou[1], S. Heymsfield[3], B. Goldstein[2], M.L. Reitman[4], L. Tacconi[5], N. Thornberry[2].* 1) Merck Research Laboratories,Informatics and Analysis, Merck & Co., Inc., Boston, MA; 2) Merck Research Laboratories, Whitehouse Station, NJ, USA; 3) Pennington Biomedical Research Center, Baton Rouge, LA, USA; 4) NIDDK, National Institutes of Health, Bethesda, MD, USA; 5) Vree Health, a wholly owned subsidiary of Merck and Co., Inc., Whitehouse Station, NJ, USA.

Diet and exercise are the first-line treatments for obesity. Since the probability of successfully achieving long-term (> 12 months) weight-loss is modest, clinicians and patients would benefit from development of robust early biomarkers of weight-loss response. Here we report on the first genome-wide association study in placebo-treated subjects pooled from multiple weight-loss trials in order to identify genetic predictors of long-term weight-loss with diet and exercise intervention. DNA samples were available from 355 placebo-treated subjects who also followed a diet which targeted a 25% caloric restriction and an exercise regimen designed according to increase the patient's current level of physical activity. This cohort had a mean baseline BMI 35.0 (s.d. 4.0) and mean age 49.6 (s.d. 11.9) an overall weight-change of -3.3kg (CI -4.0, -2.7) over the 52 week treatment period. DNA samples were genotyped on the Illumina Human 1M DNA Beadchip and additional marker data were inferred using imputation approaches with the 1000genomes as reference. Depending on the population there were approximately 6 million markers available for analysis after imputation. After adjustments for multiplicity and undetected stratification, the analysis yielded two significant and three suggestive loci in the White population. The largest signal was observed at Chr19p13.1 for the SNP rs147773401 lying intronic to gene LPHN1 with a p-value of 5.88×10-12 and covariate adjusted genotypic means of -2.7kg (CI -3.3,-2), -6.7kg (CI -8.4,-5.1), and -29.8kg (CI -38.9,-20.7). A sensitivity analysis of this marker after removing extreme weight-loss subjects yielded an association signal that still met suggestive thresholds (p-value<5×10-7). Another single nucleotide deletion at chr10.109889967 was also associated with a p-value 5.85×10-8 with a covariate adjusted mean weight-loss of -13.9kg (CI -17.2,-10.5) for the most extreme genotype group. Genetic predictors of large effect may aid in helping to distinguish responders from non-responders in order to make trial designs more efficient and improve benefit/risk/cost ratios in medical practice. However obesity is a multifactorial process and our data suggest that it is unlikely that any single genetic marker will robustly characterize the response to a diet plus exercise intervention for a large segment of the patient population.

963W

Plasma Lipids, Chromosome 11q23.3, and the Risk of Infantile Hypertrophic Pyloric Stenosis. *B. Feenstra[1], F. Geller[1], L. Carstensen[1], P.A. Romitti[2], I. Baranowska Körberg[3], B. Bedell[4], C. Krogh[1], R. Fan[5], A. Svenningsson[3], M. Caggana[6], A. Nordenskjöld[3], J.L. Mills[5], J.C. Murray[4], M. Melbye[1].* 1) Dept. of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Dept. of Epidemiology, University of Iowa, Iowa City, Iowa; 3) Dept. of Women's and Children's Health and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 4) Dept. of Pediatrics, University of Iowa, Iowa City, Iowa; 5) Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, Maryland, USA; 6) Wadsworth Center, New York State Department of Health, Albany, New York.

Infantile Hypertrophic Pyloric Stenosis (IHPS) is a serious condition in which hypertrophy of the pyloric sphincter muscle layer leads to gastric outlet obstruction. Currently, it is the most common condition requiring surgery among newborns. To search the genome for genetic associations with IHPS, we used reference data from the 1000 Genomes Project for imputation into a genome-wide dataset of 1,001 Danish surgery-confirmed cases and 2,371 disease-free controls. The five most strongly associated loci were tested in replication sets from Denmark, Sweden, and the United States with a total of 1,663 cases and 2,315 controls. We found a novel genome-wide significant locus for IHPS at chromosome 11q23.3 in a region harboring the apolipoprotein (APOA1/C3/A4/A5) gene cluster. The most significant SNP at the locus (odds ratio, 1.59; $P = 1.9\times10^{-10}$) is strongly correlated with SNPs previously found to be associated with levels of circulating cholesterol. For these SNPs, the cholesterol lowering allele consistently conferred increased risk of IHPS. To further investigate the role of lipid levels at birth for IHPS risk, we conducted a functional follow-up study based on umbilical cord blood samples from 46 cases and 189 controls of Danish ancestry sampled from the Danish National Birth Cohort. We found that IHPS risk was inversely associated with total cholesterol levels at birth. There was a 64% (95% confidence interval -83% to -26%; P = 0.005) lower risk per mmol per liter and the 25% of newborns with the lowest levels of total cholesterol were at almost five times higher risk compared with the 25% with the highest levels. In conclusion, low levels of plasma cholesterol, partly genetically-driven, were associated with higher risk of IHPS. This association suggests the possibility that dietary modifications to ensure normal lipid levels could reduce the risk.

964T

Meta-analysis of genome-wide association studies identifies three new susceptibility loci for intracerebral hemorrhage. *G.J. Falcone[1,2,3,4], D. Woo[5], C.D. Langefeld[6], J. Rosand[1,2,3], International Stroke Genetics Consortium.* 1) Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA; 2) Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Hemorrhagic Stroke Research Group, Massachusetts General Hospital, Boston, MA; 4) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 5) Department of Neurology, University of Cincinnati College of Medicine, Cincinnati, OH; 6) Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest University, Winston-Salem, NC.

BACKGROUND: Intracerebral hemorrhage (ICH) is the stroke type with the worst prognosis and has no established treatment. ICH is classified as lobar and nonlobar based on histopathological differences, with genetic variation contributing to risk of both subtypes. We report the first genome wide association study (GWAS) of this condition. **METHODS**: The discovery phase included 6 ICH GWASs that enrolled subjects from European ancestry in the US and Europe. Cases were ascertained by neurologists blinded to genotype data and classified as lobar or nonlobar based on brain computed tomography. ICH-free controls were selected from ambulatory clinics or by random digit dialing, depending on the study. Genotyping was completed in Illumina 610 or Affymetrix 6.0. Standard quality controls filters were applied and principal component analysis was implemented to account for population structure. Unobserved genotypes were inferred by imputation to 1000 genomes references panels, with subsequent filtering of single nucleotide polymorphisms (SNPs) with minor allele frequency (MAF) <1% or low imputation quality. Association testing was carried out using logistic regression assuming additive genetic effects and including age, gender and principal components in the model. Meta-analyses were completed using the inverse normal method, weighting by sample size. Replication of signals with p<5×10E-6 was pursued in an independent multi-ethnic study utilizing direct genotyping. **RESULTS**: The discovery phase included 1545 cases (664 lobar and 881 nonlobar) and 1481 controls and identified 4 possible loci: for lobar ICH, chromosome 12q21.1 (rs11179580: MAF=0.22, odds ratio [OR]=1.54, p=6.7×10E-8) and 8p23.1 (rs1630950: MAF=0.37, OR=1.33, p= 7.1×10E-7); for nonlobar ICH, chromosome 1q22 (rs2984613: MAF=0.31, OR=1.45, p=1.8×10E-8) and 3q13.13 (rs4682240: MAF=0.07, OR=2.04, p= 1.4×10E-06). Replication included 1256 cases (329 whites, 506 blacks and 421 Hispanics) and 760 controls (237 whites, 346 blacks and 177 Hispanics) and confirmed 3 loci: 12q21.1 (p=0.037, meta-analysis p=2.3×10E-8), 1q22 (p=0.009, meta-analysis p=2.0×10E-9) and 3q13.13 (p=0.048, meta-analysis p=6.7×10E-7). Chromosome 8p23.1 did not replicate (p=0.72). **CONCLUSION**: Three susceptibility loci for ICH were identified, all showing subtype-specificity. These results confirm the presence of biological heterogeneity across ICH subtypes and highlight the importance of ascertaining ICH cases accordingly.

965F

Association between Obsessive-Compulsive Disorder (OCD) and Phenotypes in a Genome-Wide Association Study. *G. Zai[1, 2, 3], C. Zai[1, 2], J. Knight[1, 3], J.L. Kennedy[1, 2, 3], M.A. Richter[2, 3, 4].* 1) Neurogenetics Section, CAMH, Univ Toronto, Toronto, ON, Canada; 2) Department of Psychiatry, University of Toronto, ON, Canada; 3) Institute of Medical Science, University of Toronto, ON, Canada; 4) Department of Psychiatry, Sunnybrook Health Science Centre, Toronto, ON, Canada.

Background: Obsessive-compulsive disorder (OCD) is a chronic and debilitating psychiatric disorder with a strong genetic etiology. A recent genome-wide association study (GWAS) reported interesting candidate gene variants with potential susceptibility to developing OCD. Hypothesis: We hypothesized that genetic variation(s) may be associated with OCD symptom severity using the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) severity score and age at onset (AAO). Method: We investigated 357 individuals with OCD and their nuclear families or matched healthy control (age, gender, and ethnicity) for an association of gene variant(s) and phenotypes including Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) severity score and age at onset (AAO) in a GWAS. GWAS was conducted by the OCF Genetic Collaborative Group. Quality control and analyses were conducted using PLINK and R programs. Results: Several suggestive genome-wide association signals were detected with a p-value of 10E-6 on chromosome 3 and 22 for Y-BOCS severity score and chromosome 3 and 9 for AAO. Conclusion: Preliminary findings from our GWAS suggested possible involvement of several regions in OCD symptom severity and age of onset. Further analyses are required to characterize these results.

966W

Genome-wide association analysis in a 23andMe cohort identifies novel associations with uterine fibroids. *A.S. Shmygelska, N.K. Eriksson, J.Y. Tung, J.L. Mountain, U. Francke, A.K. Kiefer, D.A. Hinds.* 23andMe Inc., Mountain View, CA.

Uterine leilomyomata, commonly known as uterine fibroids, are benign tumors derived from smooth muscle and fibrous tissue in the uterus, and are the leading cause of hysterectomy in the United States. The lifetime risk for a woman to develop fibroids has been estimated to be as high as 25%. Fibroids tend to grow under the influence of estrogen and shrink when estrogen levels are reduced. The underlying causes of uterine fibroids are not well understood, but twin studies suggest that approximately 55% of the variation in susceptibility to fibroids is genetic. To investigate the genetic factors underlying uterine fibroids, we conducted a genome-wide association study (GWAS) of over 4,000 cases and 12,000 controls of unrelated individuals with European ancestry from the 23andMe cohort. We imputed genotypes against 1000 Genomes reference haplotypes. Participants reported *via* a web-based survey whether or not they had had uterine fibroids. We report one novel genome-wide association and four suggestive associations. The most significant finding is a variant in the spectrin repeat containing nuclear envelope 1 (SYNE1) gene (rs71575922: odds ratio=1.3, $p = 4.8 \times 10^{-10}$), which is 18 kbp upstream of the ESR1 gene that encodes estrogen receptor alpha. Variants in ESR1 have been previously shown to be associated with breast and endometrial cancer. SYNE1 is expressed in skeletal and smooth muscle, and peripheral blood lymphocytes, and localizes to the nuclear membrane. Other suggestive associations include (1) the lipoprotein(a) (LPA) gene (rs1800769: odds ratio=1.2, p=8.8x 10^{-7}), which has been shown to promote smooth muscle cells proliferation in transgenic rabbits and (2) the rs12789861 variant (odds ratio=0.9, p=6.7x 10^{-7}) 40 kbp downstream of the Wilms tumor protein (WT1), whose product interacts with the estrogen receptor alpha in breast cancer cells, and downstream of the reticulocalbin 1 (RCN1) gene which is involved in muscle development and tumorigenesis. We also detect an association between uterine fibroids and the recoverin (RCVRN) gene (rs6503269: odds ratio=1.2, p=4.4x 10^{-7}) that appears non causal. RCVRN encodes a neuronal calcium sensor that plays role in the phototransduction cascade in the retina and is implicated in endometrial cancer associated retinopathy. In our European cohort we replicate two of the three previously identified associations reported in women of Japanese ancestry, however, not on the genome-wide significance level.

967T

Identification of susceptibility genes of adult asthma in French Canadian women. *J.C. Berube[1], E. Lavoie-Charland[1], N. Gaudreault[1], L. Sbarra[1], C. Henry[1], L.P. Boulet[1], Y. Bosse[1,2].* 1) Institut universitaire de cardiologie et de pneumologie de Québec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada.

Introduction: Asthma is a heterogeneous disease in which several susceptibility genes have been identified, but together, those genes explain only a small fraction of the heritability. It was postulated that additional genetic factors may be found by studying subgroups of phenotypically similar asthmatic patients. **Objective:** The goal of this study is to identify single nucleotide polymorphisms (SNPs) associated with asthma in adult women of European ancestry. **Methods:** A pooling-based genome-wide association study was performed in 240 allergic asthmatic women and 120 non-asthmatic allergic women. The genotyping was performed in six replicates for both pools using the Ilumina HumanOmniExpress BeadChip. The 730 525 interrogated SNPs were ranked based on their likelihood of being associated with asthma using the silhouette score calculated by the GenePool software. The top 20 ranked SNPs plus 18 SNPs ranked among the top 2000 and near candidate genes were selected for individual genotyping in an extended cohort of 349 asthmatic women and 261 non-asthmatic women using the Illumina BeadXpress platform. **Results:** 21 of the 38 SNPs tested by individual genotyping showed P values lower than 0.05 for association with asthma. SNPs rs17655581 (P = 8.2 × 10-5) and rs7980829 (P = 2.4 × 10-4) were the most significant. These SNPs are located in intergenic regions uncharacterized for asthma. The third most important genetic signal was found with rs803010 (P = 4.4 × 10-4) located in the promoter of PTGDR, which is a gene previously associated with asthma in individuals of European ancestry. **Conclusion:** This pooling-based GWAS in French Canadian adult women identified new loci associated with asthma. This study also supported the role of PTGDR in this more homogenous subgroup of asthmatic patients. Further validation is required in independent cohorts of adult asthmatic women of European ancestry.

968F

Familial history of chronic rhinosinusitis predicts more severe disease. *L. Mfuna Endam[1], A. Filali-Mouhim[1], P. Boisvert[2], LP. Boulet[3], Y. Bossé[3], M. Desrosiers[1,4].* 1) Department of Otolaryngology, CHUM HOTEL DIEU, CRCHUM, Montreal, PQ, Canada; 2) Department of Otolaryngology, Saint-François d'Assise Hospital, Québec, PQ, Canada; 3) Centre de Recherche, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec, PQ, Canada; 4) Department of Otolaryngology-Head and Neck Surgery, Montreal General Hospital, McGill University, Montreal, QC, Canada.

Introduction: Genetic association studies have linked chronic rhinosinusitis (CRS) with a number of biologically-plausible genes. However, effects are often small and account for only a portion of CRS cases. Recent evidence suggested that pathogens contributing to CRS development share similar pathways as those revealed by genetic association studies. Cohorts of individuals with CRS may thus contain individuals with genetically-determined and acquired disease, thereby diluting the strength of the genetic signal. We wished to determine whether CRS populations would be enriched for genetic basis to disease by restricting analysis to the sub-group of subjects with a positive family history of CRS. More specifically we tested whether this sub-group was associated with i) a distinct disease evolution and ii) increased strength of previously identified genetic associations. Methods: An existing population of 198 French-Canadian patients with CRS refractory to medical and surgical treatment were screened for incidence of CRS in relatives and impact on population demographics and previously reported genetic associations. PLINK software was used to perform genetic association tests. Results: 38% of subjects had a family history of CRS. This group had more severe disease as suggested by an earlier age at onset, an earlier age at first endoscopic sinus surgery and a greater number of surgeries. However, analysis of previously reported SNPs did not show a great variation in strength of previously reported genetic associations. Conclusion: Family history of CRS appears to select for a sub-group of patients with more severe disease suggesting a large genetic component. Exclusion of individuals with a lower risk of a genetic component by restricting patient recruitment or analysis to individuals with a family history of CRS may help better identify candidate genes implicated in development of CRS.

969W

Refinement of whole-body bone mineral density measures of children assists the identification of genetic variants associated with skeletal site specificity and bone mass attainment. *J.P Kemp[1,2], C. Medina-Gomez[3,4,5,6], K. Estrada[3,5,6], B. St-Pourcain[1], D.H.M. Heppe[4,5,7], N.J. Timpson[1,2], L. Oei[3,5,6], S.M. Ring[1], C.J. Kruithof[4,5], L.E. Wolber[8], F.M.K. Williams[8], M.C. Zillikens[3,6], A. Hofman[4,5,6], A.G. Uitterlinden[3,4,5,6], G. Davey-Smith[1,2], V.W.V. Jaddoe[4,5,7], J.H. Tobias[9], F. Rivadeneira[3,4,5,6], D.M. Evans[1,2].* 1) MRC Centre for Causal and Translational Epidemiology, University of Bristol, Bristol, United Kingdom; 2) School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 3) Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) The Generation R Study Group, Erasmus University Medical Center, Rotterdam, The Netherlands; 5) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 6) Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), The Netherlands; 7) Department of Paediatrics, Erasmus University Medical Center, Rotterdam, The Netherlands; 8) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 9) School of Clinical Sciences, University of Bristol, Bristol, UK.

The heritability of bone mineral density (BMD) varies across skeletal sites, reflecting different relative contributions of environmental and genetic influences. To quantify the degree to which common genetic variants and environmental factors influence BMD at different sites, we estimated the genetic (r_g) and environmental (r_e) correlations of BMD measured at the upper limbs (UL), lower limbs (LL) and skull (S), using whole-body DXA scans of ~4890 participants recruited by Avon Longitudinal Study of Parents and their Children (ALSPAC). Point estimates of r_g indicated that appendicular sites shared a greater proportion of shared genetic architecture (LL-/UL-BMD r_g= 0.78) when compared to the skull (UL-/S-BMD r_g=0.58 and LL-/S-BMD r_g= 0.43). Likewise, environmental factors influencing BMD at the appendicular sites were broadly similar to each other (r_e=0.55), however they were appreciably different from the factors influencing S-BMD (r_e=0.20 - 0.24). To explore the basis for the observed differenced in r_g and r_e, genome-wide association meta-analyses were performed (n~9395), combining data from ALSPAC and the Generation R study (GEN-R) to identify genetic signals associated across different skeletal regions. Twelve known BMD associated variants differed in the strength of their association and magnitude of effect with each sub-region. In particular, effect sizes of variants situated closest to *EYA4* (6q23.2), *COLEC10* (8q24.12), *LIN7C* (11p14.1) and *TNFRSF11A* (18q21.33) appeared to be largest for S-BMD when compared to UL- and LL-BMD. Furthermore variants at the *WNT16* and *CENPW* showed considerable site-specificity as indicated by very strong levels of association with S-BMD and/or UL-BMD but not with LL-BMD. In addition, we report a novel association between a variant in *RIN3* and LL-BMD (rs754388: β=0.13, SE= 0.02, P=1.4×10^{-10}) and highlight its prior association with Pagets disease. We suggest that different skeletal sites as measured by whole body-DXA are to a certain extent under distinct environmental and genetic influences. Allowing for these differences may help to uncover new genetic influences on BMD, particularly those involved in bone accrual, for which S-BMD appears to be particularly well suited.

970T

GWAS SNPs impact gene expression through inheritance of multiple enhancer variants. *O. Corradin[1], A. Saiakhova[1], B. Akhtar-Zaidi[1], L. Myeroff[2], J. Willis[2,3], S. Markowitz[1,2,4], P.C. Scacheri[1,2].* 1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH 44122; 2) Department of Pathology, Case Western Reserve University, Cleveland, OH 44122; 3) Department of Medicine, Case Western Reserve University, Cleveland, OH 44122; 4) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106.

Since 2005 over 1350 genome-wide association studies (GWAS) have been published, identifying thousands of SNPs associated with more than 600 traits. More than 90% of these SNPs are located outside protein coding genes, and are hypothesized to influence gene expression by disrupting the function of non-coding regulatory elements such as enhancers. Consistent with this hypothesis, the ENCODE consortium and other groups have localized thousands of GWAS variants to enhancer elements identified through epigenomic profiling studies. However, the genes whose expression these enhancer variants influence and their effect on target transcript levels often remains unknown. Moreover, because GWAS-loci often contain multiple SNPs in linkage disequilibrium (LD) with the 'lead' SNP, any one of which may be 'causative', pinpointing the culprit is often a major challenge. We provide evidence that for six common autoimmune disorders (rheumatoid arthritis, Crohn's disease, celiac disease, multiple sclerosis, lupus, and ulcerative colitis), the GWAS-association arises from a group of polymorphisms in LD that map to multiple enhancer elements active in the same cell type. This finding suggests that for common traits, there is not necessarily a single disease-causing variant that underlies the association signal, but rather several variants that impact multiple enhancers and cooperatively affect gene expression. We call this the 'multiple enhancer variant' hypothesis. Using a novel method to delineate enhancer-gene interactions, we show that multiple enhancer variants within a given locus typically target the same gene. Using available data from HapMap and B lymphoblasts as a model system, we provide evidence at numerous loci that multiple enhancer variants cooperatively contribute to altered expression of their gene targets. The effects on target transcript levels can be either gain or loss of function, and overall tend to be modest. Overall, the multiple enhancer variant hypothesis offers a new paradigm by which non-coding variants confer susceptibility to common traits.

971F

Novel Genetic Association of Primary Severe Localized Provoked Vulvodynia with TRPV1 and NGF; Possible Common Predisposition with Other Pain Syndromes. *T. Falik-Zaccai[1,2], L. Kalfon[1], A. Azran[3], Y. Farajun[3], E. Tubin[1], O. Hemo[1], L. Abramov[4,6], A. Yeshaya[5,6], J. Bornstein[2,3].* 1) Institute of Human Genetics, Western Galilee Medical Center, Nahariya, Israel; 2) The Galilee Faculty of Medicine, Bar Ilan, Safed, Israel; 3) Department of Obstetrics & Gynecology, Western Galilee Medical Center, Nahariya, Israel; 4) Lis Women's Hospital, Tel Aviv, Israel; 5) Shnaider Women's Hospital, Beilinson Medical Center, Petah-Tikva ,Israel; 6) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Localized Provoked Vulvodynia (LPV) is a multifactorial syndrome affecting 10-16% of women worldwide, causing pain of variable nature and intensity at every attempt to have sexual intercourse. Familial occurrences of LPV suggest genetic susceptibility in the etiology of this condition. We have studied possible associations between LPV and several SNPs in genes previously hypothesized to be involved in the pathophysiology of LPV by coding for proteins involved in mast cells degradation and hyper-innervation in the vestibule of affected women. We have analyzed and compared the prevalence of seven SNPs in each of three genes between 70 women presenting severe primary LPV and 132 healthy, ethnically matched controls. Each participating women have answered a detailed questionnaire, addressing possible familial occurrence of LPV and the existence of pain co-morbidity conditions. SNP analyses were performed for the genes Transient Receptor Potential Vanilloid type-1 (TRPV1), the Nerve Growth Factor (NGF) and the Heparanase (HPSE), using multiplex PCR and SNapShot assay and by restriction fragment length polymorphism. We found a significant difference in familial occurrences of LPV, Temporo mandibular joint (TMJ) symptoms, recurrent vaginitis and irritable bowel syndrome (IBS) between affected women and the healthy controls. Novel statistically significant differences in the prevalence of non-synonymous polymorphism c.945G>C (rs222747) of TRPV1 and a SNP located in the promoter region of NGF (rs11102930) were identified between the affected women especially from Ashkenazi Jewish ancestry and the control group. Data was also analyzed to compare genetic variation with a variety of pain condition among the 202 women. Interestingly, rs222747 minor allele of TRPV1 was found in association with women presenting TMJ and women with recurrent vaginitis, suggesting possible common genetic predisposition to pain co-morbidities. Our results imply possible genetic susceptibility to LPV associated with specific alleles in the genes TRPV1 and NGF, and propose a genetic predisposition for LPV and other pain syndromes through a yet unknown biological pathway involving TRPV1.

972W

Predicting regions associated with complex traits using Multi-kernel Support Vector Machines. *D. Kostka[1], J.A. Capra[2].* 1) Developmental Biology and Computational & Systems, University of Pittsburgh, Pittsburgh, PA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Genome-wide association studies (GWAS) have identified thousands of genomic loci associated with hundreds of complex traits in different populations. However, these associations likely represent only a fraction of the loci contributing to the genetic basis of most complex traits. To aid the search for additional loci and improve our understanding of known associations, we developed an algorithm to distinguish genomic regions containing SNPs associated with a trait from regions associated with other traits and from the genomic background. For 59 traits with more than 40 unique associated SNPs in the NHGRI GWAS Catalog, we computed a tight linkage disequilibrium (LD) block (using appropriate 1000 Genomes populations) for each trait-associated SNP and then characterized these blocks' evolutionary histories, sequence motif patterns, and overlapping functional genomics data (from ENCODE and Roadmap Epigenomics). We then trained a multiple kernel support vector machine to distinguish regions associated with each trait from those associated with other traits and regions with no known associations. The algorithm obtains strong performance at distinguishing regions for many traits from regions with no associations (ROC AUCs above 0.8), and it performs significantly better than random for all but five traits. Our approach is also able to distinguish among regions associated with different traits; however, there is considerable variation in performance: a maximum ROC AUC of 0.78 for celiac disease and a minimum of 0.50 for adverse metabolic events. Using the trained classifiers for each trait, we predicted thousands of loci across the human genome that resemble known trait-associated regions (at a 5% FDR). The predicted regions for nearly all traits are enriched for nearby annotations and gene expression relevant to the trait. In addition, traits known to share common pathways overlap in their predicted regions more than less related traits. These genome-wide predictions can extend our knowledge of the genetic basis for complex traits. To illustrate this potential, we show how they can aid in prioritization of SNPs with low p-values that do not reach the genome-wide significance level. We also demonstrate the use of our predictions in developing hypotheses about the mechanisms behind these associations by identifying several hundred known enhancers and eQTLs that fall in novel predicted trait-associated regions.

973T

A search for genetic risk factors of age-related macular degeneration informed by the genetics of Alzheimer disease. *M.W. Logue[1,5], M. Schu[1], B.N. Vardarajan[1], J. Farrell[1], K.L. Lunetta[5], G. Jun[1], C.T. Baldwin[1], M. DeAngelis[6], L.A. Farrer[1,2,3,4,5].* 1) Biomedical Genetics, Boston Univ Sch Med, Boston, MA; 2) Dept of Neurology, Boston Univ Sch Med, Boston, MA; 3) Opthalmology, Bost Univ Sch Med, Boston, MA; 4) Epidemiology, Boston Univ Sch of Public Health. Boston, MA; 5) Biostatistics, Boston Univ Sch of Public Health, Boston MA; 6) Moran Eye Center, University of Utah, Salt Lake City, UT.

Background: Several lines of evidence point to overlapping molecular mechanisms between late-onset Alzheimer disease (AD) and age-related macular degeneration (AMD), including the molecular composition of the extra-cellular deposits which are the hallmarks of each, shared cardiovascular risk factors, and significant risk associated with *APOE* in both disorders. **Methods:** We evaluated summary statistics from large-scale genome-wide meta-analyses of AD (Naj et al. 2011) and AMD (Fritsche et al. 2013), treating the AD GWAS as a discovery dataset and the AMD dataset as a replication dataset to focus our search for genetic risk factors of AMD at the SNP, gene (significance based on peak SNP adjusted using the Li and Ji method), and pathway level. Results: First, we examined all SNPs associated with AD at a 10% false discovery rate (FDR) cutoff~excluding regions previously associated with AMD. None of the 226 SNPs tested for association with AMD were significant at the .05/226=.00022 level. Next, we examined all putative AD genes significant at a 10% FDR level (excluding the *APOE* region) for association with AMD. Two of 24 genes examined were significant (p<.05/24=.0021). The gene *PILRA* on chromosome 7q22 (FDR_{AD}=.057) was associated with AMD (p_{gene}=7.0E-6; peak SNP rs7792525, MAF=19%, OR=1.14, p=2.3E-6). Second, the known AD risk gene *ABCA7* (FDR_{AD}=.006) was also associated with AMD (p_{gene}=.0012, peak SNP rs3752228, MAF=5% OR=1.22, p=1.2E-4). The AMD-associated *ABCA7* SNP is not in LD with the previously identified AD risk variants rs3764650 and rs3752246 (r2<.01). Finally, we examined association of AMD with genes in the clathrin-mediated endocytosis signaling pathway (191 genes) and the atherosclerosis signaling pathway (132 genes), which emerged from an INGENUITY canonical pathway analysis of AD risk genes. Multiple genes in these pathways were significantly associated with AMD including several previously established AMD loci. Novel associations for AMD were observed with *HGS* (p_{gene}=7.5E-5, peak SNP rs8070488, MAF=23%, OR=.91, p=1.8E-5), which encodes a protein regulating endosomal sorting, and gene *TNF* (p_{gene}=1.2E-5, peak SNP rs2071590, MAF=34%, OR=.89, p=9.8E-7), a proinflammatory cytokine. LD in the regions around *TNF* and *PILRA* makes it impossible to localize the association to those genes specifically. Conclusions: We identified several novel modest-effect AMD risk variants in genes with a direct or peripheral role in AD pathogenesis.

974F

X-linked genes associated with cleft lip with or without palate in a family-based genome wide association study in Patagonia. *R.F. Fonseca[1], F.M. de Carvalho[1], F. Poletta[2,3], D. Montaner[4], J.C. Mereb[5], M.A.M. Moreira[6], H.N. Seuanez[6], A.R. Vieira[7], E.H. Castilla[8], I.M. Orioli[1].* 1) ECLAMC (Latin American Collaborative Study of Congenital Malformations) INAGEMP (National Institute of Populational Medical Genetics) at Department of Genetics, Institute of Biology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil;; 2) ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP-CNPq (National Institute of Population Medical Genetics), Rio de Janeiro, Brazil;; 3) ECLAMC at Center for Medical Education and Clinical Research (CEMIC-CONICET), Buenos Aires, Argentina;; 4) Bioinformatic Department, Research Center Principe Felipe, Valencia, Spain;; 5) ECLAMC and INAGEMP at Hospital Zonal El Bolsón, El Bolsón, Argentina;; 6) Divisão de Genética, Instituto Nacional de Câncer, Rio de Janeiro, Brazil;; 7) Departments of Oral Biology and Pediatric Dentistry and Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh,; 8) ECLAMC and INAGEMP at Congenital Malformations Epidemiology Laboratory, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil.

Cleft lip with or without palate (CL±P) are one of the most common birth defects, worldwide. Its birth prevalence ranges from 1/500 to 1/2000, depending on geographical origin, and ethnic background. The etiology of CL±P is complex and very heterogeneous with multiple genetic and environmental factors playing a role. The identification of genes that participate in the complex etiology of CL±P is an important step considering the morbidity that accompanies this relatively common birth defect. In this scenario, genome wide association studies are important tools in the search for candidates genes associated with complex diseases. Among the candidate genes described to be associated with CL±P some of them are on the X chromosome. Recently two genes (OFD and DMD) were associated with isolated CL±P in Asians and European populations. In this study we selected a CL±P high prevalence population from Patagonia identified by the ECLAMC and 99 individuals were until now ascertained and studied. SNPs from chromosome X were genotyping using the Affymetrix Plataform (Genome-Wide 6.0) and tested (TDT) for association with isolated CL±P in 24 trios, 2 duos, and 23 sibs using PLINK. We found statistical evidence of association with CL±P in 57 SNPs localized in six genes KAL 1 (06), DMD (04), SYN1 (02), CFP (02), TBL1X (01), TLR8 (01), uncharacterized region LOC100873065 (01) and intergenic regions (40) ($P \leq 0.008151$). To assess the role of these 57 SNPs on CL±P phenotype, we performed the haplotypes analysis with Haploview Software. Fourteen haplotypes under $P \leq 0.0068$ were identified whereas the most significant involves the following SNPS rs16988033-rs7055843-rs16997889, all of them in intergenic regions on locus Xp21.3 ($P= 0.0009$). The results for DMD and KAL1 genes are consistent with recent data about the involvement of these genes in the etiology of oral clefts while the other genes were described for the first time. In order to confirm the current association and also to improve our statistical power, it's advisable increase the number of trios affected by CL±P in further studies.

975W

Linkage analysis: genomic regions contributing to the expression of type 1 diabetes microvascular complications. *E.M. Lipner[1], Y. Tomer[2], J.A. Noble[3], M.C. Monti[4], J.T. Lonsdale[5], B. Corso[4], D.A. Greenberg[6].* 1) Department of Epidemiology, Columbia University, New York, NY; 2) Department of Medicine, Mount Sinai Medical Center, New York, New York; 3) Children's Hospital Oakland Research Institute, Oakland, CA; 4) Department of Public Health, Neurosciences, Experimental and Forensic Medicine, University of Pavia, Pavia, Italy; 5) National Disease Research Interchange, Philadelphia, PA; 6) Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Department of Pediatrics, Wexner Medical Center, Ohio State University, Columbus, Ohio.

We conducted a linkage analysis to identify susceptibility loci for microvascular complications of type 1 diabetes (T1D). We used 415 families from the Human Biological Data Interchange (HBDI) who were diagnosed with T1D before age 30. After finding significant HLA and non-HLA linkage peaks on chromosome 6, we restricted further analyses to that chromosome. Using 402 SNP markers, our analysis used the phenotypes of 1) any microvascular complication, 2) only retinopathy, 3) nephropathy alone, 4) neuropathy alone. When using 'any complication' as the phenotype, we identified two linkage peaks: one located at the HLA region (HLOD=2.90) and another, novel locus telomeric to HLA (HLOD=3.13). These same two peaks were also evident when retinopathy was the phenotype (HLODs of HLA region=2.69, telomeric locus=3.30). We did not find evidence for linkage for nephropathy alone or neuropathy alone We stratified on families whose probands were positive for DRB1*03:01 and DRB1*04:01 because previously published evidence suggest that DRB1 locus alleles affect complications expression. When using 'any complication' as the phenotype and including only DRB1*03:01-positive families, the HLA peak decreased (HLOD=1.82) from the unstratified analysis (HLOD=2.90), and a peak centromeric to HLA appeared (HLOD=1.27). When we stratified on DRB1*04:01-positive families, not only was there was no evidence of heterogeneity (LOD=HLOD) but the linkage evidence for HLA (HLOD=3.83) and the telomeric locus (HLOD=3.69) both went up, despite the drop in sample size with stratification. When using retinopathy as the phenotype, we observed the same increase in linkage peaks (HLOD at HLA region=3.62, telomeric locus=3.76). These observations suggest that DRB1*04:01 interacts with the telomeric locus to produce complications' susceptibility. Simultaneously, the drop in linkage evidence for the DRB1*03:01 confirms a protective effect seen in our previously reported association analysis (Lipner et al, 2013). Based on large differences in the linkage evidence, we argue that the DRB1*03:01-positive and DRB1*04:01-positive groups are genetically distinct, a finding in accordance with the observation that DRB1*03:01 is protective for retinopathy. The evidence also suggests there is an interaction between specific HLA alleles and a locus telomeric to HLA that influences the expression of some complications.

976T

Gene-environment interactions between obesity gene variants and body size in multiple sclerosis. *M.A. Gianfrancesco[1], B. Acuna[2], L. Shen[2], F.B.S. Briggs[1], H. Quach[1], A. Bernstein[3], A.K. Hedstrom[4], I. Kockum[5], L. Alfredsson[4,6], T. Olsson[5], C. Schaefer[2,7], L.F. Barcellos[1,2].* 1) School of Public Health, Dept of Epidemiology, UC Berkeley, Berkeley, CA; 2) Kaiser Permanente Division of Research, Oakland, CA; 3) Palm Drive Hospital, Sebastopol, CA; 4) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 5) Neuroimmunology Unit, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; 6) Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden; 7) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS) is a demyelinating autoimmune disease affecting over 2.5 million people worldwide. Although studies have confirmed a strong genetic component for susceptibility, evidence for environmental risk factors, such as childhood and adolescent obesity, has also been reported. We investigated the relationship between early obesity and MS while controlling for effects of several established genetic and environmental risk factors. A gene-environment (GxE) investigation also assessed whether variation within established obesity genes, including APOA2, APOA5, FTO, MC4R, SCG3, TUB, and PCSK1, modifies the MS risk conferred by body size and body mass index (BMI) during various periods of life. Participants included female members of Kaiser Permanente Medical Care Plan, Northern California Region (985 MS cases, 10,000 controls). Common and rare variant data for each candidate gene were obtained for each individual through genome-wide association profiling and imputation. Being overweight at age 10 and 20 were associated with MS onset ($p<0.02$) after adjusting for age, smoking status, education, history of infectious mononucleosis, HLA-DRB1*15:01 and a weighted genetic risk score. Pooled analyses with a secondary dataset from a Swedish cohort indicated strong evidence for dose-effect of BMI at age 20 and MS (N= 1,364 cases and 1,639 controls) ($p=9.6\times10$-4). Results, thus far, show that of 1,212 SNPs across the fat mass and obesity-associated (FTO) gene (chr16q12.2), 7 SNPs were associated with MS risk ($p<0.05$: rs741300, rs9931900, rs2540772, rs8051873, rs7185735, rs2192869, rs708278). Inclusion of carrier status for the common FTO rs9939609 variant in the model did not modify the relationship between body size/BMI and MS at age 10. There was no evidence for interaction between FTO rs9939609 and body size/BMI at various ages. These findings convincingly show that childhood and adolescence obesity are associated with an increased risk of MS after controlling for known genetic and environmental risk factors, as well as obesity-related variants. Obesity is modifiable; therefore, these findings have large clinical and public health implications. Furthermore, MS risk conferred by obesity may be independent of predisposing genetic factors for obesity, suggesting that alternative mechanisms may mediate disease onset.

977F

Genetic linkage analysis using large consanguineous pedigrees from South India suggests new loci for ocular quantitative traits. *B. Fan[1], P. Ferdina Marie Sharmila[2], N. Soumittra[2], S. Sripriya[2], J. Madhavan[2], D.S. Friedman[3], L. Vijaya[4], J.L. Haines[5], R. George[4], J.L. Wiggs[1].* 1) Department of Ophthalmology, Harvard Med Sch, Massachusetts Eye & Ear Infirmary, Boston, MA; 2) Vision Research Foundation, Sankara Nethralaya, Chennai, India; 3) Johns Hopkins Medical School, Wilmer Eye Institute, Baltimore, MD; 4) Medical Research Foundation, Sankara Nethralaya, Chennai, India; 5) Center for Human Genetic Research, Vanderbilt University School of Medicine, Nashville, TN.

Many ocular quantitative traits are highly heritable. Mapping genes contributing to these quantitative traits may effectively identify genetic risk factors predisposing to related complex diseases, such as glaucoma. Compared to outbred pedigrees, consanguineous pedigrees are expected to have more power to identify quantitative trait alleles with additive or subtractive effects. Additionally, inbreeding due to social customs or geographic constraints may reduce genetic heterogeneity. In this study we performed genome-wide linkage analyses for selected ocular quantitative traits using 240 members of 16 consanguineous pedigrees from South India (Chennai). All study participants underwent a comprehensive ophthalmic examination for ocular quantitative traits including axial length (AXL), central corneal thickness (CCT), cup depth (CD) and vertical cup-disc ratio (VCDR). Genotyping was performed using the Illumina HumanOmni2.5-8 platform. Multipoint linkage analysis of 1,223,314 SNPs was performed using the variance components analysis in MERLIN (v1.1.2). All traits were adjusted for age and sex, and inverse-normal transformation of ranks was applied before analysis. Significant novel linkage signals were identified for AXL on 6q24.1-q24.2 (LOD_{max} = 3.45), CD on 12q24.31 (LOD_{max} = 3.92) and VCDR on 14q32.12 (LOD_{max} = 3.62). In addition, suggestive evidence of linkage was identified for CCT on 11q22.3-q23.2 (LOD_{max} = 2.37). Further evaluation of interesting candidate genes located within these linkage intervals is ongoing. This study shows that a relatively small number of consanguineous pedigrees can provide sufficient power to identify quantitative trait loci. Grant support: NEI Grants R21EY018149; P30EY014104; Research to Prevent Blindness; Massachusetts Lions Eye Research Fund.

978W

Genome-Wide Family-Based Linkage Analysis of Coding Variants and Cardiometabolic Risk. *J.N. Hellwege[1,2,3], N.D. Palmer[2,3,4], L.M. Raffield[1,2,3], M.C.Y. Ng[2,3], G.A. Hawkins[2], J. Long[5], C. Lorenzo[6], J.M. Norris[7], J.I. Rotter[8], C.D. Langefeld[9], L.E. Wagenknecht[9], D.W. Bowden[2,3,4].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN; 6) Department of Medicine, University of Texas Health Science Center, San Antonio, TX; 7) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO; 8) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California; 9) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC.

In contrast to Mendelian disorders, family-based linkage analysis of complex traits has had limited success in identifying trait-influencing or disease loci. Recently, however, low frequency coding variants have been shown to be the basis for some biomedical associations. We tested the hypothesis that low frequency coding variants are the basis for linkage peaks identified in the analysis of complex traits from 42 African American (N=586) and 88 Hispanic (N=1270) families in the Insulin Resistance Atherosclerosis Family Study (IRASFS). DNAs were genotyped using the Illumina HumanExome Beadchip. A total of 92,157 variants in African Americans (~34%) and 81,559 (~31%) in Hispanics were polymorphic and passed quality control metrics, including Mendelian error checking. Family-based two-point linkage and conventional association analysis were performed using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) program package. Two-point linkage analyses in families in each ethnic group were performed for 38 cardiometabolic phenotypes. Analyses yielded greater than 3 million LOD scores in each ethnic group. In African Americans, 1127 LOD scores greater than 2 and 77 greater than 3 were observed. The highest LOD score was 4.91 with the *APOE* SNP rs7412 (MAF=0.13) and plasma levels of apolipoprotein B (ApoB). In association analysis, rs7412, one of the SNP determinants of the ApoE ε2/3/4 isoforms, was also powerfully associated with ApoB (p-value= 4×10^{-19}). This *APOE* variant accounts for 16.2% of the variance in ApoB in African Americans. In Hispanic families, 1218 LOD scores were greater than 2 and 104 were greater than 3. Excluding a previously identified linkage peak on chromosome 3 for adiponectin, the strongest evidence of linkage (LOD=4.29) was with rs5882 (MAF=0.46) in *CETP* with HDL levels. Variants in and near *CETP* were strongly associated with HDL levels (p-values= $0.00049\text{-}4.6\times10^{-12}$) accounting for up to 4.5% of the variance in HDL. Both of these loci have previously been shown to have effects on the biomedical traits evaluated in this study. Thus, variants that contribute appreciably to the variance in a trait show evidence of sample-wide linkage. Analysis of individual families provides a powerful approach for identifying additional coding variants linked to complex traits in IRASFS.

979T

Genome-wide screen for self-reported physical disability loci in the oldest-old Amish. *J.E. Hicks[1], J.R. Gilbert[1], L. Caywood[1], L. Reinhart-Mercer[1], D. Fuzzell[2], R. Laux[2], M.A. Pericak-Vance[1], J.L. Haines[2], W.K. Scott[1].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Center for Human Genetics Research, Vanderbilt University, Nashville TN.

As the population ages, age-related disability will increasingly become an important public health concern. However, many objective measures of aging differ from self-reported measures of impairment. To identify patterns of perceived disability in the oldest old, a latent class analysis was performed on four scales assessing perceived disability in 263 cognitively intact participants aged 80 or above. Participants were enrolled through a population-based door-to-door study of community-dwelling Amish individuals aged 65 and older from Indiana and Ohio. Individuals were placed into a 13-generation pedigree using the Anabaptist Genealogy Database. These individuals were assessed with the Activities of Daily Living scale (basic self-care tasks such as dressing and bathing), Instrumental Activities of Daily Living scale (higher-level tasks such as cooking and shopping), the Nagi scale for musculoskeletal function, and the Rosow-Breslau scale for lower-extremity function. Data analysis was performed with the poLCA package in R with age and sex included as covariates. The optimal model identified two classes. The first (n=209) was largely unimpaired on all disability questions, while the second class (n=54) reported some degree of physical disability. Random forest analysis of disability found that the strongest predictor of class was ability to perform heavy housework. A genome-wide screen was performed with Affymetrix Genome-Wide Human SNP Array 6.0. PLINK was used to select a set of 4,966 SNPs in linkage equilibrium (pairwise r2 < 0.16) for linkage analysis. PedCut created 11 subpedigrees suitable for linkage analysis, each containing at least 2 of the 54 physically impaired individuals. Multipoint nonparametric linkage analysis (LOD*) was performed in MERLIN and identified no regions of significant linkage to disability but did find two regions providing suggestive evidence of linkage on chromosomes 1 (LOD*=2.43, 18,922,283 Mb) and 3 (LOD*=2.28, 30,502,223 Mb). While tests of association in the 2-LOD-down support intervals (SI) surrounding these peaks did not identify markers significantly associated with disability after correction for multiple testing ($p<10^{-6}$), the strongest associations were at rs12071805 (p=0.000052) and rs1552391 (p=0.000029). The chromosome 3 SI contains rs11177, which has been previously associated with osteoarthritis (OA). These results suggest that multiple loci may influence disability in these older adults, possibly through OA.

980F

Linkage and association analysis of von Willebrand factor propeptide levels provides mechanistic insight into the genetic control of plasma von Willebrand factor levels. *K.C. Desch[1], A.B. Ozel[2], D. Siemieniak[4], J.Z. Li[2], D. Ginsburg[1,2,3,4].* 1) Pediatrics and Communicable Disease, University of Michigan, Ann Arbor, MI; 2) Human Genetics, University of Michigan, Ann Arbor, MI; 3) Internal Medicine, University of Michigan, Ann Arbor, MI; 4) Howard Hughes Medical Institute, Ann Arbor, MI.

von Willebrand Factor (VWF) is a plasma glycoprotein playing a central role in hemostasis. Common alleles at the ABO locus account for 24.5% of VWF variance and operate through altered VWF clearance rates. The VWF propeptide (VWFpp) is secreted into circulation with mature VWF where unlike VWF, it is rapidly cleared from circulation. Therefore plasma VWFpp levels are largely determined by the rate of synthesis/secretion of VWF. To gain mechanistic insight into genetic determinants of VWF levels, we performed linkage and genome-wide association studies (GWAS) of VWFpp in two healthy cohorts (combined n=3462) and compared the results to our previous analyses of VWF in these cohorts.(Desch et al., 2013) VWFpp levels were determined with monoclonal antibodies in an AlphaLISA assay. 800K SNPs, that were available after extensive QC in our previous study, were used in association studies (PLINK, EMMAX, METAL) and linkage analysis (MERLIN). Heritability (h2) of VWFpp was 74% compared to 65% for VWF. Our previous meta-analysis GWAS of VWF revealed signals at ABO (P=7.9E-139) and VWF (P=5.5E-16). Here, preliminary GWAS for VWFpp in one cohort revealed a single significant SNP rs1800378 (p-value= 8.94E-13) at 12p13, overlying the VWF gene itself. This SNP encodes a His484Arg variant in the VWF propeptide domain. However, interference of His484Arg with antigen detection was suggested after a repeat GWAS in this cohort using VWFpp levels determined with a different set of monoclonal antibodies was performed and did not reveal any significant SNPs at the VWF locus. Meta-analysis GWAS for VWFpp identified a signal in ABO (rs8176746, P=1.01E-12) suggesting that the plasma clearance of VWFpp, like VWF, is weakly altered by ABO glycosylation patterns. Previous linkage analysis of VWF identified a novel locus at 2p12-2q13 (19.2% variance explained) that was not detected by GWAS, strongly suggesting a high level of allelic heterogeneity. Here, linkage analysis for VWFpp did not reveal significant signals suggesting the 2p12-2q13 locus harbors variants affecting VWF clearance rates. In summary, we identify a common non-synonymous SNP in VWFpp domain resulting in a 'false positive' GWAS signal that may have important implications in the analysis of GWAS data for other antibody based quantitative traits. Futhermore, by examining VWFpp, a marker of VWF synthesis/secretion rates, we provide mechanistic insight into a previously identified linkage signal for VWF.

981W

Genetic variants on 17q23.2 and 10q11.21 are associated with variation in telomere length: The Long Life Family Study. *J.H. Lee[1,3], R. Cheng[1], L.S. Honig[1,2], M. Feitosa[4], C. Kammerer[6], M.S. Kang[1], N. Schupf[1,3], R. Lin[4], J.L. Sanders[6], H. Bae[7], T. Druley[5], T. Perls[7], K. Christensen[8], M. Province[4], R. Mayeux[1,2,3].* 1) Sergievsky Ctr/Taub Inst, Columbia Univ, New York, NY; 2) Depts of Neurology and Psychiatry, Columbia University, New York, NY; 3) Dept of Epidemiology, Columbia University, New York, NY; 4) Div of Statistical Genomics, Depr of Genetics, Washington University, St. Louis, MO; 5) Dept of Pediatrics & Genetics, Washington University, St. Louis, MO; 6) Dept of Epidemiology & Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, PA; 7) Div of Geriatrics, Dept of Medicine, Boston University Medical Center, Boston, MA; 8) The Danish Aging Research Center, Epidemiology, University of Southern Denmark, Odense, Denmark.

Telomere length is hypothesized to measure cellular aging in humans, and short telomere length is associated with increased risks of late onset diseases, including cardiovascular disease, dementia, cancer, and other age-related disorders. Many studies have shown that telomere length is a heritable trait, and several candidate genes have been identified, including *TERT, TERC, OBF1*, and *CTC1*. Most studies have focused on genetic causes of chronic diseases such as heart disease and diabetes in relation to telomere length. This investigation involved a genome wide family-based association analyses that identified genetic variants that contribute to variation in telomere length among families with exceptional longevity. From the genome wide association analysis in 4,289 LLFS participants, we identified a novel intergenic SNP rs7680468 near *PAPSS1* on 4q24 (p=4.7E-8). From our linkage analysis, we identified two additional loci with HLOD scores>3, including one at 17q23.2 (HLOD=4.77) and another at 10q11.21 (HLOD= 4.36). Subsequent gene-wise association analysis identified multiple candidate genes from these two regions, including *DCAF7, POLG2, CEP95*, and *SMURF2* at 17q23.2; and *RASGEF1A, HNRNPF, ANF487, CSTF2T,* and *PRKG1* at 10q11.21. Among these genes, we identified one contiguous haplotype in *CEP95* and *SMURF2* that was significantly associated with telomere length. In addition, previously reported genes - *TERC, ARPM1, MYNN, OBFC1*, and *ZNF729* - showed weak to modest association with telomere length variation. Here we report novel genetic variants and confirm previously reported genes that are associated with variation in telomere length. These may reveal additional insight into cellular aging.

982T

Combined genome-wide linkage and association studies of centenarians identifies several new candidate genes for longevity. *P. Sebastiani[1], H. Bae[1], T.T. Perls[2].* 1) Dept Biostatistics, Boston Univ Sch Public Health, Boston, MA; 2) Dept Medicine, Boston University School of Medicine, and Boston Medical Center, Boston MA.

The genetic basis of extreme longevity increases as more extreme lifespans are examined. However, genome-wide association studies (GWAS) of longevity have failed to discover highly significant variants. Several studies have confirmed a role of APOE in lifespan and rs2075650 in TOMM40/APOE is the only single nucleotide polymorphisms(SNP) that reached genome-wide significance in several GWAS of nonagenarians and centenarians. In a GWAS of 801 centenarians (median age at death 104 years) versus population controls, we showed that a set of 281 SNPs that did not reach genome-wide significance could accurately distinguish centenarians from controls, with increasing sensitivity in older and older people. SNPs that failed to reach genome-wide significance but showed consistent effects in different studies have also been published. These findings suggest that the genetic makeup of exceptional longevity is determined by a combination of common and rare variants with modest genetic effects, and that the relatively small sample size of centenarian studies is a serious limitation to their discovery power. We hypothesized that linkage analysis could be used to prioritize variants that reached statistical significance in association studies but failed to reach genome-wide significance. We conducted a non parametric linkage analysis of affected sib-pairs enrolled in 172 families of the New England Centenarian Study (NECS) enrolled between 1995 and 2010. Affected pairs were defined as two siblings with age at death exceeding a fixed percentile survival from sex and birth year matched life table from the US Social Security Administration. Percentile survivals π between 0.25 and 0.001 were used to derive nested sets of sib-pairs that were analyzed in Merlin. Regions in chromosomes 2, 3, 4, 7, 10 and 19 had linkage peaks with LOD score >2.4 for at least one threshold π, or LOD score >2 for more than one threshold π and 1.5LOD support intervals were annotated by significant results from two GWAS of longevity in NECS centenarians. The region under the linkage peak in chromosome 10 includes ADARB2 that we found previously associated with extreme longevity. The region in chromosome 2 includes SNPs in MYTL1 and its regulatory region that is enriched of SNPs associated with insulin resistance, insulin variability and diabetes complications. Three regions in chromosome 19 point to additional genes that are associated with extreme longevity independent of APOE.

983F

Analyzing patterns of IBD sharing in Oceanic Palau to identify genomic regions harboring risk for schizophrenia. *C.A. Bodea[1], F. Middleton[2], L. Klei[3], S.V. Faraone[2], S. Vinogradov[4], J. Tiobech[5], V. Yano[5], S. Kuartei[5], K. Roeder[1], N. Melhem[3], B. Devlin[3], M. Myles-Worsley[2], W. Byerley[4].* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 2) Department of Psychiatry, SUNY Upstate Medical University, Syracuse NY; 3) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 4) Department of Psychiatry, University of California San Francisco, Pittsburgh, PA; 5) Palauan Ministry of Health, Republic of Palau, Pittsburgh, PA.

The genetics of schizophrenia is complex and, despite recent successes, it remains challenging to identify individual risk loci. We have been studying the genetics of schizophrenia in Oceanic Palau, an island nation some 500 miles east of the Phillipines. One of the features we have been analyzing is the co-segregation pattern of genetic variation and schizophrenia status to identify regions of the genome shared identical-by-descent (IBD) among related, affected individuals. These IBD segments are shared as haplotypes and we search for inherited haplotypes that increase the risk of developing schizophrenia. A challenge in this small and until recent-generations isolated Oceanic population is that pedigrees are large and even those subjects who do not fall in the same close-knit pedigree are often found to be genetically related to some degree. Motivated by this observation we build a statistical test that can be applied to large pedigrees whose structure need not be known, one that evaluates the association between the extent of IBD sharing within specific genomic regions and the presence of schizophrenia. We first estimate a matrix of pairwise relatedness between subjects by measuring IBD sharing over the entire genome using Germline. This results in a matrix with genetic estimates of kinship amongst all pairs of subjects in the study, but the matrix is noisy. Applying a recently developed algorithm to the kinship matrix -- treelet covariance smoothing -- we reduce the noise by using the information inherent in clusters of individuals with estimated positive kinship coefficients. To form a test, we apply a kernel-based approach that compares local IBD sharing between cases and controls, while controlling for estimated relatedness. Within a given region, every pair of subjects is characterized by the segment-length inherited from a common ancestor (IBD). The statistic seeks to identify regions of the genome where the amount of IBD sharing amongst cases and/or amongst controls is greater than the sharing between cases and controls, because these regions have the potential to harbor mutations that are associated with schizophrenia. Our approach identifies regions of the genome as potentially harboring risk variants and some of those loci harbor known schizophrenia-associated genes, suggesting the results are likely to be meaningful. We continue to explore these data to identify variants affecting risk for schizophrenia.

984W

Cumulative genetic load for known multiple sclerosis risk variants in Sardinia. *A. Hadjixenofontos[1], L. Foco[2], P-A. Gourraud[3], A. Tikka[5], P. Bitti[6], R. Pastorino[2], L. Bernardinelli[2,4], J.L. McCauley[1].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Public Health, Neurosciences, Experimental and Forensic Medicine, Medical Statistics and Epidemiology Unit, University of Pavia27100 Pavia, Italy; 3) Department of Neurology, School of Medicine, University of California at San Francisco, San Francisco, CA, USA; 4) Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK; 5) Divisione di Neurologia, Ospedale S. Francesco, Nuoro, Italy; 6) Immunohaematology and Blood Transfusion Department, Ospedale S. Francesco, Nuoro, Italy.

Multiple Sclerosis (MS) is an autoimmune, demyelinating disease for which the prevalence rates follow a latitude gradient with the lowest prevalence in countries closest to the equator. Sardinia, in Insular Italy, is an exception to this trend with unenviable prevalence rates that range between 150 and 200 per 100,000 individuals. The geographical isolation of Sardinia presents a valuable opportunity to study a relatively homogeneous population, for which the underlying genetic load is not known. Recent collaborative investigations by the International Multiple Sclerosis Genetics Consortium (IMSGC) have increased the list of MS risk variants to 113 tag SNPs. We use these variants in the calculation of a MS genetic risk score with the goal of assessing the extent to which NE derived variants allow differentiation between Sardinian MS cases and controls. The genotyping data on our dataset of 16 Sardinian families and 94 unrelated Sardinian controls originated from a custom Illumina genotyping chip dubbed the 'Immunochip'. Using Generalized Estimating Equations (GEE) to correct for correlations between relatives, we built a logistic regression model with disease status as the outcome, and the MS risk score, and sex as the predictors. An increase in the NE derived MS risk score was associated with MS case status in our preliminary analysis of 70 Sardinian cases and 244 Sardinian unaffected family members and population controls (OR=1.68, 95% CI=1.33-2.11, P<0.001). This is in contrast to our previous results with a version of the MS risk score that included fewer tag SNPs and which was not associated with affection status. No differences in the MS risk score were detected between MS patients from Sardinian vs. Caucasian population (OR=1.11, 95% CI=0.81-1.58, P= 0.55). In light of our latest results a more extensive analysis is currently underway. These will seek to capitalize on the relatedness in the sample by examining whether variants in the known regions are shared identical by descent (IBD) in cases from each family, and additionally will include information on rare variants from exome sequencing data which may be present in the known regions.

985T

Interaction between Adiponectin and Adiponectin Receptor 1 is Associated with Age-related Hearing Impairment. *C. Wu[1,2], C. Tsai[3], J. Hwang[4], Y. Lu[1,5], Y. Lin[1,6], P. Chen[7], W. Yang[2,7], W. Liao[8], Y. Lee[3,9], T. Liu[1], C. Hsu[1].* 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 4) Department of Otolaryngology, Buddhist Dalin Tzu-Chi General Hospital, Chiayi, Taiwan; 5) Institute of Biomedical Engineering, National Taiwan University, Taipei, Taiwan; 6) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 7) Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 8) Health Management Center, National Taiwan University Hospital, Taipei, Taiwan; 9) Research Center for Genes, Environment and Human Health, College of Public Health, National Taiwan University, Taipei, Taiwan.

Age-related hearing impairment (ARHI) is a complex disease caused by an interaction between environmental and genetic factors. Recently, several studies confirmed obesity as an independent risk factor for ARHI. In our previous study investigating the underlying mechanisms, we demonstrated that plasma adiponectin might protect peripheral hearing function (Clin Endocrinol. 2011). It has been revealed that polymorphisms of the adiponectin gene, ADIPOQ, might affect plama adiponectin levels; and polymorphisms of both ADIPOQ and its type 1 receptor gene, ADIPOR1, have been related to obesity-related morbidities. Hence, we postulated that genotypes of ADIPOQ and ADIPOR1 might be associated with the development of ARHI. A total of 1682 volunteers (Han Chinese, aged 40 to 80 y) were included in the clinical analyses, and their audiological phenotypes were determined according to the Z scores converted from their original frequency-specific hearing thresholds. By using the database of Chinese Han haplotypes in International HapMap Project and NIEHS, followed by analyses with the Haploview software, 9 tagSNPs and 4 tagSNPs in ADIPOQ and ADIPOR1, respectively, were selected for genotyping. The genotypes were then correlated to the audiological phenotypes under the assumption of various inheritance models. Significant association was identified between certain ADIPOQ tagSNPs and Z scores under dominant, co-dominant, or additive models; whereas no association was identified between the ADIPOR1 tagSNPs and Z scores. Of note, the association between specific ADIPOQ tagSNPs and Z scores appeared to exist in subjects with specific ADIPOR1 genotypes only, indicating a gene-gene interaction between ADIPOQ and ADIPOR1. These findings were then validated on haplotype analyses. Further measurement of plasma adiponectin level in 736 subjects revealed ADIPOQ genotypes might exert their effects on hearing levels through modulating the plasma adiponectin levels. In conclusion, the ADIPOQ genotypes were associated with ARHI, and might exert effects on hearing levels through modulating the plasma adiponectin levels. In contrast, although the ADIPOR1 genotypes were not directly associated with ARHI, the association between ADIPOQ and hearing thresholds were influenced by the ADIPOR1 genotypes. In other words, the development of ARHI might result from an interaction between adiponectin and type 1 adiponectin receptor.

986F

On the use of cis-eQTL analyses to identify novel associations between *OPRM1* polymorphisms and heroin abuse.
D.B. Hancock[1], J.L. Levy[1], N.L. Saccone[2], N.C. Gaddis[1], L.J. Bierut[2], G.P. Page[3], E.O. Johnson[1]. 1) RTI International, Research Triangle Park, NC; 2) Washington University in St. Louis, St. Louis, MO; 3) RTI International, Atlanta, GA.

The *OPRM1* (opioid receptor, mu 1) gene has long been suspected to influence risk of heroin abuse. The 118A>G polymorphism, in particular, has been widely studied for its association with heroin and other addictions and its functional consequences, but the evidence for association is inconsistent. To identify other polymorphisms that may regulate *OPRM1* and in turn influence risk of heroin abuse, we evaluated single nucleotide polymorphism (SNP) associations with *OPRM1* expression to nominate cis-expression quantitative trait loci (cis-eQTL) in the human brain and subsequently tested the nominated SNPs for association with the heroin abuse phenotype. For the cis-eQTL analyses, we utilized SNP genotypes (Illumina Human1M-Duo or HumanHap650Y) and gene expression data from prefrontal cortex samples of normal human subjects in the publically available BrainCloud. Our cis-eQTL analyses used 112 African American and 109 Caucasian subjects from infancy through adulthood (ranging in age from <1 to 78 years old). In each ancestral group, 103 SNPs located within 100kb of *OPRM1* were tested for association with *OPRM1* expression level using linear models that accounted for sex and age within developmental stage (childhood vs. adulthood). Of the 103 SNPs, 14 were nominated as cis-eQTLs based on their association with *OPRM1* expression (P<0.05) in either ancestral group. The 14 SNPs were then tested for association with heroin abuse using cases from the Urban Health Study and public controls (total N=7,095 African Americans and 3,824 Caucasians), who were genotyped on Illumina arrays and imputed using IMPUTE2 with reference to the 1000 Genomes ALL haplotype panel. In each ancestral group, SNP associations with heroin abuse were tested using logistic models accounting for sex and 3 ancestry-specific eigenvectors. Convergence of the SNP association results revealed that 4 of the nominated cis-eQTLs (P ranging from 0.015 to 0.046) were also associated with heroin abuse (P ranging from 6.17×10^{-5} to 3.85×10^{-3} in a meta-analysis combining the ancestry-specific results). The 4 newly identified regulatory SNPs had minor allele frequencies between 16% and 38% and consistent directions of association with heroin abuse across the ancestral groups. Our findings offer new insights into the role that *OPRM1* may play in developing heroin abuse. Further studies are needed to confirm the observed SNP associations in other data sets and to elucidate their mechanisms.

987W

Evidence for a polygenic contribution to androgenetic alopecia.
S. Heilmann[1,2], F.F. Brockschmidt[1,2], A.M. Hillmer[3], S. Hanneken[4], S. Eigelshoven[5], K.U. Ludwig[1,2], C. Herold[6], E. Mangold[1], T. Becker[6,7], R. Kruse[8], M. Knapp[7], M.M. Nöthen[1,2]. 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life & Brain Centre, University of Bonn, Bonn, Germany; 3) Genome Technology and Biology, Genome Institute of Singapore, Singapore, Singapore; 4) Department of Dermatology, University of Düsseldorf, Düsseldorf, Germany; 5) Private Dermatology Practice, Solingen, Germany; 6) German Centre for Neurodegenerative Diseases (DZNE), Bonn, Germany; 7) Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany; 8) Private Dermatology Practice, Paderborn, Germany.

Male pattern baldness (androgenetic alopecia, AGA) is a highly heritable trait and the most common form of hair loss in humans. Genome-wide association studies (GWASs) including a meta-analysis and large-scale replication study have identified twelve genome-wide significant risk loci for AGA, which explain a major proportion of the genetic risk for AGA. However, the efficiency of contemporary GWASs for detecting disease associations is restricted by the burden of multiple testing due to the large numbers of single nucleotide polymorphisms (SNPs) considered in these analyses. Thus, only highly significant individual disease associations ($P<5\times10^{-8}$) can be detected. It is therefore likely, that additional genetic risk factors with smaller individual effect sizes still await identification. Their sensitivity for detecting individual disease associations is limited but GWAS datasets can be used to analyze whether there is a collective contribution (polygenic contribution) of common SNPs that show disease association above the *P*-value threshold for genome-wide significance ($P>5\times10^{-8}$). Irrespective of their small individual effect sizes, these loci might jointly explain a significant proportion of the overall genetic risk for AGA. Polygenic score analysis is a method for determining this collective contribution. This study used a German case-control sample for AGA, which comprised 581 severely affected patients and 617 controls, to determine the contribution of polygenic variance to genetic risk for AGA. The sample was divided evenly into discovery and test samples. An additive polygenic risk score was calculated from risk alleles with increasingly liberal *P*-values in the discovery dataset, which was then used to test for the enrichment of AGA risk score alleles in the independent test samples. The analysis provided significant evidence for the specific contribution of a polygenic component to AGA where the amount of variance explained was 1.4-4.5%. It is likely that the contribution of a polygenic component and the large number of genes involved reflect the complexity of the AGA associated biological pathways. Further studies are required to progress from evidence for a polygenic contribution to understanding the specific genetic factors that comprise this polygenic component.

988T

Integrated eQTL and Genetic Association Analysis in Smokers with COPD Identifies Disease-Associated eQTL Loci.

P.J. Castaldi[1], M.H. Cho[1,2], W. Qiu[1], B.R. Celli[2], J.H. Riley[3], S.M. Fox[3], D. Singh[4], R. Tal-Singer[5], B.A. Raby[1,2], V.J. Carey[1], E.K. Silverman[1,2], C.P. Hersh[1,2], ECLIPSE Investigators. 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 3) GlaxoSmithKline, Uxbridge, UK; 4) Medicines Evaluation Unit, University of Manchester, Manchester, UK; 5) GlaxoSmithKline, King of Prussia, PA, USA.

Background: Recent GWAS have identified novel loci associated with chronic obstructive pulmonary disease (COPD). Two of these loci have been shown to be associated with the expression level of nearby genes. We hypothesized that integrating eQTL and GWAS results from subjects with COPD would allow for the identification of novel COPD-associated loci, and that cis and trans eQTL analysis could characterize transcriptional networks of disease-relevant genes. **Methods:** Gene expression profiles from whole blood and induced sputum samples were generated for 121 subjects with COPD from the ECLIPSE study using the Affymetrix HG-U133 Plus 2.0 GeneChip. Genome-wide identification of cis eQTLs was performed using a cis window of 500kb. Multiple comparison adjustment was performed by controlling the false discovery rate at 10%. eQTL SNPs were tested for association with COPD susceptibility in a meta-analysis of 4 COPD case-control studies (NETT/NAS, ECLIPSE, GenKOLS, COPDGene). Targeted trans-eQTL analysis was performed for significant cis eQTL SNPs that were also associated with COPD. **Results:** Cis eQTL analysis in subjects with COPD identified 42,020 SNPs associated with expression of 7,223 unique genes in whole blood and 18,474 SNPs associated with 4,474 genes in sputum samples. Cis eQTL SNPs were significantly enriched in the top GWAS results (p<0.001 from both sources of expression data). When these SNPs were tested for association with COPD, SNPs associated with expression of 10 unique genes were also associated with COPD susceptibility (9 genes from blood, 2 genes from sputum with one gene present in both). All previously identified COPD GWAS loci harbored a significant cis eQTL. Trans eQTL analysis of these SNPs identified two SNPs associated with the expression of distal genes. rs1828591 was associated in cis with HHIP (q value = 0.09) and in trans with RTN4RL1 (q value = 0.03). rs8034191 was associated in cis with CHRNA3 and IREB2 (q value = 0.0001 and 0.002) and in trans with SLC35E3, GALE, and SLC22A7 (q value = 0.08 for all three genes). **Conclusions:** Using a gene-based eQTL approach in blood and sputum, we identified ten SNP-gene functional units associated with COPD. Some previously described COPD loci are eQTL for more than one gene, and trans eQTL analysis identified that cis eQTL SNP for HHIP and IREB2 are also associated with regulation of genes in trans. Funding: K08HL102265, K08 HL097029, P01 HL105339, R01HL094635, R01NR013377, NCT00292552.

989F

Using an integrative mapping approach to identifiy novel genetic factors contributing to pulmonary arterial hypertension (PAH).

W.C. Nichols[1,5], M.W. Pauciulo[1], D. Koller[2], C. Tolentino[1], T.D. Le Cras[3,5], P. Pastura[3], B. Aronow[4,5], D. Li[2], T. Foroud[2]. 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Department of Medical and Molecular Genetics, Indiana University Medical Center, Indianapolis, IN; 3) Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 5) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH.

While mutations in 5 genes have been identified in patients with PAH, data suggest that additional genetic factors contribute. Lacking the necessary large patient cohorts to power genomewide studies, we are using an integrative mapping approach in mice to identify other genetic factors. Mice from 33 strains as well as F_2 animals from 2 parental crosses were housed in hypoxic conditions (10% O2) for 4 weeks. After hypoxia, mice underwent right heart catheterization to measure right ventricular systolic pressure (RVSP). Hearts were dissected to assess the degree of RV hypertrophy. Total RNA was extracted from left lung (LL) and RV of 51 selected animals. Haplotype association mapping (HAM) was performed to identify regions associated with the quantitative traits measured. QTL analysis was performed after genotyping 668 F_2 animals from 2 crosses: PL/J X MRL/MpJ (PM model) and PL/J X FVB/NJ (PF model) using the MUGA panel of 7,851 SNPs. RNASeq was performed using the LL and RV total RNA. A LOD >15 was obtained on chromosome 17 (F_2 threshold LOD = 3.8) for the PM model that accounts for 17% of the total variation in RVSP. There is no convergence of results between the HAM and QTL analyses on this chromosome. The PF model demonstrated no evidence of a QTL for RVSP on chromosome 17. Also, a significant RVSP QTL was identified on chromosome 11 (LOD 5.8) for the PM model. This region on chromosome 11 was identified in the HAM analysis as an association with increases in RVSP after hypoxia. For the PF model, several regions had LOD scores for RVSP above the F_2 threshold LOD (chromosomes 1, 4, 9, and 16). The most significant, LOD of ~7.0, occurred on chromosome 9. HAM analysis of this region approached, but did not reach, the HAM threshold LOD of 5.1. Any overlap between the HAM and QTL analyses for RVSP in the PF model is under investigation. Preliminary RNASeq analyses identified expression differences for 3,232 genes between the high responder PL/J strain and the low responder MRL/MpJ and FVB/NJ strains. Analysis of PL/J at 4 different hypoxia time points identified 1,467 differentially expressed genes clustered into 5 groups with different expression patterns. These different expression patterns can be used to identify genes contributing to PAH when integrated with the HAM and QTL analyses. Identification of additional PAH genes can ideally reveal novel pathways involved in disease pathogenesis that will lead to the development of new patient treatments.

990W

Investigation of classical human leukocyte antigens (HLA) and expression quantitative trait loci (eQTL) within the HLA region in association with Parkinson's disease. *H. Payami[1,2], W. Wissemann[1], E. Hill-Burns[1], C. Zabetian[3], S. Factor[4], N. Patsopoulos[5,6,7], B. Hoglund[8], C. Holcomb[8], G. Thomson[9], H. Erlich[8,10].* 1) Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, NY, USA; 2) Department of Biomedical Sciences, State University of New York Albany, NY, USA; 3) VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA, USA; 4) Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA; 5) Program in Translational NeuroPsychiatric Genomics, Neurosciences Institute, Departments of Neurology, Brigham and Women's Hospital, Boston, MA, USA; 6) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 7) Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA; 8) Human Genetics Dept, Roche Molecular Systems, Pleasanton, CA, USA; 9) Department of Integrative Biology, University of California, Berkeley ,CA, USA; 10) Childrens Hospital of Oakland Research Institute, Oakland, CA, USA.

Historically, tests of association between HLA and disease have been conducted with the 'classical' HLA alleles; i.e., polymorphisms that encode differential ability of HLA molecules to bind and present antigens. Association of PD with HLA, however, was discovered in a hypothesis-free GWAS and replicated using SNP-based studies. The original SNP that reached genome-wide significance, rs3129882, is in intron 1 of DRA and the others map to intergenic sequences near DRA, DRB1 and DRB5. Our aim was to determine if these SNPs are tagging specific HLA alleles. We HLA-typed 2000 PD cases and 1986 controls that were used in the original GWAS (PMID:20711177), and 843 cases and 856 controls (phs000126.v1.p1) for replication using two SNP-based imputation algorithms (de Bakker PMID:21051598 and HLA*IMP:02). We also sequenced 194 cases and 204 controls using 454-next-generation-sequencing, and calculated allele-specific sensitivity and specificity for each algorithm. For association studies we used the imputed data to increase sample size, mindful of allele-specific imputation accuracy. We tested association of HLA alleles and haplotypes with PD, and replicated alleles that passed P<0.05 (set low to minimize false negatives). We then conducted conditional and stratified analyses to test interdependence of the classical variants identified here and the SNPs identified before. PD risk was associated positively with B*07:02_C*07:02_DRB5*01_DRB1*15:01_DQA1*01:02_DQB1*06:02 and negatively with B*40:01_C*03:04_DRB4*01_DRB1*04:04_DQA1*03:01_DQB1*03:02 alleles and haplotypes. However, when conditioned on the SNPs, the signals for the risk variants were abolished, but the ones for the protective variants remained nominally significant (Pconditioned≥1E-4). The SNPs, on the other hand, stayed strongly significant when conditioned on classical variants (Pconditioned≥3E-10). With in-silico studies (SCAN, eqtlBrowser, Genevar), we found that PD-associated SNPs rs3129882 (original GWAS PMID:20711177), rs2395163 (US_Meta analysis PMID:22451204) and rs660895 (French study PMID:22807207) are eQTLs for HLA-DR and HLA-DQ (4E-5≥PeQTL≥2E-79). In summary: Classical HLA-alleles do not explain association of PD with non-coding SNPs. The SNPs are likely eQTLs. Whether the protective factor that is embedded in the protective haplotype is a classical variant affecting antigen binding or a regulatory element affecting gene expression levels is unknown.

991T

Identification of candidate psoriasis susceptibility alleles within an HLA-C enhancer element. *R.C. Trembath[1,2], A. Clop[2], A. Bertoni[2], S.L. Spain[2], M.A. Simpson[2], V. Pullabhatla[2], R. Tonda[3], C. Hundhausen[2], P. Di Meglio[2], P. De Jong[4], A. Hayday[5], F.O. Nestle[2], J.N. Barker[2], R.J.A. Bell[6], F. Capon[2].* 1) Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AD, United Kingdom; 2) Division of Genetics and Molecular Medicine, King's College London, London SE1 9RT, UK; 3) Centre for Research in Agricultural Genomics (CRAG), Campus Universitat Autònoma de Barcelona, 08193 Cerdanyola del Valles, Spain; 4) BACPAC Resources Centre, Children's Hospital Oakland, Oakland, CA 94609, USA; 5) Division of Immunology, Infection & Inflammatory Diseases, King's College London, London SE1 9RT, UK; 6) Department of Neurosurgery, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA 94115, USA.

Psoriasis is a common inflammatory skin disorder that is inherited as a complex genetic trait. Although genome-wide association scans (GWAS) have uncovered more than 30 disease susceptibility regions, up to 50% of the genetic variance is accounted for by single Major Histocompatibility Complex (MHC) locus, known as *PSORS1*. *HLA-C* is widely regarded as the strongest *PSORS1* candidate gene, since markers tagging *HLA-Cw*0602* consistently generate the most significant association signals in GWAS. However, the question as to whether *HLA-Cw*0602* is the causal *PSORS1* allele has not been resolved, especially as the role of SNPs that may affect its expression has not been investigated. Here, we have undertaken a comprehensive molecular characterization of the *PSORS1* region, with the aim of identifying regulatory variants that may contribute to psoriasis susceptibility. We first analysed high-density SNP data and refined the *PSORS1* locus to a 179kb region encompassing *HLA-C* and the neighbouring *HCG27* pseudogene. Next, we compared multiple MHC sequences spanning the 179kb critical interval and catalogued 144 candidate susceptibility variants, which were unique to Cw*0602-bearing chromosomes. In parallel, we determined the epigenetic profile of the critical *PSORS1* interval, uncovering three enhancer elements likely to be active in T lymphocytes. Finally, we showed that nine candidate susceptibility SNPs map within a *HLA-C* enhancer and that three co-localise with binding sites for immune-related transcriptional activators. These data indicate that SNPs affecting the expression of *HLA-Cw*0602* are likely to contribute to psoriasis susceptibility and highlight the importance of integrating multiple experimental approaches in the genetic analysis of complex genomic regions such as the MHC.

992F

Joint eQTL Analysis in Multiple Tissues, Accounting for Heterogeneity and Incomplete Subject Overlap. *G. Li[1], A.A. Shabalin[2], I. Rusyn[3], F.A. Wright[4], A.B. Nobel[1,4].* 1) Statistics and Operations Research, Univ North Carolina, Chapel Hill, NC., Select a Country; 2) Department of Pharmacotherapy & Outcomes Science, Virginia Commonwealth University; 3) Department of Environmental Sciences and Engineering, University of North Carolina; 4) Department of Biostatistics, University of North Carolina.

The analysis of expression quantitative trait loci (eQTL) aims to identify genetic variants that regulate gene transcription, and help dissect complex transcriptional-based mechanisms of disease. Patterns of transcriptional variation are highly tissue-specific, and identifying common and specific eQTLs among various tissues is of great interest. However, most existing methodologies for multiple-tissue eQTL analysis are based on post-hoc analysis of eQTL from individual tissues, and therefore do not fully utilize the data structure of eQTL datasets consisting of multiple tissues. We propose a hierarchical model for the observed correlations of gene-SNP pairs across the available K tissues in a multi-tissue experiment, with a full specification of all 2^K possible patterns of null vs. alternative outcomes across the K tissues. The model allows for incomplete overlap of subjects across the tissues by fitting a flexible correlation structure for genotype-expression associations. The model is fit using an empirical Bayes approach, and provides interpretable posterior probabilities for eQTLs across the range of tissues, as well as control of the local FDR. Simulations demonstrate the approach is stable and computationally feasible for numerous tissues. We further illustrate using RNA-Seq expression data on 9 tissues and genotypes on 3.4 million SNPs, available in pilot form from the Genotyping-Tissue Expression project (GTEx) Consortium, with sample sizes per tissue ranging from 80-200.

993W

Role of common genetic polymorphisms in transgenerational inheritance of inherent as well as acquired traits in budding yeast. *Z. Zhu, Q. Lu, D. Yuan, Y. Li, X. Man, Y. Zhu, S. Huang.* Central South University, Changsha, China.

Both inherent and acquired traits can be transmitted through multiple generations with some traits more stable than others. But the relationship between the stability of such transgenerational inheritance and the genetic variations in an individual or cell has yet to be explored. We studied the effect of genetic polymorphisms on transgenerational inheritance of yeast segregants that were derived from a cross between a laboratory strain and a wild strain of Saccharomyces cerevisiae. For each of 2835 SNPs analyzed, the parental allele present in less than half of the 124 segregants panel was called the minor allele (MA). We found a nonrandom distribution of MAs in the segregants, indicating natural selection, as segregants with high MA content or amount (MAC) were not enriched with MAs from the parental strain that contributed significantly more to the whole set of MAs. We compared segregants with high MAC relative to those with less and found a more dramatic shortening of the lag phase length for the high MAC group in response to 14 days of ethanol training. Also, the short lag phase as acquired and epigenetically memorized by ethanol training was more dramatically lost after 7 days of recovery in ethanol free medium for the high MAC group. Sodium chloride treatment produced similar observations. Using public datasets, we found MAC linkage to mRNA expression of hundreds of genes. Finally, by analyzing a recently published datasets of 1009 yeast segregants that identified numerous additive QTLs for 46 traits, we found by multivariate regression analysis preferential MAC linkage to traits with high number of known additive QTLs (average 16 for the 5 MAC-linked traits vs 12 for the whole set of 46 traits), suggesting that MAC may affect traits via the additive effects of a large number of minor effect SNPs and may thereby account for the 'missing heritability' in most complex traits. These results provide evidence for the slightly deleterious nature of most MAs and a lower capacity to maintain inheritance of traits in individuals or cells with greater MAC, which have implications for disease prevention and treatment. Individuals with high MAC may be more susceptible to environmental pathogens, but they may also be more treatable if treatment was administered relatively early before the disease has progressed past the threshold of no return, because the acquired disease trait may be less stably maintained in these individuals.

994T

Evidence for a genetic contribution to rotator cuff disease in a new population genealogy at the Veterans Administration. *C.C. Teerlink[1], J.M. Farnham[1], L.J. Meyer[2,3], L.A. Cannon-Albright[1,2].* 1) Dept Internal Medicine, Univ Utah Sch Med, Salt Lake City, UT; 2) George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT; 3) Department of Dermatology, University of Utah, Salt Lake City, UT.

The Utah Population Data Base (UPDB) and the Icelandic genealogy resource, combining genealogic and phenotypic data, have proven valuable to genetic studies. We are creating a similar resource including a genealogy of United States residents and their ancestors. We have begun to record link to the national Veteran's Administration (VA) population of 25 million retired service men and women beginning with the genealogy of Utah (UT) and Massachusetts (MA) by obtaining genealogical data from public sources using common published formats. The genealogy currently includes 38 million individuals. We have begun record linking to 11M VA patients using any national VA facility and have currently linked 500,000 (4.5%) of these patients to the genealogy. For genetic analyses we restrict our use to the 52,332 VA patients who link to genealogy data including at least both parents and all 4 grandparents. We can analyze the relationships of patients with phenotypes of interest to define relatedness and relative risks. The recent initiation of the VA Million Veterans Project, which will sample DNA for all Veterans, will soon result in the largest genetic repository in the world. When combined with this VA genealogy, the resource will have unlimited potential to the genetics of health-related phenotypes. As an example, among the 52,332 VA patients who link to genealogy data we identified 1,412 patients with an ICD9 in their medical record consistent with rotator cuff tear. We used the Genealogic Index of Familiality (GIF) statistic to assess the presence of excess familial clustering among these cases compared to 1,000 sets of matched controls (matched on birthplace, age, and sex). The GIF statistic computes the average genetic distance between all pairs of people in a set using the Malecot kinship coefficient. The GIF statistic was computed once for the set of cases and once for each of the 10,000 sets of matched controls. The distribution of the sets of matched controls provides an empirical distribution for a hypothesis test. According to the Gif analysis, rotator cuff tear exhibited strong evidence for excess familial clustering ($p<1E-4$) in the VA genealogy, which further supports prior evidence for a genetic contribution to this phenotype in the Utah Population Database and validates, at least in part, the reliability and utility of the new VA genealogy.

995F

High-dimensional genetic prediction of type-2 diabetes susceptibility. *Y. Klimentidis[1], A. Vazquez[2], G. de los Campos[2], D. Allison[3].* 1) Division of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ; 2) Section on Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 3) Office of Energetics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL.

Recent findings from genome-wide association studies (GWAS) have uncovered approximately 65 loci confidently associated with type-2 diabetes (T2D) risk. However, these loci explain less than ten percent of the expected heritability of T2D, and have limited predictive value. Different statistical methods along with the inclusion of additional variants may increase the utility of genomic information, and ultimately lead to a more confident identification of individuals at highest genetic risk of T2D. In this study, we use genotype information from up to approximately 2.5 million SNPs (single-nucleotide polymorphisms), along with information about each SNP's potential importance in T2D risk, as determined by the summary statistics of a recent meta-analysis of GWAS. We build and test various polygenic prediction models in the context of early-life prediction of T2D susceptibility using 1) G-BLUP (genomic best linear unbiased predictor) and 2) multi-SNP risk scores. Prediction models are developed and tested within the Framingham Heart Study dataset, and the main results are replicated in the ARIC (Atherosclerosis Risk in Communities) dataset. We find that prediction accuracy increases substantially with the inclusion of additional SNPs, from an AUC (area under the receiver operating characteristic curve) of 0.689 for the commonly used 'top hits' genetic risk score, up to 0.814 using a multi-SNP risk score that includes over 3,700 SNPs. Our findings are successfully replicated in the ARIC dataset. We have shown that T2D genetic risk prediction models can be markedly improved with the inclusion of thousands of SNPs, given prior information about relative SNP importance.

996W

Investigating missing heritability and improving risk prediction with maximally free marginal models of phenotype. *N. Zaitlen[1], T. Wingo[2], A. Gusev[3], D.J. Cutler[4].* 1) Department of Medicine, University of California San Francisco, San Francisco, CA; 2) Department of Neurology, Emory University School of Medicine, Atlanta, GA; 3) Department of Genetic Epidemiology, Harvard School of Public Health, Boston, MA; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America.

Genetic association studies have successfully identified thousands of loci in hundreds of phenotypes. In order to increase power studies typically utilize one degree of freedom statistical tests, making a restrictive additive (or multiplicative) assumption about the underlying genetic basis of the phenotype of interest. In this work we examine the validity of this choice, introducing new methods to determine how much better a given phenotype is modeled with unrestricted effect sizes for zero, one, or two copies of each causal variant. We know from both human and model organism genetics that many variants behave in a dominant, recessive, or over-dominant manner, but the total extent to which these richer marginal models of disease affect complex phenotypes is currently unknown. Our approach is an extension of the recent mixed linear model approaches to estimating the phenotypic variance driven by genotyped variants (Yang et al Nat Gen 2010), which also assume a one degree of freedom model. We propose multi-variance component methods, that estimate the additional phenotypic variance captured by increasingly freer disease models. The variance components are constructed by measuring the genome-wide similarity of individuals according to different encodings of genotypes. The method has a corresponding likelihood ratio test with a null model of additivity, allowing a formal examination of the additive assumption. We show via extensive simulation over real and simulated data that our test statistic has the appropriate type-I error rate and accurately estimates the fraction on phenotypic variance driven by each disease model. We apply our method to phenotypes from the National Finnish Birth Cohort and WTCCC data sets, show several potential sources for bias including ascertained data sets, relatedness, and population structure, and confirm the known dominance structure in Type 1 Diabetes (p-value $< 5e-5$). We derive a BLUP based phenotypic prediction method from our mixed linear model, which will improve our prediction ability over existing additive models for those phenotypes with more complex underlying structure. This work will add to our fundamental understanding of the genetic basis of complex human phenotypes and offer insights into the missing heritability problem.

997T

Life course variations in the heritability of body size. *J. Zhao[1], J.A. Luan[1], S.J. Sharp[1], R. Hardy[2], A. Wong[2], Q. Tan[3], N.J. Wareham[1], D. Kuh[2], K.K. Ong[1,2].* 1) MRC Epidemiology Unit, University of Cambridge, Institute of Metabolic Science, Box 285, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK; 2) 2MRC Unit for Lifelong Health and Ageing, University College London, London WC1B 5JU, UK; 3) Epidemiology and Biostatistics, University of Southern Denmark, J.B. Winslows Vej 9, DK-5000 Odense C, Denmark.

Background: It has been shown recently that whole genome data can facilitate estimation of genetic contributions to a variety of traits via a mixed model framework as implemented in GCTA and R/SAS (Yang et al. Nat Genet 2010, 42:565-9; Zhao & Luan. J Prob Stat 2012, doi 10.1155/2012.485174). Our aim was to use this approach to investigate the life course variations in heritability of body size. Methods: We analysed height, weight and body mass index variables at 11 time-points in 2,452 individuals (1,225 men, 1,227 women) born in 1946 and enrolled in the MRC National Survey of Health and Development (NSHD), with genotypes at 147,949 single nucleotide polymorphisms (SNPs) on Metabochips which were subsequently imputed to 506,255 according to the 1000Genomes project. We obtained genome-wide kinship matrices using genotypes at SNPs on Metabochips and genotypes at all SNPs, which were used in mixed models as implemented in the computer program GCTA. Results were also compared to those obtained using an alternative procedure of kinship estimation in PLINK and mixed models in R. Results: In line with earlier findings that specific genetic variants have variable temporal effects in this cohort (Hardy et al. Hum Mol Genet. 2010; 19:545-552), we observed age-related variations in heritability estimates. Estimates based on genotypes at SNPS on Metabochips and genotypes at all SNPs were comparable but generally lower than recently reported GCTA estimates with mean(range) being 0.09(0-0.50), 0.11(0-0.20), 0.10(0-0.22) for height, weight and body mass index, respectively. Variation in estimates was also seen between alternative procedures. Conclusion: This work supports the utility of large-scale genotype data in heritability estimation and highlights the age-related variability in genetic contributions to body size across the life course. Further work will be to distinguish the effects of established variants and to consider estimates in a unified longitudinal model including contrast with models assuming various degrees of temporal homoscedasticity.

998F

Quantifying missing heritability from known GWAS loci and rare coding variants. *A. Gusev[1,2], B.M. Neale[2], G. Bhatia[1,2], N. Zaitlen[3], B.J. Vilhjalmsson[1,2], D. Diogo[4,2], E.A. Stahl[4,2], P.K. Gregersen[5], J. Worthington[6], L. Klareskog[7], S. Raychaudhuri[4,2], R. Plenge[4,2], B. Pasaniuc[8], P.F. Sullivan[9], A.L. Price[1,2].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America; 2) The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America; 3) Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco, California, United States of America; 4) Division of Rheumatology, Immunology, and Allergy and Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; 5) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, United States of America; 6) Arthritis Research UK Epidemiology Unit, Musculoskeletal Research Group, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom, NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Fou; 7) Rheumatology Unit, Department of Medicine, Karolinska Institute at Karolinska Hospital, Stockholm, Sweden; 8) Interdepartmental Program in Bioinformatics Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, California, United States of America; 9) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America.

GWAS currently explain only a small fraction of common disease heritability. Possible sources of missing heritability include rare coding variants throughout the genome, or additional causal variants at known GWAS loci, each of which may be poorly tagged by GWAS associations. We describe variance component methods to estimate these components of heritability while adjusting for LD both within and between variant classes, which can bias estimates even when all causal variants are typed. Our simulations on real genotypes show that our methods produce unbiased estimates. We applied the methods to two data sets: a 23,000-sample study of rheumatoid arthritis (RA) with ImmunoChip data from 10 known RA loci (excluding HLA), and a 6,400-sample study of schizophrenia typed on both exome and GWAS chips. For RA, known associations at these loci (including those identified in these data) explained 0.006 of the variance in liability of RA, while including all SNPs at the loci explained 0.014±0.002, increasing further to 0.032±0.003 when including data from 17 known autoimmune loci not previously associated to RA. These findings are consistent with the presence of additional causal variants at previously known loci. In the schizophrenia analysis, we partitioned total heritability explained by all typed SNPs of 0.38±0.04 into 0.09±0.03 from coding SNPs and 0.30±0.03 from noncoding GWAS-chip SNPs, demonstrating significant exonic heritability. However, the contribution of 0.04±0.03 from rare coding SNPs (after adjusting for LD between variants) was non-significant, even after collapsing rare variants to reduce statistical noise. The remaining independent contribution from common coding SNPs was significant but largely tagged by GWAS-chip SNPs. Partitioning the coding variants further by functional classification revealed a 4.6x enrichment in heritability at severe loss-of-function sites (P=0.02 by permutation) but no significant enrichment at other predicted-damaging coding variants. Our results shed light on components of missing heritability for these traits and provide insights into their genetic architecture.

999W

Assessing the Impact of Coding Variants on Lipid Levels with the Exomechip. *K.E. Stirrups[1], N. Masca[2,3], T.V. Varga[4], R. Scott[5], L. Southam[1], W. Zhang[6], D. Pasko[7], A. Mahajan[8], S. Kanoni[1], UK Exomechip Lipids Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 2) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 3) NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK; 4) Genetic and Molecular Epidemiology Unit, Lund University Diabetes Center, Malmo, Sweden; 5) MRC Epidemiology Unit , Institute of Metabolic Science, Cambridge, UK; 6) School of Public Health, Imperial College London, London, UK; 7) Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK; 8) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

Multiple rounds of Genome Wide Association (GWA) meta-analyses have so far implicated mainly common variants underlying the molecular basis of blood lipid levels in humans, explaining only a fraction of the genetic component of these traits. The aim of the Illumina HumanExome chip (247,870 SNPs) was to provide a cost-effective way to assess the genetic contribution of low frequency (minor allele frequency (MAF) < 5%) and rare (MAF < 1%) exonic variants which are likely to be functional. An estimated 97% of nonsynonymous variation and 94% of splice and stop variation detected in an average genome through exome sequencing are covered by the design. The array also includes a scaffold of common tag SNPs, HLA and mitochondrial markers as well as all known GWAS lead SNPs (NHGRI catalogue).

We have set up a collaborative effort to perform an inverse variance weighted meta-analysis in European and South-East Asian cohorts (current analysis includes ~20,000 samples) for high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides (TG) and total cholesterol (TC) levels. Genotype calling was performed using GenCall followed by zCall. Meta-analysis (autosomal data) was performed in GWAMA, testing fixed and random effects. We excluded subjects on lipid lowering medication and analysed disease and control cohorts separately.

We implemented additional quality control steps for filtering genotypes obtained for rare variants. In total, ~175,000 variants passed QC criteria. We then examined SNPs which were associated with lipid traits at a significance threshold of P < 1×10-4 after inspection of cluster plots. The majority of these SNPs were located within 500kB of a known lipid locus (~40% of reported loci replicated at nominal significance), though in 5 loci the known association was to a lipid trait other than the one analysed. We identified a coding variant in 12 known loci showing a stronger association than the reported sentinel SNP, these are good functional candidates. Interestingly, for 21 known loci we found a variant which was not in LD with the sentinel SNP and remained significant upon conditional analysis, most of them (76%) were low frequency / rare variants. We are currently pursuing replication approaches for 74 (78% are coding SNPs) putative new signals of which ~64% have MAF <1%.

1000T

Narrowing the gap on heritability of common disease by direct estimation in case-control GWAS. *D. Golan, S. Rosset.* Department of Statistics, Tel-Aviv University, Tel-Aviv , Israel, 69978.

One of the major developments in recent years in the search for missing heritability of human phenotypes is the adoption of linear mixed-effects models (LMMs) to estimate heritability due to genetic variants which are not significantly associated with the phenotype. A variant of the LMM approach has been adapted to case-control studies by Lee et al. (2011) and applied to many major diseases, successfully accounting for a considerable portion of the missing heritability. For example, for Crohn's disease the estimated heritability was 22%, considerably higher than the heritability accounted for by significantly associated SNPs. We propose to estimate heritability of disease directly by regression of phenotype similarities on genotype correlations, corrected to account for ascertainment (the enrichment of cases in the study compared to the population). We refer to this method as genetic correlation regression (GCR). Using GCR we estimate the heritability of Crohn's disease at 34% using the same data as Lee et al., much closer to the estimated heritability of 50-60% obtained from family studies. GCR estimates of heritability are also considerably higher than LMM for bipolar disorder (54% compared to 43%, respectively). To understand the gap between GCR and LMM estimates, we conduct extensive simulations to evaluate the performance of both heritability estimation methods. Contrary to previous simulations conducted in the context of heritability estimation, our simulations are the first to simulate the entire generative process of data collection, including generating genotypes and selecting cases and controls. This realistic simulation scheme captures some essential features of case-control GWASs which were overlooked by previous simulation schemes, for example, when a highly heritable polygenic phenotype is rare, randomly sampled cases tend to be more genetically similar than random controls, even when they are unrelated. Our simulations demonstrate that GCR produces accurate and unbiased estimates of heritability for a wide range of scenarios, while LMM estimates are considerably and consistently negatively biased under ascertainment. We thus conclude that GCR estimates of heritability are more reliable than LMM estimates. Lastly, we develop a heuristic correction to LMM estimates, which can be applied to published LMM results. Applying our heuristic correction increases the estimated heritability of multiple sclerosis from 30% (Lee et al. 2013) to 52.6%.

1001F

GWAS-data based heritability estimation of memory related phenotypes. *C. Vogler[1,2,4], V. Freytag[2], A. Milnik[2], L. Gschwind[2], D. Coynel[3], A. Heck[1,2,4], D. de Quervain[1,3], A. Papassotiropoulos[1,2,4].* 1) Psychiatric University Clinic, University of Basel, Basel, Basel City, Switzerland; 2) Moleuclar Neuroscience, University of Basel, Basel, Basel City, Switzerland; 3) Cognitive Neuroscience, University of Basel, Basel, Basel City, Switzerland; 4) Department Biozentrum, Life Sciences Training Facility, University of Basel, Basel, Switzerland.

Background: Genome-wide association studies provide the means to identify the molecular underpinnings of complex heritable traits. Heritability is a concept that summarizes how much of the phenotypic variation in a trait is attributable to genetic factors. Conventionally, heritability estimates in humans are derived from phenotypic data solely by comparing correlations between relatives. However, those methods imply that the amount of resemblance that is due to shared environment is the same for different degrees of relationship. Further, the genetic identity between sib pairs is assumed to be 50%, whereas it has been shown to vary from ~40 to ~60% (Visscher, 2006). Alternatively, recently developed methods, propose to infer genetic identity from high-throughput SNP data and to correlate these estimates with phenotypic resemblance among unrelated individuals. This allows heritability estimation for any specific GWAS dataset. Here we present genetic marker based heritability estimates for working and episodic memory phenotypes.

Methods: A healthy young Swiss sample comprising a total of N=1789 individuals was subjected to phenotypic assessment with a cognitive testing battery and subsequent genotyping using the Affymetrix SNP Array 6.0. Working memory (WM) performance was measured with the n-back paradigm. Episodic memory was assessed with the International Affective Picture System (IAPS) and lists of semantically unrelated words taken from the collections of Hager and Hasselhorn (Hager and Hasselhorn, 1994). In order to obtain the heritability estimates, we performed restricted maximum likelihood (REML) analyses implemented in the GCTA software package (Yang et al, 2011).

Results: Heritability estimates occupy a range from 26 to 65% with a mean standard error of 19% The average heritability estimate for the attention corrected WM performance is 60% (SE: 19%) and the mean heritability for the different episodic memory tasks is 49% (mean SE: 20%). These estimates are consistent with previously reported heritability for WM (Ando, 2001) and episodic memory (Papassotiropoulos, 2011). As this reflects that heritability for the cognitive phenotypes is represented in the GWAS data, this design can serve to unravel the molecular players that act in concert to form these complex traits.

1002W

Polygenic analysis of type II diabetes in the DIAGRAMv3 GWAS meta-analysis, and partitioning of polygenic signal in tissue-specifically active genes. *A.L. Dobbyn[1], B.F. Voight[2], E.E. Kenny[1,3,8,9], J. Jeff[1,3], P. Roussos[1,6,7], O. Gottesman[3], A. Morris[4], R. Loos[1,3], E. Bottinger[3], S. Raychaudhuri[5,6], M. McCarthy[4], E.A. Stahl[1,6,7], the DIAGRAM Consortium.* 1) Institute for Genomics and Multiscale Biology, Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Departments of Pharmacology and Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 5) Partners HealthCare Center for Personalized Genetic Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA; 6) Broad Institute, Cambridge, MA; 7) Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY; 8) Center for Statistical Genetics, Icahn School of Medicine at Mount Sinai, New York, NY; 9) Dept. of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

INTRODUCTION: Polygenic risk score analyses have shown that a substantial portion of the missing heritability of common, complex diseases can be explained by weakly associated GWAS SNPs that fail to reach genome-wide significance. Separate analyses have demonstrated that tissue specific chromatin marks can be used to identify causal tissues by enrichment in GWAS associations (Trynka 2013). We combine these approaches, partitioning polygenic scores across liver- and pancreatic islet-specific gene sets, in order to dissect the genetic architecture of type 2 diabetes (T2D), a common metabolic disorder that affects up to 8% of the adult population. METHODS: We first conducted polygenic risk score analysis of T2D using the DIAGRAMv3 GWAS (Morris 2012; 12,171/56,862 European cases/controls). Polygenic risk scores based on LD-pruned SNPs stratified by discovery P_{GWAS} and risk allele frequency (RAF) were tested in either WTCCC1 (1,924/2,938 cases/controls) or the Mount Sinai Biobank (4,095/9,832 multiethnic cases/controls). We then identified MACS-called H3K4me3 ChIP-seq peaks from the NIH Epigenomics Roadmap project (34 tissues) in the promoter regions of Refseq transcripts (TSS +/- 1Kb), and normalized to obtain genic tissue-specificity scores. We have explored distance from transcription start and end, and nearest recombination hotspots, to define SNP sets for partitioned polygenic analyses. RESULTS: Polygenic scores were significantly associated with T2D in the test data, with the strongest score having $P=3\times10^{-29}$ and $R^2=0.035$ (841 SNPs with $P_{GWAS} < 0.001$), and with many additional scores having $P < 0.01$ (SNPs with $P_{GWAS} < 0.1$ and $0.1 < RAF < 0.88$). 761 (1.80%) and 442 transcripts (1.05% of all transcripts) constitute primary liver and pancreatic islet gene sets (specificity score > 0.5), respectively, each corresponding to 3.3% of LD-pruned SNPs (transcripts +/- 100Kb). Top polygenic scores for these SNP sets are significant (liver $P=0.002$, $R^2=0.002$, 28 SNPs with $P_{GWAS} < 10^{-3}$; pancreas $P=0.001$, $R^2=0.001$, 817 SNPs with $P_{GWAS} < 0.05$), and show 1.5- to 2.1-fold greater R^2 than expected for the proportion of SNPs represented. CONCLUSIONS: Tissue-specific compartments of polygenic signal can be derived from gene harboring chromatin marks, and will be applied in further studies of T2D causal mechanisms including the relative roles of liver versus pancreatic islets, and pleiotropy with T1D and cardiac/metabolic phenotypes.

1003T

Bivariate analysis of blood pressure traits in Danish and Chinese twins. *S. Li[1], T.A. Kruse[1], K.O. Kyvik[2].* 1) Human Genetics, Institute of Clinical Research, University of Southern Denmark, Odense, Denmark; 2) Institute for Regional Health Research, University of Southern Denmark and Odense Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark.

Studying the multiplex genetic basis of blood pressure regulation can help with elucidating the etiology of hypertension and related diseases as well as with the development of more efficient prevention and treatment strategies. This study aims at investigating the genetic and environmental backgrounds for blood pressure variation as well as for phenotype correlation between systolic and diastolic blood pressure by fitting bivariate models to Danish and Chinese twins and comparing ethnic difference between the two samples. The Danish sample consists of 756 twin pairs (309 monozygotic or MZ pairs, 447 dizygotic or DZ pairs) and Chinese sample includes a total of 325 pairs of twins (183 MZ twin pairs and 142 DZ twin pairs). Our parameter estimates revealed moderate to high additive genetic and unique environmental influences on both systolic and diastolic blood pressure and on their correlation with an overall pattern of higher genetic regulation in Danish than in Chinese twins while higher contribution by unique environment in Chinese than in Danish twins. The estimated genetic components in blood pressure and in blood pressure correlation could serve to guide new researches aiming at looking for heritable genetic variants that affect systolic and/or diastolic blood pressure. Meanwhile, the estimated high contribution to blood pressure by unique environment suggests that promoting healthy lifestyle could provide an efficient way for the control of hypertension especially in the Chinese population.

1004F

The genetic architecture of schizophrenia in the Swedish Schizophrenia Study. *E.A. Stahl[1], S.H. Lee[2], G. Moser[2], S. Ripke[3,4], D.M. Ruderfer[1], S.M. Purcell[1,3,4], P. Sklar[1,4], C. Hultman[5], P.F. Sullivan[6].* 1) Division of Psychiatric Genomics in the Department of Psychiatry, Friedman Brain Institute, and Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NY, NY 10029; 2) The University of Queensland, Queensland Brain Institute, Brisbane, 4072, Queensland, Australia; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, 02114; 4) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, 02142; 5) Department of Medical Epidemiology, Karolinska Institutet, Stockholm, Sweden; 6) Departments of Genetics, Psychiatry, and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599.

INTRODUCTION: Schizophrenia has long been recognized to be familial, and more recently to be highly polygenic. The substantial contribution to phenotypic variance of SNPs falling short of genome-wide significance in GWAS is well appreciated, particularly in schizophrenia. Here we describe inference analyses of schizophrenia genetic architecture using the comprehensive genomic data collected on Swedish national registry samples. METHODS: We analyzed GWAS data of 5,001 schizophrenia cases and 6,243 healthy controls from the Sweden (imputed to 1000 Genomes, 5.7M SNPs after QC) as test data in polygenic risk score (PRS) analyses with PGC schizophrenia GWAS meta-analysis results for training data. Approximate Bayesian Polygenic Analysis (ABPA) was performed under a mixture model of independent associated and null SNPs (Stahl 2012 PMID 22446960). Linear mixed models of normally distributed effects were fit using GCTA (Yang 2011 PMID 21167468, Lee 2011 PMID 21376301), and LD-based SNP weighting (LDAK, Speed 2013 PMID 23217325) was considered. Results were compared using the actual genotype data, with real and simulated phenotypes under a range of genetic architectures. RESULTS: ABPA yields an estimate of 8,300 independent SNPs (95% CI 6,300-10,200) underlying risk of schizophrenia, almost all common (MAF>0.1), and together accounting for heritability of 52% (95% CI 0.45-0.54). GCTA yielded a somewhat lower estimate, 33% (0.27-0.39), still a majority of heritability estimated in the Swedish national pedigree study (64%, Lichtenstein 2009 PMID 19150704). CONCLUSION: These results answer a question that has been debated most of the past century: the majority of the phenotypic heritability of schizophrenia can be explained by GWAS data, overwhelmingly by common SNPs of small effect. These results strongly suggest that additional GWAS will yield many more genetic discoveries. Furthermore, our results suggest that schizophrenia may be more polygenic than other common, complex diseases.

1005W

Heritability Contribution of Gut Microbiome to Metabolic Traits in Mice. *E. Kostem[1], E. Org[2], B. Parks[2], R. Knight[4,5], J. Lusis[2,3], E. Eskin[1,3].* 1) Computer Science Department, UCLA, LOS ANGELES, CA; 2) Department of Medicine/Division of Cardiology, David Geffen School of Medicine, UCLA, LOS ANGELES, CA; 3) Department of Human Genetics, David Geffen School of Medicine, UCLA, LOS ANGELES, CA; 4) Department of Chemistry and Biochemistry, University of Colorado Boulder, CO; 5) Howard Hughes Medical Institute, Boulder, CO.

Over the last few years, the impact of the gut Microbiome composition in complex traits has received a lot of attention. The gut Microbiome hosts hundreds of bacterial species with a tremendous genetic diversity and has been shown to regulate complex traits in addition to the host genome. Using the Hybrid Mouse Diversity Panel (HMDP) that consists of more than 100 inbred strains of mice, we show that the gut Microbiome composition explains a considerable portion of the variation in metabolic traits. We further partition the heritability of metabolic traits to the contribution of genomic regions and show that these estimates may be different if we account for the gut Microbiome. Our findings suggest that gut Microbiome may be one of the key factors in the missing heritability of complex diseases.

1006T

Proteomic and bioinformatics analysis of the endolysosomal pathway: Deciphering the insulin receptor Golgi/endosome subnetwork (IRGEN).
M.B. DJIDJOU[1], C. LANDRY[2], R. Laframboise[1], R. Faure[1]. 1) PEDIATRICS, CHUQ Research Center CRCHUQ (Centre-Mère-Enfant-Soleil)/University Laval, Quebec, QUÉBEC, Canada; 2) Institut de Biologie Intégrative des Systèmes (IBIS), Université Laval, Québec, PQ, Canada.

While insulin resistance can be a heritable trait in lean and obese subjects, it remains unclear which cellular mechanism is the dominant driver that determines insulin sensitivity under baseline conditions. The study of the proteomes of lower organisms recently confirmed that proteins are not randomly connected; most interact with just a few others, whereas a limited number (hubs) interact with many others. This nodular architecture is conserved and provides a robust defense against random alterations, but connected proteins-which tend to play essential roles in organisms-can also make networks vulnerable to environmental attacks that target nodes, suggesting that the study of interactomes in rodent models and human is important to understand complex diseases such as diabetes. It is plausible that changes in insulin receptor (IR) traficking will alter circulating insulin concentrations, which in turn could affect insulin-sensitive tissues thus supressing peripheral insulin action. A proteome of a Golgi/endosomes (G/E) fraction prepared from the mouse liver was previously described by us. Bioinformatics analyses using well documented databases such as MGI, STRING and BIOGRID, completed by powerful analytical algorithms now show that they enriched for specific groups of proteins and depleted for others. Proteins identified in the G/E fraction can be grouped and linked according to their functional and molecular associations in the cell whereby they form tight and interconnected clusters, visualized through Cytoscape software, in a robust scale-free protein-protein interaction network (PPIN) of 927 nodes for 4617 interactions. This represents the G/E network (GEN), having a general vesicle-like topology. A close analysis of the insulin receptor subnetwork (IRGEN) identified physically associated proteins, which are involved in the cell cycle machinery (Cdk2, cyclin E), vacuolar ATPase elements, the de novo purine biosynthetic pathway (ATIC), a putative protein tyrosine phosphatase (PTPLAD1), actin cytoskeleton (β-catenin, CEACAM1), cargos (LRP-1) and kinases (AMPK). These proteins were confirmed to have a role in IR regulation. The results indicate the presence of unanticipated pathways subtending both IR signalling and trafficking favoring the idea that insulin clearance may be a primary event, which could generate hyperinsulinemia that negatively act on insulin-sensitive tissues, thus generating insulin resistance.

1007F

Transcriptomics of natural resistance to SIV infection in wild African green monkey populations in Africa.
A. Jasinska[1], D. Ma[2], N. Tran[1], C. Blum[1], P. Turner[3,4], T. Grobler[4], Y. Jung[1], C.A. Schmitt[1], M. Dion[5], M. Antonio[5], G. Coppola[6], I. Pandrea[2], C. Apetrei[2], N. Freimer[1]. 1) UCLA Center for Neurobehavioral Genetics, Univ California, Los Angeles, Los Angeles, CA; 2) Canter for Vaccine Research, University of Pittsburgh, Pittsburgh, Pennsylvania USA; 3) Department of Anthropology, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201; 4) Department of Genetics, University of the Free State, Bloemfontein, South Africa,; 5) Medical Research Council, Fajara, The Gambia; 6) Department of Neurology, Univ. of California, Los Angeles.

Resistance to acquiring the human immunodeficiency virus (HIV) infection has been observed in some individuals, who remain uninfected despite their exposure. These people are known as exposed seronegative or exposed uninfected (ESN/EU) individuals. This phenomenon offers a unique opportunity to identify host factors protecting against HIV infection. While this trait is rare in humans, we have recently discovered that it is present at a high frequency (~20%) in wild populations of African green monkeys (AGM, *Chlorocebus*), also called vervets, which are natural hosts of simian immunodeficiency virus (SIV), a virus closely related to HIV. We hypothesized that host genetic factors underlie resistance to acquiring SIV infection among AGMs. In order to identify host genes, which are critical to the evasion of infection in the face of massive exposure among AGMs, we used RNA-seq to analyze blood transcriptomic profiles in 19 ESN, 4 acutely SIV-infected and 26 chronically SIV-infected AGMs from wild *Ch. pygerythrus* and *Ch. sabaeus* populations from South and West Africa, respectively. Transcriptome profiling in the *pygerythrus* population indicated that a large group of genes transiently deregulated during the acute (early) stage of infection encompasses interferon-regulated genes, and genes associated with defense, stress response immunity, and inflammation. We also identified 23 transcripts deregulated between ESNs and individuals chronically infected with SIV (in the later stages of infection), with the strongest statistical evidence for IL1F7 (interleukin-1 family member 7), RPS2 (40S ribosomal protein S2), C4BPA (complement component 4 binding protein, alpha), RAD21L1 (RAD21-like 1), and APOE (Apolipoprotein E). We identified the deregulation of several genes previously linked to the control of HIV infection in humans: RPS2, which codes ribosomal protein regulated by HIV tat expression, APOE, whose common alleles are associated with the accelerated progression to AIDS in humans, and genes from the family of defensins (DEFA1, DEFA1B, and DEFA3), which are natural anti-HIV peptides. Among the deregulated genes in pygerythrus, three genes were also top-ranked in *sabaeus* AGMs: RPS2, APOE, and IFIT1B. Our transcriptomic analysis revealed genes that are deregulated in ESN AGMs and therefore possibly involved in the mechanism of defense against infection. These are candidates for functional follow-up studies and potential targets for preventive therapy.

1008W

Knockdown of zebrafish *crispld2* results in craniofacial defects and abnormal NCC migration.
Q. Yuan, E.C. Swindell, J.T. Hecht. Department of Pediatrics, Pediatric Research Center, University of Texas Medical School at Houston, Houston, TX.

We have previously found that variation in the CRISPLD2 (cysteine-rich secretory protein LCCL domain containing 2) gene is associated with nonsyndromic cleft lip and palate (NSCLP) and that mouse Crispld2 is expressed in the developing craniofacies, suggesting that CRISPLD2 may play a role in craniofacial morphogenesis. Crispld2 was first identified in mice (as a gene called Lgl1) and plays a role in early branching morphogenesis of the developing lung and kidney and in epithelial/mesenchymal interaction in lung tissue. Defects in the epithelial to mesenchymal transition in craniofacial tissue are thought to underlie nonsyndromic orofacial clefting. To define the role of CRISPLD2 in zebrafish craniofacial development, antisense morpholinos (MO) directed against the start site (MO3) and exon/intron boundaries for exons 3 and 4 (MO1 and MO2, respectively) of zcrispld2 were injected into early embryos. These embryos showed a range of craniofacial abnormalities involving both the jaw and palate, with the most severe phenotypes resulting from knockdown of MO3. Moreover, several marker genes for neural crest cells (NCCs) show abnormal expression in MO3 injected embryos, suggesting that knockdown of zcrispld2 specifically affects NCCs. These results show that zcrispld2 is involved in palate and jaw morphogenesis. We are now examining the migration of NCCs in live embryos to determine when zcrispld2 plays a role in craniofacial development. RNAseq is being utilized to identify potential downstream effectors of zcrispld2. Importantly, our results demonstrate that genes involved in complex birth defects, such as NSCLP, can be identified in candidate gene studies and functionally assessed in zebrafish to define the biological processes that are the basis for these complex birth defects.

1009T

Whole genome analysis in fibromyalgia suggests a role for the central nervous system in disease susceptibility. *E. Docampo[1,2,3], G. Escaramis[1,2,3], M. Gratacos[1,2,3], S. Villatoro[1,2,3], A. Puig[1,2,3], M. Kogevinas[4], A. Collado[5], J. Carbonell[6], J. Rivera[7], J. Vidal[8], J. Alegre[9], R. Rabionet[1,2,3], X. Estivill[1,2,3].* 1) Genetic Causes of Disease Group, Centre for Genomic Regulation (CRG), Barcelona, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; 4) CREAL, Centre de Recerca en Epidemiologia Ambiental, CREAL, Barcelona, Spain; 5) Rheumatology Service, Hospital Clinic, Barcelona, Spain; 6) Rheumatology Service, Hospital del Mar, Barcelona, Spain; 7) Rheumatology Unit, Instituto Provincial de Rehabilitación, Madrid, Spain; 8) Rheumatology Unit, Rheumatology, Guadalajara, Spain; 9) Chronic Fatigue Syndrome Unit, Internal Medicine, Barcelona, Spain.

Fibromyalgia (FM) is a highly disabling syndrome defined by a low pain threshold and a permanent state of pain. The mechanisms explaining this complex disorder remain unclear and its genetic factors have not been identified. The aim of this study was to elucidate genetic susceptibility factors for FM. We used the Illumina 1 million duo array to perform a genomewide association study (GWAS), and Agilent's 2X400K platform for array comparative genomic hybridization (aCGH) to identify regions varying in copy number that could be involved in FM susceptibility. GWAS was performed in 300 FM cases and 203 controls. No SNP reached GWAS significance, but 21 of the most associated SNPs were chosen for replication in over 900 cases and 900 pain free-controls. Four of the strongest associated SNPs selected for replication showed a nominal association in the joint analysis, and one, rs11127292 (MYT1L) was found to be associated to FM with low comorbidities ($p=4.28\times 10^{-5}$, OR (95%CI)=0.58 (0.44-0.75). By aCGH, an intronic deletion in NRXN3 showed to be associated to female cases of FM with low levels of comorbidities (p=0.021, OR (95%CI)= 1.46 (1.05-2.04)). Both GWAS and aCGH results point at a role for the central nervous system in FM genetic susceptibility. If the proposed FM candidate genes are further validated in replication studies, this would highlight a neurocognitive involvement in this disorder, currently considered musculoskeletal and affective. This work was supported by ENGAGE, NOVADIS (EU and Spanish Ministry of Economy and Competitiveness).

1010F

Sex-specific heritability of traits related to human obesity in rhesus macaques. *A. Vinson[1,2], A.D. Mitchell[1], D. Toffey[2], M.J. Raboin[2].* 1) Molecular and Medical Genetics, Oregon Hlth & Sci Univ, Portland, OR; 2) Oregon National Primate Research Center, Oregon Hlth & Sci Univ, Portland, OR.

Obesity is a major component of the metabolic syndrome and is associated with insulin resistance, type 2 diabetes mellitus, and increased risk for cardiovascular disease. Numerous studies in humans have demonstrated a genetic contribution to obesity. Further, obesity is sexually dimorphic in humans, and sex-specific genetic effects on human obesity have also been reported. The goals of this study were to explore the utility of the rhesus macaque as a genetic model for human obesity, by assessing the extent of additive genetic effects on traits related to human obesity and the influence of sex on these genetic effects. We measured weight, body mass index (BMI), abdominal circumference, and waist-to-thigh ratio in 584 Indian-origin rhesus macaques selected from a single 6-generation, 1,289-member pedigree unascertained with respect to phenotype, and designed for statistical power in quantitative trait analysis and gene mapping. Using a maximum likelihood-based variance decomposition approach, after removing mean effects related to housing, age, and sex, we found significant heritability for all traits in the total sample (h^2 range 0.247~0.407, P-value range 7.44×10^{-10}~2.43×10^{-3}). When stratified by sex (N=361 females, 223 males), analyses revealed apparent differences in heritability between males and females for BMI ($h^2 = 0.703$ in males, $P=7.75 \times 10^{-5}$, vs. $h^2 = 0.198$ in females, P=0.07), abdominal circumference ($h^2 = 0.258$ in females, P=0.05, vs. $h^2 = 0.101$ in males, P=0.25), and waist-to-thigh ratio ($h^2 = 0.273$ in females, P=0.03, vs. $h^2 = 0.191$ in males, P=0.06). We conclude that genetic contributions to traits related to human obesity can be detected and measured in rhesus macaques, and that genetic contributions to some traits may differ by sex. Our results are consistent with substantial similarity between macaques and humans in the genetic architecture of obesity, and support the value of the rhesus macaque as a genetic model for this human condition.

1011W

Gene expression changes in response to paraquat and caffeine in a Drosophila model of Parkinson's Disease. *E.M. Hill-Burns[1], P. Ganguly[1], W.T. Wissemann[1], W.J. Wolfgang[2,3], H. Payami[1,3].* 1) Division of Genetics, NY State Department of Health Wadsworth Center, Albany, NY; 2) Division of Infectious Disease, NY State Department of Health Wadsworth Center, Albany, NY; 3) Department of Biomedical Sciences, State University of New York Albany, NY.

Previous studies have established that paraquat exposure can induce Parkinsonism in Drosophila and serve as a model for Parkinson's disease (PD) in humans. Consistent with findings in humans that coffee drinking correlates with lower risk of PD, our lab extended the Drosophila paraquat model, observing that exposure to caffeine can partially rescue paraquat-induced mortality. To better understand the mechanisms at work in these Drosophila models and how they might relate to PD pathology in humans, we attempted to identify genes whose expression changed significantly in response to paraquat and/or caffeine. Flies were kept on food containing either no caffeine or 1mM caffeine for ten days, and then transferred to food containing either no paraquat or 5mM paraquat (in addition to the original concentration of caffeine), for a total of four treatment conditions with 3-4 replicates each of 30 female flies per vial. Fly heads were dissected after six days of paraquat treatment (or no paraquat treatment), RNA was extracted from the pooled tissue of each replicate, and cDNA was synthesized and hybridized onto arrays. Expression (of 18,954 transcripts, using Affymetrix expression arrays) was assayed in dissected heads of untreated flies, flies exposed to paraquat, flies exposed to caffeine, and flies exposed to both paraquat and caffeine. Array data were corrected for background signal and normalized. We tested for differences in expression associating with paraquat (in untreated vs. paraquat-only-treated flies), caffeine (in untreated vs. caffeine-only-treated flies), or for paraquat-caffeine interaction (in all four treatment conditions). We found transcripts of two genes with significantly different expression in response to paraquat exposure: *CG31288* ($P_{unadjusted}$=3E-6; $P_{adjusted}$=0.046) and *sugarbabe* ($P_{unadjusted}$=5E-6; $P_{adjusted}$=0.046). We also found transcripts of two genes with significantly different expression upon caffeine exposure: *Hr38* ($P_{unadjusted}$=3E-6; $P_{adjusted}$=0.04) and *CG3016* ($P_{unadjusted}$=4E-6; $P_{adjusted}$=0.04). No transcripts showed significant evidence for paraquat-caffeine interaction on expression when adjusted for genome-wide tests. Human orthologs for these Drosophila genes include *Nr4a2* (with *Hr38*) and *USP30* (ubiquitin specific peptidase 30, with *CG3016*). Both *Nr4a2* (aka *Nurr1*) and ubiquitin-related genes have been previously linked to PD in humans, making these genes interesting candidates for further study.

1012T

Probing the Gaucher/Parkinson Link By Crossing Different Mouse Models. *N. Tayebi, E. Maniwang, R. Tamargo, Y. Blech-Hermoni, N. Moaven, E. Aflaki, B. Berhe, E. Sidransky.* Section on Molecular Neurogenetics, Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

Gaucher disease, the most common lysosomal storage disorder, results from mutations in the glucocerebrosidase gene (GBA1) causing pathologic accumulation of glucosylceramide and glucosylsphingosine. Mutations in GBA1 are identified at an increased frequency in patients with Parkinson disease (PD) and Dementia with Lewy Bodies (DLB), although the mechanism for this association is not fully established. Among the heterozygous GBA1 mutations identified in subjects with PD are the loss-of function mutations, c.84dupG and IVS2+1. To explore how loss-of-function mutations on one allele might impact PD pathogenesis, we attempted to model this situation in mice using null allele gba carrier knock-out mice (gba-/wt) crossed with transgenic (Tg) mice over-expressing mutant A53T human alpha-synuclein (α-syn).Wild type (wt/wt) gba-/wt, TgA53Tα-syn/wt and TgA53Tα-syn/Tgα-syn mice were used as controls. The α-syn copy number in the Tg mice was estimated using Real-Time PCR. After breeding, gba and TgA53Tα-syn genotypes were confirmed, and nine sets of mice were followed over two years, monitoring for weight and neurological symptoms. Weight loss was a common initial sign, followed by arched back, impaired axial rotation, and finally lack of any movement. Mice with progressive neurologic deterioration were euthanized with their respective controls, and midbrain, frontal cortex, and spinal cord were collected. Gba expression, glucocerebrosidase (GCase) levels and enzyme activity were determined. Over-expression of α-syn was confirmed by western blot. Levels of RNA expression and GCase protein in the brain samples were similar in all mice studied. However, GCase activity was lower in the gba- /TgA53Tα-syn mice compared to gba-/wt mice. Brain glucosylsphingosine levels and neuropathology are also being evaluated. In a second set of animals with the same genotypes, 12 month old mice were treated with thioglycolate injections intraperitoneally for one week. The mice were then sacrificed and peritoneal macrophages were collected. Here too GCase activity was significantly lower in the gba- /TgA53Tα-syn macrophages compared to those from gba-/wt mice. These results suggest that over-expression of mutant α-syn could impair lysosomal function or interrupt the interaction between GCase and its activator, SAP-C. We postulate that the loss of function of one GBA1 allele may alter lysosomal function, and contribute to neuronal degeneration.

1013F

eQTL mapping of genes differentially expressed in the lung of COPD patients reveals new genes associated with disease susceptibility. *M. Lamontagne[1], C. Couture[1], M. Laviolette[1], Y. Bossé[1,2], The Merck-Laval-UBC-Groningen.* 1) Inst Univ de Cardiologie et de Pneumologie de Québec, Quebec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada.

Introduction: COPD is a complex disease characterized by a progressive airflow obstruction that is not fully reversible. Cigarette smoking is responsible for 85 to 90% of COPD cases, but only 15 to 20% of the smokers develop the disease. A genetic component to develop COPD is thus persuasive. Methods: Genome-wide gene expression profiles of 500 non-tumor lung specimens were obtained from patients undergoing lung surgery. Blood-DNA from the same patients were genotyped for 1,2 million SNPs. Following quality control filters, 409 samples were analyzed in three steps. First, analyses were performed to identify transcripts differentially expressed between patients with and without COPD. Second, expression quantitative trait loci (eQTL) mapping was performed on these transcripts. Finally, eQTL SNPs were tested for association with COPD. Results: 33 transcripts mapping 28 unique genes were differently expressed between cases and controls (p-value < 1e-04). 160 eQTLs (145 SNPs) were associated with 19 of the 33 transcripts (p-value < 1e-06). 57 of the 145 eQTL SNPs were nominally associated with COPD (p-value < 0.05). The 57 disease-associated variants were correlated with five differently expressed genes (BRAP, DAAM1, SEPT3, SOST, SPINK5). Disease-associated eQTL SNPs for BRAP, SEPT3, SOST, SPINK5 were trans-acting variants. Conclusion: By combining genome-wide gene expression in the lung, eQTL mapping and genetic association in a case-control series of patients with COPD, we identified five new genes that are likely to play a role in the pathogenesis of COPD. This study confirms the importance of well-powered eQTL mapping design to find disease-related trans variants that cannot be identified by GWAS alone.

1014W

Generation of a cre recombinase-conditional *Nos1ap* overexpression transgenic mouse for modeling QT interval variation. *D.R. Auer[1], P. Sysa-Shah[2], D. Bedja[2], J. Simmers[1], E. Pak[3], A. Dutra[3], R. Cohn[1], K.L. Gabrielson[2], A. Chakravarti[1], A. Kapoor[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD; 3) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Polymorphic noncoding variation at the *NOS1AP* locus is the major genetic regulator of the human electrocardiographic QT interval explaining ~2% of the phenotypic variation, is associated with the risk for sudden cardiac death (hazard ratio ~ 1.4) and is a potent genetic modifier (increasing risk up to 10X) of cardiac phenotypes in patients with Mendelian long QT syndrome. *in vitro* and *ex vivo* gene targeting studies by our group have indicated that gene expression levels of *NOS1AP* alter cardiac electrophysiology through variation at an enhancer element. To further enhance the understanding of *NOS1AP* function in physiology and cardiac disease, we report here the generation and characterization of cre recombinase conditional *Nos1ap* overexpression transgenic mice (*Nos1ap*^{Tg}). Conditional transgenic mice were generated by pronuclear injection and three independent single site, multiple copy integration event-based founder lines were selected. For heart-restricted overexpression, *Nos1ap*^{Tg} mice were crossed with *Mlc2v-cre* and *Nos1ap*^{Tg}; *Mlc2v-cre* F1 mice were evaluated for transcript levels, electro- and echo-cardiographic parameters. Surface electrocardiographic recordings were performed at baseline and 20 minutes after injection with the sodium channel blocker flecainide acetate at dosages of 20 or 30 milligrams of drug per kilogram body weight. *Nos1ap* transcript over-expression was observed in left ventricles from *Nos1ap*^{Tg}; *Mlc2v-cre* mice. However, no significant difference was observed between *Nos1ap*^{Tg}; *Mlc2v-cre* mice and wild type control littermates in electro- and echocardiographic measurements. Steady-state Nos1ap protein levels, as assessed by Western blotting, in left ventricles from *Nos1ap*^{Tg}; *Mlc2v-cre* mice were similar to those in control wild type mice, indicating that Nos1ap protein levels are tightly regulated, at least in cardiac tissue. We hypothesize that the discordance between *Nos1ap* transcript and protein expression leads to absence of phenotypic consequence in *Nos1ap*^{Tg}; *Mlc2v-cre* mice.

1015T

Generation and characterization of NLGN2 R215H knock-in mice as a model of schizophrenia. *C.H. Chen[1], H.M. Liao[2], C.C. Yeh[3], Y.L. Chao[4].* 1) Dept. of Psychiatry, Chang Gung Memorial Hospital at Linkou, Gueishan, Taiwan; 2) Dept. of Psychiatry, Taiwan University Hospital, Taipei, Taiwan; 3) Graduate Institue of Life Sciences, National Defense Medical Center, Taipei, Taiwan; 4) Dept. of Psychiatry, Tzu-Chi General Hospital, Hualien, Taiwan.

Schizophrenia is a severe chronic mental disorder with a high genetic component in its etiology. Several mutations of synaptic genes have been found to be associated with schizophrenia in the literature, including neuroligin gene family. In our previous study, we identified a rare missense mutation R215H of the neuroligin-2 gene (NLGN2) in a patient with schizophrenia. The missense mutation was shown to be a loss-of-function mutant in inducing GABAergic synaptogenesis in cell-based assay (Hum Mol Genet 201;20(15)3042-51) . To further characterize the pathogenesis of this mutation in schizophrenia in vivo, we generated a line of R215H knock-in mice using homologous recombinant gene transferring technology. Immunohistochemistry and Western blot analysis of the animal showed significant reduction of neuroligin-2 protein in the cerebral cortex, hippocampus, and cerebellum of the transgenic mice. A series of behavioral tests showed that the male transgenic mice had significant increased pre-pulse inhibition (PPI) compared to the wild type animal. This enhanced PPI was not observed in female transgenic mice. Morris water maze test showed no impaired memory in both male and female transgenic animals compared to the wild type animal. Furthermore, enhanced social interaction activities and contact time were observed in both male and female transgenic mice compared to the wild type. The transgenic animals seemed to be resistant to the methamphetamine-induced impaired PPI, but more susceptible to alcohol-induced impaired PPI. Several functional and behavioral characterization tests are under way. Our preliminary data suggest that the R215H transgenic mice may have impaired GABAergic interneuron function in their brain, and the behavioral characteristics of the animal might to some extent mimic the positive symptoms rather than negative symptoms of schizophrenia, which is consistent with the clinical symptoms of the affected patient who carried this mutation.

1016F

Identifying pathway-level pleiotropic effects in pancreatic cancer and Crohn's disease. *K. Aquino-Michaels[1], E.R. Gamazon[1], P. Evans[1], V. Trubetskoy[1], F. Innocenti[3], K. Owzar[4], H.K. Im[2], N.J. Cox[1].* 1) Department of Medicine, University of Chicago, Chicago, IL; 2) Department of Health Studies, University of Chicago, Chicago, IL; 3) Institute for Pharmacogenomics and Individualized Therapy, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Duke Department of Biostatistics & Bioinformatics, Duke University School of Medicine, Durham, NC.

Inflammation has long been considered the underlying basis to several complex traits and plays well-established roles in elevating risk to various cancer types. Given the clear heritability of complex traits, we hypothesized that there may be pleiotropic loci for both inflammatory diseases and cancer. To test this hypothesis, we used the SNP effects from the Wellcome Trust Case Control Consortium (WTCCC) Crohn's disease (CD) dataset to generate a polygenic CD liability score which we correlated to case control status in a set of 304 CALGB 80303 pancreatic cancer (PC) patients and 1425 controls. We report significant predictive ability of the Crohn's disease liability score in pancreatic cancer (area under the receiver operating characteristic (ROC) curve of 0.63). To better understand the biology driving this inflammatory signature built from CD, we developed SNP-pathways by annotating SNPs from the SCAN Database (scandb.org) and Regulome Database (regulomedb.org) to canonical pathways from the Molecular Signatures Database (broadinstitute.org/gsea/msigdb) based on their association to transcript levels or overlap with transcription factor binding regions. Using this annotation approach, a given SNP-pathway contained typically between 2-8% of the genome-wide SNP set. We computed liability scores (containing only subsets of β-values) for each SNP-pathway and assessed the correlation between a pathway specific score and PC case control status. Top pathways yield comparable area under the curve (AUC) to the full set (0.58-0.60 versus 0.63) and are all characterized as immune mediated.

1017W

Cell-specific enrichment metrics for overlap of signals from GWAS with DNase hypersensitivity sites. *D. Chasman, F. Giulianini on behalf of the CKDGen and BP Metabochip-ICBP consortia.* Division of Preventive Medicine, Brigham & Women's Hospital, Boston, MA.

SNP associations from GWAS are enriched in non-coding regions mapping to chromatin marks. Quantitative characterization of this enrichment by cell type may identify causal SNPs and help reveal the physiologic basis of a phenotype of interest. We propose quantifying enrichment as an odds-ratio (OR) comparing the odds of a phenotype's GWAS SNPs overlapping the chromatin marks of a particular cell with a reference odds. The reference odds may pertain either to reference SNPs mapping to chromatin marks of the same cell or to the same phenotype's SNPs mapping to chromatin marks of a reference cell. We also considered a related notion of specificity quantified by relative entropy (H), the amount of information in bits needed to distinguish the enrichment in chromatin marks of phenotype SNPs compared with reference SNPs. We explored these metrics in publicly available data from 123 cell types, in which 1.1-6.7% of autosomal HapMap SNPs used by GWAS studies map to DNase hypersensitivity sites (DHSs). Applied to the NHGRI GWAS catalog, the enrichment measures were consistent with known biology. For example, across all 123 cell types and using reference SNPs from the GWAS catalog excluding the phenotype of interest, breast cancer SNPs were most enriched in DHSs from mammary ductal carcinoma epithelial cells (62% SNPs, OR= 4.4, p=0.012, H=0.38) while SNPs for celiac disease or rheumatoid arthritis were most enriched in DHSs from T-regulatory cells (67-72% SNPs, OR=3.4-4.9; p=0.002-0.04, H=0.23-0.38). In a new scan of eGFR (i.e. kidney function) among a total of >130,000 samples, genome-wide significant SNPs were most enriched in DHSs from renal proximal tubule and cortical epithelial cells with ORs up to 3.7 -7.1 (45-72% SNPs, p=0.02-0.001, H=0.30-0.64). Similarly, genome-wide significant SNPs from a new combined metabochip+GWAS scan for blood pressure (BP) in up to 201,200 samples were most enriched in DHSs from microvascular endothelial cells with ORs up to 6.8-7.4 (75% SNPs, p=0.01-0.009, H= 0.60-0.66). For both eGFR and BP, SNPs with larger association p-values showed less cell-type specific enrichment in DHSs compared with reference SNPs from the GWAS catalog. This trend may be interpreted as evidence for a lower proportion of true associations or associations with pleiotropic effects shared with other phenotypes among less significant SNPs. The proposed framework may be extended to other choices for reference SNPs or chromatin marks, e.g. H3K4me3 sites.

1018T

Down-regulation of the acetyl CoA metabolic network in visceral and subcutaneous adipose tissue of diabetic but not healthy obese individuals. *H. Dharuri[1], P.A.C. 't Hoen[1,5], J.B. van Klinken[1], P. Henneman[2], J.F.J. Laros[1,4,5], M. Lips[3], H. Pijl[3], K. Willems van Dijk[1,3], V. van Harmelen[1].* 1) Department of Human Genetics, Leiden University Medical Center, Leiden, RC, Netherlands; 2) Department of Clinical Genetics, DNA Diagnostics Laboratory, University of Amsterdam, Amsterdam, The Netherlands; 3) Department of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands; 4) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands; 5) Netherlands Bioinformatics Center, Nijmegen, The Netherlands.

Obesity, a growing world-wide epidemic, is associated with decreased life expectancy due to associated metabolic and cardiovascular disorders. The expanded adipose tissue is thought to serve as the pathogenic link between obesity and type-2 diabetes. Furthermore, it has been shown that the accumulation of visceral adipose tissue (VAT) poses a greater metabolic risk than the subcutaneous adipose tissue (SAT). While a majority of obese individuals develop insulin resistance and type-2 diabetes, some remain metabolically healthy. We hypothesize that the biochemical mechanisms that underlie the function of adipose tissue can help explain the difference between healthy and diabetic obese subjects. To address this hypothesis, we utilized RNA-sequencing to analyze the transcriptome of samples extracted from VAT and SAT of obese female subjects undergoing bariatric surgery. A differential gene expression analysis between diabetic and non-diabetic individuals within the cohort pointed to important differences both in the VAT and SAT. Bioinformatic visualization techniques implicated acetyl coA metabolism. A workflow management software called Taverna was used to generate a gene set for acetyl coA reaction network using the KEGG database. Gene set analysis using the global test R package pointed to a statistically significant association with the phenotype. Multiple genes in the immediate vicinity of acetyl coA in the reaction network are down-regulated in diabetic obese subjects. To ascertain if the down-regulation of these genes is correlated to health status, we compared the pre and post-surgery (3 months after) expression levels of these genes in the SAT by qPCR. At this time the majority of diabetic obese women had a significantly improved metabolic health status. We observed a statistically significant up-regulation of acetyl-CoA acetyltransferase 1 (ACAT1), acetyl-CoA carboxylase alpha (ACACA), aldehyde dehydrogenase 6 family, member A1 (ALDH6A1), and methylenetetrahydrofolate dehydrogenase (MTHFD1) post-surgery in diabetic subjects, and the expression level of these genes increased to those observed in non-diabetic subjects. These represent novel loci associated with type-2 diabetes and recovery in the VAT and SAT of obese subjects. Moreover, our data suggest that acetyl coA metabolic dysregulation plays a role in the pathophysiology of type-2 diabetes in obese subjects.

1019F

Primary Progressive versus Bout Onset Multiple Sclerosis: GWAS and pathway-network analysis. *G. Giacalone[1], F. Clarelli[1], P. Brambilla[1], A. Osiceanu[1], M. Sorosina[1], V. Martinelli[1], M. Leone[2], S. D'Alfonso[2], G. Comi[1], F. Martinelli Boneschi[1].* 1) San Raffaele Scientific Institute, Milan, Italy; 2) University of Eastern Piedmont, Novara, Italy.

Background: Genetic contribution to the two different Multiple Sclerosis (MS) courses, Primary Progressive MS (PPMS) and Bout Onset MS (BOMS), has previously been hypothesized by epidemiological studies, but so far no single genetic marker association has been found. Aim: to identify single nucleotide polymorphisms (SNPs) associated to MS course by comparing PPMS and BOMS through a genome-wide association study (GWAS) and a pathway-network analysis approach. Methods: a total of 444 PPMS and 541 BOMS patients genotyped on two different platforms were merged on common autosomal SNPs (296,589). After applying standard quality controls, logistic regression analysis was carried out including age, age at onset, gender and the first principal component as covariates. A pipeline of multiple-pathway and network analysis was developed to move from SNPs, through gene-wise p-values (p), to biological enriched terms/networks and potential disease-modifier genes. In the pipeline we used Gene Ontology and KEGG databases and the following bioinformatic tools: proxyGeneLD, MetaCoreTM, GOSTAT, Genecodis, Pathway-Express. Results: ProxyGeneLD program mapped 296,589 SNPs to 16,583 genes according to linkage-disequilibrium structure (HapMap CEU II 23a/hg18) and computed gene-wise adjusted p. Genes with adjusted p <0.05 (n=958) were prioritized for those directly interacting with each other (MetacoreTM protein interactions database) and the final list (n=218) submitted to multiple-pathway and network analysis. The first 3 hubs of the network composed of 218 genes were HIF1A, ETS1, NOTCH1, known to be involved in angiogenesis, immunity and myelination. Multiple-pathway analysis produced 32 significant terms, the majority of which related to immune functions ('B cell mediated immunity', 'T cell receptor signaling', 'Leukocyte transendothelial migration', 'Cytokine production'), 'Canonical Wnt receptor signaling', 'L-glutamate transport'. Network analysis produced 11 significant networks, some of them enriched also in immune functions and Wnt signaling. Conclusions: in the comparison of PPMS vs BOMS patients, pathway and network analyses approach suggest a possible role of genes related to immune functions, Wnt signaling and myelination processes, which should be further investigated in additional studies.

1020W

Genome-wide analyses of TCF7L2 DNA occupancy across multiple cell lines point to genetic networks underpinning complex traits. *M.E. Johnson[1], J. Zhao[1], J. Schug[2], S. Deliard[1], Q. Xia[1], V. Guy[1], J. Sainz[3], K.H. Kaestner[2], A.D. Wells[4], S.F.A. Grant[1,2,5].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics and Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), Spanish National Research Council (CSIC), Santander, Spain; 4) Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Variation at the *TCF7L2* locus is strongly implicated in the pathogenesis of type 2 diabetes (T2D) and cancer. We previously mapped the genomic regions bound by TCF7L2 using ChIP-seq in the colorectal carcinoma cell line, HCT116, revealing a highly significant over-representation of GWAS loci associated with T2D and cardiovascular disease among the TCF7L2-bound targets. We carried out ChIP-seq in HepG2 hepatoma cells, where TCF7L2 is abundantly expressed, and leveraged these data against six additional cell lines for which genome-wide TCF7L2 binding is now available from ENCODE. Analyzing each cell line in turn, we found strong evidence for enrichment of endocrine (including T2D), cardiovascular and cancer GWAS categories. Ingenuity pathway analysis detected enrichment of the 'T2D mellitus signaling' category in TCF7L2-bound genes in all cell lines except in HepG2 and HEK293, suggesting that neither the liver nor the kidney is the primary site of TCF7L2 influence on type 2 diabetes pathogenesis. In addition, we observed consistent under-representation of factors involved in beta-cell related functions within this 'T2D mellitus signaling' category but over-representation of members of the AKT pathway in other tissues within the same category. Finally, the lists of genes within key pathways that are consistently bound by TCF7L2 across multiple cell lines were also highly significantly enriched for genes implicated by GWAS. As such, our data strongly suggests that knowledge of TCF7L2 occupancy gives us insight in to multiple genetic vulnerabilities in important pathways influencing the risk of several complex diseases.

1021T

Pathway analysis reveals new insights to genetic risk for multiple sclerosis among smokers. *H. Quach[1], F.B.S. Briggs[1], L. Shen[2], C. Schaefer[2,3], L.F. Barcellos[1,2].* 1) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Division of Research, Kaiser Permanente, Oakland, CA; 3) Research Program in Genes, Environment, and Health (RPGEH), Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS) is a complex neuro-inflammatory autoimmune disease. Despite progress through genome-wide association studies (GWAS), identifying the complete heritable risk component has proven challenging. There has been less success determining the environmental contribution, though tobacco smoke (TS) is a definitive risk factor. Collectively, much of the MS risk remains unknown and prior studies have not explored genetic heterogeneity in an environmental exposure context. Therefore, we sought to identify the biological mechanisms mediating pathogenicity among tobacco smokers. A multi-stage investigation was conducted using a large non-Hispanic White MS study in Northern California (1,084 cases, 10,832 controls), genotyped on the Illumina Infinium 660K BeadChip or Affymetrix Axiom array. Detailed TS exposure was available; MS cases who smoked before onset were considered smokers. Two GWAS were conducted: smokers (46% of cases, 40% of controls) and nonsmokers. Stringent quality control was imposed; 4571 and 4543 genes were associated ($p<0.05$) with MS risk among smokers and nonsmokers, compared to control groups, respectively. 2268 genes were specific to risk among smokers, and these loci underwent functional classification (DAVID 6.7). Using Gene Ontology, genes participating in ion transport (fold enrichment [FE]=1.4, p=0.0005) and acute inflammatory response (FE=2, p=0.006) were overrepresented. There was an enrichment of genes in several KEGG and PANTHER pathways: metabotropic glutamate receptor group III (FE=2, p=0.001); complement and coagulation cascades (FE=2, p=0.01); butanoate metabolism (FE= 2.4, p=0.02); nicotinic acetylcholine receptor signaling (FE=1.8, p=0.03) and *WNT* signaling pathway (FE=1.5, p=0.03). These pathways have not been previously implicated in MS, but play a role in physiological responses to TS. For example: TS negatively influences the coagulation cascade and activates thrombi formation; *WNT4* exacerbates TS induced proinflammatory cytokine release in bronchial endothelium. Recent work in MS murine models shows that upon local lung stimulation, autoreactive T cells migrate to the CNS with the aid of *Ninj1* receptors and induce inflammation. Interestingly, a missense *NINJ1* SNP was significantly associated with MS risk among smokers only (rs2275848: OR=1.4, $p<0.01$; PolyPhen: possibly damaging). Thus, our results suggest variation within genes related to physiological response to TS contributes to MS onset among smokers.

1022F

Regulatory networks reveal directed pathways enriched in disease-associated SNVs in complex diseases. *G. Quon[1,2], S. Feizi[1,2,3], D. Marbach[1], M. Mendoza[1,4], M. Kellis[1,2].* 1) CSAIL, MIT, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Research Laboratory of Electronics, MIT, Cambridge, MA; 4) Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

A major challenge in understanding the role of regulatory variation in human disease is to identify regulatory pathways that may be disrupted by single nucleotide variants (SNVs) located in regulatory elements in the genome. To address this challenge, we have identified candidate regulatory elements and inferred a directed human regulatory network using the ENCODE and Roadmap Epigenomics datasets. We then developed a probabilistic model that takes as input the aforementioned network and regulatory element locations, as well as the SNV association P-values from genome-wide association studies (GWAS). The model identifies SNVs highly ranked (by P-value) whose target genes form surprisingly long, or dense, pathways in the regulatory network, and computes posterior probabilities for disease relevance for each SNV and gene in the network.

We applied our model to two GWAS datasets from dbGaP: type 1 diabetes (T1D) and multiple sclerosis (MS). For T1D, 1276 (0.26%) SNVs and 423 (1.8%) genes were predicted as disease-relevant; for MS, 141 (0.044%) SNVs and 232 (1.0%) genes were predicted as disease-relevant. The disease genes formed five modules in T1D and three modules in MS.

For each disease, the model identified one key regulator that is not in *cis* with any disease-associated SNV, but is predicted to be disease-relevant because it connects the disease gene modules together. The MAZ regulator identified in the T1D dataset is an important regulator of the insulin gene: a variable number tandem repeat (VNTR) sequence several hundred base pairs upstream of the insulin transcription start site contains high-affinity binding sites for MAZ, and specific VNTR alleles have protective effects on T1D risk.

The SP3 regulator connects the disease gene modules identified in the MS dataset. Previous work has identified SP3 as substantially depleted in expression in several immune cell types in MS patients relative to control cell populations, consistent with MS as an autoimmune disease.

These results suggest our probabilistic model is a powerful tool for identifying disease-relevant pathways from GWAS data.

1023W

Currarino syndrome - Searching for factors modifying expression of MNX1. *I. Holm[1], T. Monclair[2], B. Stadheim[1], K.L. Eiklid[1].* 1) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 2) Department of Pediatric Surgery, Oslo University Hospital, Oslo, Norway.

Currarino syndrome (CS), (OMIM#176450) is an autosomal dominantly inherited developmental disorder caused by incomplete rostral separation of the endo- and ectoderm. The condition is characterized by three main findings; a sacral bony abnormality, anal stenosis and a presacral mass. CS is associated with mutations in the MNX1 transcription factor gene (previously known as HLXB9) in familial cases. Penetrance is reduced and expression is variable within families. In sporadic cases less than one third of the patients have detectable MNX1 mutations. Clinical signs and symptoms are similar in familial and sporadic patients. We have detected mutations in MNX1 in all affected members of four Norwegian families, but in none of the six sporadic cases. MNX1 is one of many transcription factors working together in regulation of embryonal development. The transcription factor is expressed early in foetal life and is especially important for pancreatic development and development in the sacral region. It is assumed to play a role in neonatal diabetes (Bonnefond et al., 2013) although patients with CS do not have diabetes. We wanted to examine if transcription factors known to interact with MNX1 could lead to defective expression of MNX1 and thereby causing clinical CS features in the non-familiar cases. We searched for sequence variation in both translated and untranslated regions in the transcription factors already known to interfere with MNX1. We have so far looked at SOX17 (OMIM # 610928), SHH (OMIM # 600725), CDX1 (OMIM # 600746), ISL1 (OMIM # 600366) and PDX1 (OMIM # 600733). Preliminary results show no common SNPs shared by all six sporadic patients. We also searched for sequence variation in the highly conserved region upstream of the MNX1 to see if that could influence expression of MNX1 in non-familiar CS patients. aCGH with SNP-microarray were used to look for other regions with factors or genes explaining the symptoms in non-MNX1 patients. The complete result will be presented in the poster.

1024T

Analysis of GTF2IRD1's role in the craniofacial and neurological features of Williams-Beuren syndrome indicates an epigenetic control function. *S.J. Palmer[1], C.P. Canales[1], P. Carmona-Mora[1], F. Tomasetig[1], J. Widagdo[1], A.C.Y. Wong[2], G.D. Housley[2], P. Kaur[3], I. Smyth[4], A.J. Hannan[5], P.W. Gunning[1], E.C. Hardeman[1].* 1) School of Medical Sciences, Cellular and Genetic Medicine Unit, University of New South Wales, Sydney, Australia; 2) School of Medical Sciences, Department of Physiology and Translational Neuroscience Facility, University of New South Wales, Sydney, Australia; 3) Epithelial Stem Cell Biology Laboratory, Peter MacCallum Cancer Centre, Melbourne, Australia; 4) Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia; 5) The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Australia.

Williams-Beuren syndrome (WBS) is a neurodevelopmental disorder resulting from a hemizygous microdeletion within chromosome 7q11.23 involving up to 28 genes. Genotype-phenotype correlations in patients with atypical deletions have mapped the typical craniofacial dysmorphologies and most aspects of the cognitive profile to a pair of genes that encode the evolutionarily-related transcriptional regulators GTF2IRD1 and GTF2I. The gene *GTF2IRD1* was originally cloned and characterized in our laboratory and we have generated *Gtf2ird1* knockout mouse lines that have phenotypes reflecting aspects of the human disease. Similar to WBS patients, knockout mice have large lips and this phenotype correlates with the pattern of *Gtf2ird1* expression in facial tissue during mouse development. RNA-seq analysis of knockout lip tissue suggests defects of proliferation and differentiation that implicates GTF2IRD1 in the control of patterning and development in lip epidermis. *Gtf2ird1* is also expressed in discrete brain regions and sensory organs that correlate with other abnormalities in the knockout mice and WBS patients, including defects of motor coordination, hearing, exploratory drive and anxiety. Gene expression profiling in the corpus striatum of *Gtf2ird1* knockout mice shows changes in the activity of a set of immediate-early genes, indicating possible roles in adult neuron maintenance and the control of experience-induced gene activity. To understand how this may work at the molecular level, our biochemical experiments have shown that GTF2IRD1 negatively auto-regulates its own allele though direct DNA binding and utilizes protein interaction surfaces to cooperate with other DNA binding proteins. GTF2IRD1 also interacts with a specific set of chromatin modifying proteins that would explain its role in gene silencing. The GTF2IRD1 protein is subject to tight control of abundance and activity through SUMOylation and ubiquitin-mediated proteasomal degradation. Together, these data support a role for GTF2IRD1 in the neurological and craniofacial features of WBS and suggest that it regulates gene expression via epigenetic mechanisms.

1025F

A Disease Module Captures Novel Candidate Genes and pathways for Asthma. *A. Sharma[1,2], J. Menche[1,2], S. Ghiassian[1,2], A-L. Barabási[1,2,3].* 1) Center for Complex Networks Research and Department of Physics, Northeastern University, Boston, MA 02115, USA; 2) Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA; 3) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

To understand various complex disease mechanisms, we not only require cataloging of disease genes by current high-throughput technologies, but detail of cellular components that are influenced by these genes and gene products. Here, we propose a novel concept of disease module that aims to identify the local neighborhood of cellular network for a particular disease phenotype. We illustrate the disease module detection algorithm by identifying the asthma disease module. Further, we validate the module for functional and pathophysiological relevance, using both computational and experimental approaches. Asthma module provides a deeper knowledge of the potential candidate genes and pathways that drive the pathological processes. On one hand, asthma module is enriched with disease-associated variations and on other hand; the impact of asthma specific drug is limited to the disease module. The emergence of GAB1 signaling pathway with both genetics and therapeutics (glucocorticoid) influence establishes the potential of disease module algorithm for dissecting the complex diseases.

1026W

The generation of sex-specific allergy networks: Validation that allergic disease has sexual dimorphic origins. *J. Lasky-Su[1,2], W. Qiu[1,2], V. Carey[1,2], S. Weiss[1,2], B. Raby[1,2], D. DeMeo[1,2].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA.

Childhood allergy affects over 50% of children and has sexually dimorphic features that are evidenced in early childhood and into adolescence, including a higher prevalence of allergy and atopy in boys. A key characteristic of this sex difference is the consistently higher levels of total serum immunoglobulin E (IgE) observed in males. There is growing evidence that a portion of the underlying basis for the observed sexual dimorphism in allergic disorders is molecular, as heritability estimates for allergic phenotypes such as total serum IgE vary significantly for males and females (83% vs. 63%). Previously we examined the relationship between total serum IgE levels and genome-wide gene expression in 223 CAMP individuals and found striking differences by sex: 1) Sex stratified analyses revealed a significant correlation between IL17RB and IgE in males only ($r2 = 0.19$, p value= 8×10^{-8}); 2) There were 873 compared to only 154 transcripts that mapped to gene IDs for males versus females; 3) Pathway enrichment analysis found nominal levels of enrichment in 13 pathways for males and only 5 pathways for females and none of these pathways overlapped. Given these notable differences between males and females, we constructed sex-specific allergy networks using gene expression data from the 223 CAMP individuals and applying Weighted Correlation Network Analysis (WGCNA). We first constructed sex-specific allergy networks using a general co-expression network based on the Pearson correlation coefficient by measuring the concordance of gene expression measures. These correlation measures are used to define an adjacency matrix, which indicates the degree that two expression measures are connected. We then identified important allergy and sex-specific modules within the networks using hierarchical clustering and topological dissimilarity. Once the sex-specific allergy networks were generated, we identified similarities and differences between the networks using preservation methodologies, first by comparing the sex-specific dendograms and second by calculating summary measures of preservation (median rank and Z summary statistic). We found striking differences between the male and female allergy networks, identifying sex-specific biological pathways that contribute to allergy pathogenesis.

1027T

Gene network analysis of candidate loci for human anorectal malformations. *M. Garcia-Barcelo[1,3,4], E.H.M. WONG[2], C.H. NG[1], V.CH. LUI[1,4], M.T. SO[1], S.S. CHERNY[2,3], P.C. SHAM[2,3,4], P.K. TAM[1,4].* 1) Dept Surgery, Univ Hong Kong, Hong Kong, NA, Hong Kong; 2) Dept Psychiatry, Univ Hong Kong, Hong Kong, NA, Hong Kong; 3) Center for Genomic Sciences, Univ Hong Kong, Hong Kong, NA, Hong Kong; 4) 4Centre for Reproduction, Development, and Growth Univ Hong Kong, Hong Kong, NA, Hong Kong.

Anorectal malformations (ARMs) are birth defects that require surgery and carry significant chronic morbidity. Our genome-wide copy number variation (CNV) study had provided a wealth of candidate loci. To find out whether these candidate loci are related to important developmental pathways, we have performed an extensive literature search coupled with currently available bioinformatics tools. This has allowed us to assign both genic and non-genic CNVs to interrelated pathways known to govern the development of the anorectal region. We have linked 11 candidate genes to the WNT signaling pathway and 17 genes to the cytoskeletal network. Interestingly, candidate genes with similar functions are disrupted by the same type of CNV. The gene network we discovered provides evidence that rare mutations in different interrelated genes may lead to similar phenotypes, accounting for genetic heterogeneity in ARMs. Classification of patients according to the affected pathway and lesion type should eventually improve the diagnosis and the identification of common genes/molecules as therapeutics targets.

1028F

Comparison of gene expression induced by HIV-1 GAG peptides specific to HLA-A*01:01 and B*07:02 in PBMCs by mRNA-seq analysis. *L.R. Liu[1,2], P. LaCap[1], R. Capina[1], B. Liang[1,2], B. Fristensky[1,3], B. Ball[1,2], F. Plummer[1,2], M. Luo[1,2].* 1) HIV and Human Genetics Division, National Microbiology Laboratory. 1015 Arlington Street. Winnipeg, Manitoba. Canada R3E 3R2; 2) Department of Medical Microbiology, University of Manitoba. 745 Bannatyne Avenue. Winnipeg, Manitoba. R3E 0J9; 3) Department of Plant Sciences, University of Manitoba. 222 Agriculture Building University of Manitoba. Winnipeg, Manitoba. R3T 2N2.

Introduction: A subset of sex workers (CSW) enrolled in the Pumwani cohort in Nairobi, Kenya remain HIV negative despite repeated exposure through high risk sex work. Studies on genetic factors enriched within these highly HIV-1 exposed seronegatives (HESN) suggest this natural resistance to HIV-1 is multi-factorial, and associated with specific alleles of Human Leukocyte Antigens (HLA). HLA-A*01:01 was associated with reduced risk of HIV-1 infection, whereas B*07:02 was associated with increased risk. HLAs initiate cell-mediated immunity (CMI) by presenting antigens to T-cells. Systematic comparison of A*01:01 and B*07:02 HIV-1 Gag epitopes showed that A*01:01 recognized fewer epitopes than B*07:02, and recognition of more Gag epitopes is associated with susceptibility to HIV-1 infection. However, It is unclear whether the A*01:01 or B*07:02 and GAG peptide complex could differently induce downstream gene expression, and lead to differential T cell function. We conducted mRNA-seq analysis and compared gene expression induced by A*01:01 and B*07:02 specific peptide using PBMCs of an individual express both A*01:01 and B*07:02. **Approach:** PBMCs from a single blood draw was split into equal proportions, and stimulated with peptide NSSKVSQNY (A*01:01 specific) or SPRTLNAWV (B*07:02 specific). 8-hours post-stimulation, the PBMCs were harvested for RNA-seq analysis. RSEM (RNAseq by Expectation Maximization) workflow was used for read alignment, and transcript quantification. DESeq was used for differential expression analysis. **Results:** After correction for false-discovery, one gene was significantly differentially expressed, DEQ571357 (FDR P-value = 3.8×10^{-12}), and 190 gene-isoforms (FDR P-value ≤ 0.05). **Implications:** This study aims to inform vaccine design by evaluating host gene expression induced by HLA-specific epitopes. Significant differences were identified however the results need to be validated by qRT-PCR and in other A*01:01 and B*07:02 co-expressed individuals.

1029W

Protein altering variants found in ciliary and polarity genes in biliary atresia patients. *E.A. Tsai[1], C.M. Grochowski[2], L.D. Leonard[2], R.P. Matthews[3,4], K.M. Loomes[3,4], B.A. Haber[5], N.B. Spinner[2,6], M. Devoto[3,7,8,9].* 1) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 4) Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Hepatology, Infectious Diseases Clinical Research Department, Merck Research Laboratories, North Wales, PA; 6) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia PA; 7) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 8) Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 9) Department of Molecular Medicine, University La Sapienza, Rome, Italy.

Biliary atresia (BA) is a bile duct disorder that presents within the first few months of life and causes necroinflammatory obliteration of the extrahepatic biliary tree. Children with BA have severe liver disease, and BA is the most frequent indication for pediatric liver transplantation in the United States. BA is thought to result from a combination of genetic and environmental risk factors, but no specific gene responsible for BA has been identified yet. Absence of immotile cilia on the surface of endothelial cells is apparent upon inspection of intrahepatic as well as remnant extrahepatic bile ducts, suggesting a loss of cell polarity. Manipulation of polarity genes leads to biliary defects in model organisms, and some BA patients demonstrate other anomalies consistent with polarity defects. We compiled a list of genes that participate in the establishment or maintenance of cell polarity (n=280) as well as genes with a role in the composition or function of cilia (n=291). We hypothesize that rare or novel damaging mutations in these genes may contribute to the development of BA. We performed exome sequencing with the Agilent SureSelect All Exon V4+UTR capture kit on 30 Caucasian, isolated BA patients. Variant filtration was performed to analyze only variants with frequency ≤5% in the 1000 Genomes Project Phase I and the NHLBI Exome Sequencing Project (ESP). Notably, one male patient had a nonsense variant in the X-linked polarity gene, *ATP6AP1*. This variant, rs201620814, was not observed in 1000 Genomes, but heterozygous variants in 4/1949 females and a hemizygous variant in 1/1283 males in the ESP cohort (<0.1%) was observed. Additionally, mutations in vacuolar H^+-ATPase subunits have been shown to cause biliary defects in zebrafish. Another patient had two missense variants in trans in *DNAL1*, which encodes for part of the outer dynein arm of cilia. We have identified several other changes in polarity and cilia genes that we suspect are contributing to BA in these patients. Our results support the hypothesis that polarity and cilia genes may be important in BA etiology but also suggest a high level of genetic heterogeneity in this disorder.

1030T

Vascular tone pathway in relation to primary open angle: results from the NEIGHBORHOOD consortium. *L. Pasquale[1], S. Loomis[1], J. Kang[2], J. Bailey[3], J. Haines[3], B. Yaspan[3], M. Hauser[4], J. Wiggs[1], NEIGHBORHOOD consortium.* 1) Glaucoma Service, Massachusetts Eye & Ear Infirmary, Boston, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA Boston, MA; 3) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 4) Duke University School of Medicine, Durham, NC.

Background: Reduced ocular perfusion pressure and impaired autoregulation of ocular blood flow have been implicated in the pathogenesis of primary open angle glaucoma (POAG). Genome wide association studies (GWAS) of POAG have identified CAV1/CAV2, which codes for caveolins that are involved in vascular regulation. Biologic pathway analyses can enhance the power to discover biologically meaningful genetic markers associated with complex traits. We investigated whether a set of genetic markers of factors involved in setting vascular tone are associated with POAG. Methods: We used Illumina 660W-Quad array platform genotype data and pooled p-values from 3,108 POAG cases and 3,430 controls from the combined the National Eye Institute Glaucoma Human Genetics Collaboration consortium and the Glaucoma Genes and Environment study. We compiled a set of single nucleotide polymorphisms (SNPs) in 186 genes related to vascular tone that were determined with KEGG pathways or previously identified in GWAS for blood pressure. We analyzed the association between these SNPs and POAG using both a pathway- and gene-based approach with the Pathway Analysis by Randomization Incorporating Structure analysis software package, which performs a permutation algorithm to assess the statistical significance of genes and pathways relative to randomly generated genes and pathways of comparable size, structure and linkage patterns. Results: The vascular tone pathway was not associated with POAG overall or the POAG subtypes defined by type of visual field loss (paracentral [n=224 cases] or peripheral loss [n=993 cases]) (permuted $p \geq 0.20$). In addition, results were similarly null in men and women (permuted $p \geq 0.81$). In gene-based analyses, two genes had nominally significant permuted gene p-values <0.05 in almost all analyses by subtype and gender: CAV1/CAV2 (permuted gene $p \leq 0.01$ for all outcomes), PRKAA1 (permuted gene $p \leq 0.05$ for all outcomes); in addition, 4 of the 8 genes that were associated with overall POAG at p<0.001 were those coding for factors that closely interacted with the caveolins. Conclusions: Although the overall vascular tone SNP set did not show significant associations with POAG, the gene-based analyses confirm the previously identified CAV1/CAV2 gene and suggest a possible role of caveolins and related factors in POAG pathogenesis.

1031F

Polygenetic Risk Model Suggests a Protective Effect of Lipid Genes on Plasma Glucose and HbA1c levels. *J. Fu[1], N. Li[2], M. van der Sijder[1], C. Wijmenga[1], H. Snieder[3], M.H. Hofker[2].* 1) Genetics Dept, UMCG, University of Groningen, Groningen, Netherlands; 2) Molecular Genetics, UMCG, University of Groningen, Groningen, Netherlands; 3) Epidemiology Dept, UMCG, University of Groningen, Groningen, Netherlands.

Background and hypothesis: Dyslipidemia is strongly correlated with raised glucose levels and type 2 diabetes (T2D), although the causal nature of the correlation and the role of dyslipidemia genes on T2D are still unclear. The advances made by genome-wide studies are allowing the use of a polygenetic risk score to assess the impact of lipid genes on T2D or related traits. Methods: We studied 95 common genetic variants that were robustly associated with triglyceride (TG), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C) or high-density lipoprotein cholesterol (HDL-C). We included 13,107 subjects from the LifeLines cohort and calculated the polygenetic scores of blood lipids per individual. We computed the Spearman correlation between each genetic score and fasting plasma glucose and HbA1c levels. Results: Despite the positive correlation between dyslipidemia and HbA1c, surprisingly the individuals carrying a higher triglyceride genetic risk seemed to have lower HbA1c levels (r=-0.025, p=0.009). Conditional on blood lipid levels, this negative correlation became not only stronger for HbA1c (r=-0.043, $p=2.3\times10\text{-}6$) but also significant for fasting glucose level (r=-0.051, $p=8.6\times10\text{-}8$). We also consistently observed that the higher genetic scores of TC and LDL-C were correlated with lower HbA1c levels ($p=7.2\times10\text{-}4$ and p=0.003, respectively). Conclusions: In the Lifelines cohort, our polygenetic risk model suggests that genes associated with dyslipidemia may have weak protective effects on plasma glucose and HbA1c levels. This protective effect is pleotropic and independent of all types of lipid.

1032W

Molecular intermediate phenotype mapping of IL-6 and TNFα levels reveals genes critical for chronic systemic inflammation. *S.J. Schrodi[1], D. David[2], JK. Meece[2], DA. Vasco[1], JJ. Mazza[3], MH. Brilliant[1], JA. Smith[4].* 1) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 2) Core Laboratory, Marshfield Clinic Research Foundation, Marshfield, WI; 3) Clinical Research Center, Marshfield Clinic Research Foundation, Marshfield, WI; 4) Department of Pediatrics, University of Wisconsin-Madison, Madison, WI.

Background Numerous common diseases exhibit an underlying component of chronic systemic inflammation. Discovery of genetic variants that produce susceptibility to these diseases has many pitfalls, not the least of which lies in the clinical definition of disease which may or may not reflect pathological mechanisms at the molecular level. Hence, redefining phenotypes based on molecular characteristics may elicit more robust and lucid genetic association signals. This study uses a composite metric of the circulating levels of two important and correlated inflammatory cytokines, TNFα and IL-6, in the general population as a molecular intermediate phenotype. Both cytokines are important in the immunobiology of proinflammatory processes and have been directly targeted by biological therapeutics to remediate autoimmune conditions. To identify novel genes involved in regulation of TNFα/IL-6-mediated chronic inflammation, we performed a quantitative trait GWAS on TNFα/IL-6 levels in a genetically homogeneous population in Central WI, largely derived from Bavarian immigrants in the late 1800s. **Materials/Methods** Using the Meso-Scale Discovery platform, we measured levels of TNFα and IL-6 in plasma from 1015 individuals not taking immune-modulating therapies or having evidence of acute infection. A small number of genetic background outliers were removed from the sample set. The Illumina 660W-QUAD array was applied to all samples and a linear correlation test was applied for the genotypes at each of >500,000 SNPs to transformed TNFα/IL-6 concentration data. The VEGAS program was applied to individual SNP data to obtain gene-based association P-values. **Results/Conclusions** From roughly 18,000 genes evaluated, two statistically compelling regions emerged from the analysis: 1) the *RXFP3-SLC45A2* region on the short arm of Chr5 (P<1E-06), and 2) the *GPR31-CCR6* region on the long arm of Chr6 (P<1E-06). *RXFP3* encodes for a relaxin/insulin-like receptor, and may wield anti-fibrotic, and anti-inflammatory actions. *SLC45A2* encodes for a melanocyte differentiation antigen and is intimately involved in melanin synthesis (OMIM). Interestingly, Sellick *et al* (2005) mapped the *GPR31-CCR6* region as conferring strong predisposition to small vessel lymphocytic vasculitis within an extended family. Additionally, *CCR6* encodes for a chemokine receptor expressed on lymphocytes. These results may provide unique insight into the molecular mechanisms behind proinflammatory processes.

1033T

Uncovering the genetic architecture of complex traits using network approaches. *J. Choi[1], C. Cotsapas[1,2,3].* 1) Department of Neurology , Yale School of Medicine, New Haven, CT; 2) Department of Genetics, Yale School of Medicine, New Haven, CT; 3) Medical and Population Genetics, the Broad Institute of Harvard and MIT, Cambridge, MA.

Over the last decade, genome wide association studies (GWAS) have uncovered thousands of genetic variants influencing complex traits affecting all organ systems. However, it has proven difficult to infer which genes these variants are perturbing, so the challenge of uncovering underlying biological mechanisms remains elusive. We do not yet understand the genetic architecture of traits: do variants perturb entire pathways or other groups of interacting genes? How big are these groups? Are many distinct pathways involved or is a single molecular system sufficient to alter traits? We have developed a robust and flexible framework to address these questions by detecting groups of interacting genes perturbed by associated variants. We have previously observed that genes encoded in disease risk loci tend to interact, suggesting that risk variants modulate susceptibility by perturbing gene networks representing pathways [Rossin et al PLoS Genetics 2011]. Here, we project GWAS scores onto protein-protein interaction networks and use robust network clustering algorithms to look for regions of the overall interaction network enriched for disease association signal, indicating the presence of an interacting set of genes modulating susceptibility. Within this framework, we investigate several genetic architecture models by simulation: scenarios where either a single or multiple groups of interacting genes modulate risk versus no aggregation of signal in gene groups. We are able to recover both true positive and false negative associations, increasing the heritability explained by GWAS. Surprisingly, we also find that in models of multiple distinct susceptibility pathways, these are detected as a single large connected gene network. This appears to be a property of gene interaction networks and suggests that recent reports of large sets of interacting genes underlying disease susceptibility are in fact capturing multiple biological pathways. In two disease GWAS meta-analyses (Crohn's disease and multiple sclerosis) we find gene networks of size ~100 enriched for genetic risk, consistent with an architecture of cumulative genetic burden on molecular pathways. We also find that subsets of these components are expressed in different tissues, indicating that both diseases involve multiple pathways, not a single large group of genes. We are currently extending this framework to simultaneously consider both diseases and investigate pleiotropic effects in these networks.

1034F

Topological analysis of heterogeneous domain knowledge networks prioritizes genes associated with complex diseases. *D.S Himmelstein[1], S.E. Baranzini[1,2].* 1) Biological and Medical Informatics, University of California, San Francisco, San Francisco, CA; 2) Department of Neurology, University of California, San Francisco, San Francisco, CA.

Genome Wide Association Studies (GWAS) remain the preeminent strategy for identifying disease-associated variants, although small effect sizes combined with constraints on expanded sampling limit the pace of discovery. Gene prioritization approaches can increase the power of GWAS by increasing the prior probability of association, but generally only incorporate information from a single biological domain. Here we present a method to prioritize disease-associated genes from a heterogenous network which captures the diversity of biological entities and relationships underlying complex disease pathogenesis. We created a network of 11,249 genes/proteins, 30 tissues, and 44 complex diseases. Using publicly available resources, we connected the three entity types: gene-gene relationships indicate protein-protein interactions, gene-tissue relationships indicate tissue-specific gene expression; tissue-disease relationships represent known pathogenesis; and disease-gene relationships indicate reported associations from published GWAS studies. By adapting an existing algorithm, PathPredict, we provide a framework for computing the probability that a gene-disease pair represents a true association. The method calculates the prevalence of different network topologies connecting genes, tissues, and diseases. A logistic regression model identifies influential topological features and enables prediction of gene-disease pairs with an unknown association status. We assessed the performance of our method on six diseases - rheumatoid arthritis, inflammatory bowel disease, multiple sclerosis (MS), ankylosing spondylitis, psoriasis, and systemic lupus erythematosus - with multiple available large-scale GWAS. Our algorithm scored confirmed associations highly, yielding disease-specific AUC in the range of 0.70-0.79. Using our predictions as prior probabilities, we identified candidate genes that have not been associated to date. Relying on independent GWAS for discovery and validation, we highlight REL, JAK2, and TNFAIP3 as likely candidates in MS. The probability of the observed validation rate occurring by random prior estimates is 0.002, underscoring our ability to uncover novel associations. Integrating information across multiple domains through the use of heterogeneous networks significantly improved performance. Topological analysis of heterogeneous networks provides a potentially powerful new platform for discovery in human genetics.

1035W

Novel approach using gene set enrichment based on human genome-wide association study data implicates FXR/RXR activation as a common pathway affecting blood lipids levels and nonalcoholic fatty liver disease. *Y.M. Puentes[1,2], C.C. Powell[1,2], L.M. Yerges-Armstrong[3], M.F. Feitosa[4], L.F. Bielak[5], A.V. Smith[6], T.B. Harris[7], J. Liu[8], S.K. Musani[8], I.B. Borecki[4], P.A. Peyser[5], E.K. Speliotes[1,2], GOLD Consortium.* 1) Department of Bioinformatics and Computational Biology , University of Michigan, Ann Arbor, MI; 2) Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan 48109, USA; 3) Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine , Baltimore, Maryland 21201, USA; 4) Division of Statistical Genomics, Department of Genetics, Washington University, Saint Louis, Missouri 63108, USA; 5) Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor Michigan, 48109, USA; 6) Icelandic Heart Association, Kopavogur IS-201, Iceland; 7) Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA; 8) University of Mississippi Medical Center, Jackson, Mississippi, 39213, USA.

Nonalcoholic fatty liver disease (NAFLD) is rapidly increasing in prevalence and will become the number one cause of liver disease worldwide by 2020. NAFLD is associated with high triglycerides (TG), high low-density lipoprotein cholesterol (LDL-C) levels, and low high-density lipoprotein cholesterol (HDL-C) levels. Both NAFLD and blood lipid levels are genetically influenced and may share a common genetic etiology. We aimed to identify genes and pathways enriched for genetic associations with both blood lipids and NAFLD using human genome wide association study (GWAS) data. We examined whether genome wide, significantly associated lipid SNP sets from publicly available HDL-C, LDL-C and TG GWAS analyses (N=99,000) were enriched for associations with NAFLD (N=7,126, Speliotes 2011). We then used gene set enrichment analysis (GSEA) as implemented in MAGENTA to identify pathways enriched in HDL-C, LDL-C, and TG GWAS analyses. These lipid pathways were then tested for enrichment in the NAFLD GWAS (N=7,126) and these enriched pathways were also tested for replication in an independent NAFLD GWAS (N=3,124, Palmer 2013). We found that TG (P=0.004) and LDL-C (P=0.02) but not HDL-C GWA SNP sets were enriched in NAFLD. We identified 58 pathways that were enriched in lipid GWAS data. Three of these were also enriched in the NAFLD GWAS (N=7,126) and one, FXR/RXR activation, also showed significant enrichment in another NAFLD GWAS (N=3,124). None of the three original NAFLD pathways were enriched for associations in control publically available GWAS analyses of diastolic and systolic blood pressure (N=275,000), body mass index (N= 249,796), and waist to hip ratio (N=77,167) suggesting that the enrichment was specific to NAFLD. Genes associated with NAFLD ($P<0.05$) in FXR/RXR activation fell into three functional categories: (1) VLDL Assembly: *MTTP, APOB, APOC3*, (2) Nuclear Related Processes: *PPAR-α, HNF1α, NR0B2/SHP* and (3) Hepatic Transport: *MRP2, AE2, ABCG8, ABCG5, OAT2.* Using a novel approach, we found that human genetic variation in or near genes involved in FXR/RXR activation affects both blood lipids and NAFLD in humans. These results suggest that genes that play a role in lipoprotein assembly, nuclear receptor biology, and hepatic transport when altered may affect NAFLD and thus could provide possible therapeutic targets for NAFLD prevention or treatment.

1036T

Identification of disease-relevant pathways among exome variants with a gene network based on the semantic similarity of mouse phenotypes.
F. Honti, C. Webber. MRC Functional Genomics Unit, Dept. of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK.

The identification of a functional association among a set of genes whose variants are implicated in the same disease can suggest cellular processes relevant to the disease mechanism. As the function of many genes is not known and their classification to pathways is scant, functional links between genes are often inferred from large-scale 'omics' data.

We have derived functional associations between human genes from diverse data types and examined their disease relevance. Annotation data were processed in the form of semantic similarity, which is a measure of relatedness between two genes as assessed by the similarity of their annotations. We combined the most informative data sources to form a functional-linkage network, weighting the different data types in proportion to their relative accuracy. Applying our framework, we have examined the network properties of sets of genes with *de novo* substitutions implicated in the same disease by exome sequencing studies.

We found that mouse knockout phenotypes were the most informative predictors of functional links between human genes. Accordingly, genes mutated in the same disease clustered most significantly in a gene network built on mouse phenotypes and in the integrated functional-linkage network, while genes mutated in controls did not form clusters. Combining three sets of genes implicated in autism by recent exome sequencing studies, we have observed distinct functional sub-clusters in the network, suggesting relevant biological pathways.

Functional associations between genes are often inferred from omics data, but the disease relevance of such data types, including protein-protein interactions and gene co-expression networks, has been unclear. The application of semantic similarity has enabled us to compare diverse data types and establish the value of mouse phenotypes. We present a framework for the identification of biological pathways disrupted in complex disorders and demonstrate the ability of this approach to functionally dissect the molecular variants underlying autism.

1037F

Gene-set enrichment analysis of gene expression associations with acute asthma control hints at candidate drug pathways.
D.C. Croteau-Chonka[1], W. Qiu[1], V.J. Carey[1], S.T. Weiss[1,2], B.A. Raby[1,3], Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE) Consortium. 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Partners Center for Personalized Genetic Medicine, Partners Health Care, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Maintenance of effective asthma control is critical for preventing exacerbations, or sudden declines in lung function. Loss of control results in considerable economic costs due to chronic treatment, acute hospitalization, and/or absence from work or school. By better understanding the biological underpinnings of acute asthma control (AAC) we can develop improved clinical and pharmaceutical approaches for maintaining effective disease stability.

To understand the relationships of gene expression with AAC, we performed differential expression analyses in a subset of 583 subjects from the Asthma BioRepository for Integrative Genomic Exploration (BRIDGE) cohort (62% male, mean age = 8.9 years) who had genome-wide gene expression data derived from whole blood RNA extracts and measured with the HumanHT-12 v4 Expression BeadChip. AAC was represented by a modified version of the clinically validated Asthma Control Test (mACT) score. After non-specific filtering, 10,701 probes (one per gene) were tested for linear associations with the normally distributed mACT, adjusting for batch effects using the first ten principal components of gene expression. A total of 548 probes were nominally associated with mACT ($P < 0.05$), but none were associated at a false-discovery rate (FDR) < 0.05.

To gain further biological insights, we performed a gene-set enrichment analysis (GSEA) of the ranked associations using the GSEA2 software and the 'C2' collection of 3,112 curated gene sets from the Molecular Signatures Database (MSigDB). We observed significant enrichment (FDR < 0.25) of 288 MSigDB gene sets among genes whose increased expression was associated with below-median mACT and two gene sets associated with above-median mACT. At an FDR < 0.05, the counts were 41 and zero, respectively (mean gene set size = 62). Enriched gene sets included ones related to expression in immune cells (e.g. plasma cells, eosinophils, B lymphocytes) and to changes in expression in response to non-chemotherapeutic drugs (e.g., glucocorticoids, prostaglandin E2, methoxyestradiol).

These preliminary results suggest the mACT phenotype is a biologically meaningful representation of AAC and hint at potential candidate pathways for drug treatment development. This work was supported by NHLBI grants RC2 HL101543, R01 HL086601-6.

1038W

Gene set signature of leprosy reversal reaction type 1.
M. Orlova[1], A. Cobat[1, 2], N. Thu Huong[3], N.N. Ba[3], N.V. Thuc[3], J. Spencer[4], Y. Nédélec[5], L. Barreiro[5], V.H. Thai[3], L. Abel[6,7,8], A. Alcaïs[6,7,8,9], E. Schurr[1,2]. 1) The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Departments of Human Genetics and Medicine, McGill University, Montreal, Quebec, Canada; 3) Hospital for Dermato-Venereology, Ho Chi Minh City, Vietnam; 4) Department of Microbiology, Immunology and Pathology, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO, 80523-1619, USA; 5) Department of Pediatrics, Sainte-Justine Hospital Research Centre, University of Montreal, Montreal, Quebec, H3T 1C5, Canada; 6) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale, U980, Paris, France; 7) University Paris Descartes, Imagine Institute, Paris, 75993, France; 8) St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 11065, USA; 9) URC-CIC, Hopital Tarnier, Paris, 75006, France.

BACKGROUND. While effective drug treatment of leprosy is widely available, some leprosy patients suffer from major nerve damage due to leprosy type 1 reversal reactions (T1R). T1R are sudden episodes of exacerbated local delayed-type hypersensitivity to *Mycobacterium leprae* in skin and/or nerves. The incidence of T1R varies widely from 6% to 67% of leprosy patients in different leprosy endemic settings. T1R is effectively treated by corticosteroids, yet neurological impairment persists in about one third to a half of all cases. Thus, early identification of patients at risk of T1R, and consequently at risk of neurological injury, is a major challenge in leprosy care. METHODS. We used whole blood cultures stimulated with high doses of *M. leprae* sonicate to assay transcriptional responses of leprosy patients. To avoid the problem of highly dysregulated inflammatory responses during acute episodes of T1R we enrolled cured leprosy patients who had undergone T1R more than five years before and compared their transcriptional response to *M. leprae* antigens with the one displayed by former leprosy patients that had remained T1R-free (retrospective arm). The gene sets that were preferentially regulated in former T1R patients were then validated in a cohort of 43 newly diagnosed leprosy patients of which 11 developed T1R over a three year follow-up (prospective arm). Transcriptome responses to *M. leprae* antigens at the time of enrolment were compared between patients who developed T1R and those who did not. RESULTS. We defined and replicated prospectively a T1R signature set of 44 differentially regulated genes. Three functional groups represented the majority of the T1R set genes: a) pro-inflammatory regulators; b) arachidonic acid metabolism mediators and c) regulators of anti-inflammation. The validity of the T1R gene set signature in the prospective arm directly demonstrated that T1R patients have an innate defect in the regulation of inflammatory responses to *M. leprae*. The T1R gene set appeared as molecular characteristic of T1R susceptibility and was consistent with a breakdown of communication between pro- and anti-inflammatory responses in T1R patients. The identification of the T1R gene set represents a critical first step towards a genetic profile of leprosy patients who are at increased risk of T1R and concomitant nerve damage.

1039T

Characterizing variation in inter-individual immune response by asthma and allergic disease status in a founder population. *M. Stein[1], C. Hrusch[2], G. Alkorta-Aranburu[1], K.A. Bailey[1], M. Çaliskan[1], C. Igartua[1], C. Kagan[1], K. Ross[1], E.E. Thompson[1], A. Sperling[2], C. Ober[1].* 1) Department of Human Genetics, University of Chicago , Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago IL.

Asthma is a common, heterogeneous disease of the airways, characterized by inflammation and dysregulated immune responses to inhaled allergens and early life environmental exposures. The etiology of asthma and atopy is complex, but involves both genetic and environmental factors, as well as their interactions. In particular, early life exposure to high levels of endotoxin can protect again the subsequent development of asthma and atopy in childhood. We initiated this study to characterize the genetic regulation of transcriptional response to endotoxin (lipopolysaccharide, LPS), a component of the gram-negative bacterial cell wall that stimulates the innate immune system via the TLR4/CD14 pathway. As a first step, we characterized transcriptional responses to LPS using a closed cell culture system in which 1 ml whole blood from each of 40 Hutterite school children (ages 6-14) was exposed to either 0.1/ml ug LPS + media or to media alone (null control). 15% of the children in this study were diagnosed with asthma and 30% were atopic (defined by at least 1 positive specific IgE); 12.5% of children were both asthmatic and atopic. RNA from LPS-treated and untreated samples was hybridized to the Illumina HT12v4 microarray platform. 14,442 probes, corresponding to 11,317 genes, were detected as expressed with $p<0.05$ in at least 20% of samples. At a false discovery rate (FDR) of 10%, we observed significant responses to LPS for 6,021 genes in non-asthmatic children and 909 in asthmatic children. 112 of those genes showed response to LPS only in asthmatic children, and was enriched for genes in the RIG1-like receptor pathway ($p=0.019$) (IPA, Ingenuity Systems, www.ingenuity.com). At the same FDR, we observed significant responses to LPS for 5,441 and 4,034 genes in non-atopic and atopic children, respectively. 841 of those genes showed significant response in atopic children only, and were enriched for genes in immunological disease pathways ($p= 0.0015$), inflammatory response pathways ($p=0.039$), and pathways in pattern recognition receptors of bacteria and viruses ($p=0.025$). Overall, peripheral blood leukocyte response to LPS identified differences in transcriptional response in individuals with asthma or atopy that may underlie some of the clinical features of these conditions. This work was supported by NIH grant HL085197.

1040F

Gene Network Analysis with GWAS Data Identifies Novel Lung Function Gene Set. *J. Li[1], B.A. Ong[2,3], J.M. McDonough[2], Z. Wei[4], C. Kim[1], R. Chiavacci[1], F. Mentch[1], J.B. Caboot[5], J. Spergel[6,7], J.L. Allen[2,7], P.M.A. Sleiman[1,7,8], H. Hakonarson[1,7,8].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Pulmonary Medicine and Cystic Fibrosis Center, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Pediatric Pulmonology, Tripler Army Medical Center, 1 Jarrett White Rd, Honolulu, HI; 4) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 5) Division of Pediatric Pulmonology, Madigan Army Medical Center, Tacoma, WA; 6) Center for Pediatric Eosinophilic Disorders, The Children's Hospital of Philadelphia, Philadelphia, PA; 7) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA; 8) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Pulmonary function reflects various pulmonary conditions and predicts morbidity and mortality in adults. While several studies have already been published on pulmonary function traits, these have mainly been carried out in adults of Caucasian ancestry leaving the genetic determinants of pulmonary function traits in the pediatric population and in African Americans, in particular, relatively unexplored. To identify novel genetic determinants of pediatric lung function, we conducted a genome-wide association study (GWAS) of four pulmonary function traits, including FVC, FEV1 FEV1/FVC and $FEF_{25-75\%}$ in 1556 children. Though no SNP reached genome-wide significance, we identified SNPs with notable trend towards association with the pulmonary function measures in both Caucasian and African American cohorts. We also replicated the previously reported *INTS12* locus association with FEV1 ($p_{meta}= 1.41\times10^{-7}$). We subsequently carried out gene network analysis for each trait using SNPs with P-value of $< 1.0 \times 10^{-3}$ in GWAS. We found that the glycoprotein gene set is associated with all four pulmonary function measures in our pediatric Caucasian cohort and the association was validated in our pediatric African American cohort and two reported adult cohorts. Its meta-analysis p-values range from 6.29×10^{-4} to 2.80×10^{-8} for the four pulmonary function traits tested. In our study, we provide the first comprehensive analysis of pulmonary function traits in pediatric population and African Americans and identified novel pulmonary function specific gene set in that population.

1041W

Genetics of the sphingolipid metabolism in hypertension. *M. Fenger[1,2], A. Linneberg[1,3], J. Jeppesen[1,3].* 1) University of Copenhagen, Copenhagen, Denmark; 2) Hvidovre University Hospital, Hvidovre Denmark; 3) Glostrup University Hospital, Glostrup Denmark.

Several attempts to decipher the genetics of essential hypertension have been done, but only a few genes have been identified. Unsolved population heterogeneity and the insufficiency of the prevailing monogenic approach to capture genetic effects in a polygenic condition are the main reasons for the modest results obtained. The physiologically heterogeneity of diastolic and systolic blood pressures was resolved by partition of the study population by combined latent class analysis and structural equation modelling into an ensemble of 14 physiological more homogeneous subpopulations. Two-gene interactions were evaluated for the sphingolipid metabolic network, and he phosphatidate and redox metabolic networks by variance decomposition and by a information theory approach. On average more than 5,000 highly significant interactions were detected by variance decomposition in each of the subpopulations including 160 single nucleotide polymorphisms (SNP) in 82 genes. The number of interactions were reduced to less than 0.5% by only including interactions with significant mutual information. The analysis suggests that acid ceramidase and sphingosine kinase-1 to be functional hubs in blood pressure regulation. Of the 675 interactions with significant weighted mutual information 38 increased the prevalence of hypertension (diastolic and/or systolic), while 27 decreased the prevalence of hypertension in the study population. These interactions included genes from the networks mentioned above in a complex pattern of specific genotypes. Thus, the sphingolipid metabolic network was established to be of significance in regulating the blood pressure by itself and by integrated interaction with the phosphatidate and redox genetic networks.

1042T

Characterization of gene expression biomarker signatures for use as an Alopecia Areata Disease Activity Index (ALADIN). *J.E. Cerise[1], A. Jabbari[1], J.C. Chen[2,3], M. Duvic[6], M. Hordinsky[7], D. Norris[8], V. Price[9], J. Mackay-Wiggan[1], R. Clynes[1,4,5], A.M. Christiano[1,2].* 1) Depart of Dermatology, Columbia University, New York, NY; 2) Depart of Genetics & Development, Columbia University, New York, NY; 3) Joint Centers for Systems Biology, Columbia University, New York, NY; 4) Depart of Internal Medicine, Columbia University, New York, NY; 5) Depart of Pathology, Columbia University, New York, NY; 6) Dept of Dermatology, MD Anderson Cancer Center, Houston,TX; 7) Dept of Dermatology, University of Minnesota, Minneapolis,MN; 8) Dept of Dermatology, University of Colorado, Denver, CO; 9) Dept of Dermatology, UCSF, San Francisco, CA.

Alopecia Areata (AA) is a highly prevalent autoimmune disease in which the hair follicle is attacked by cytotoxic T lymphocytes. The accessibility of the target organ within skin biopsies has provided us the opportunity to develop a novel disease activity score based on quantitative composite gene expression signatures. Using both Affymetrix microarrays and RNA-Seq data analyzed using Bioconductor/R and Ingenuity Systems' IPA, we identified three striking gene expression signatures in total skin from both human AA and the C3H-HeJ mouse, namely the IFN response, including IFN-gamma and IFN-inducible chemokines, and a cytotoxic T cell (CTL) signature including CD8 and granzymes, implicating these effectors as the dominant inflammatory cells in AA pathogenesis, as well as a keratin network. To generate a functional biomarker from the AA transcriptome, we developed the Alopecia Areata Disease Activity Index (ALADIN), a multi-dimensional quantitative composite score. ALADIN was derived from expression levels of representative genes from both the IFN and CTL pathways and KRT network to generate a measure of the distance of AA transcriptional levels from a baseline obtained from the skin of healthy individuals. We anticipate that ALADIN will be useful as a dynamic functional biomarker to stratify and longitudinally track patients enrolled in observational and interventional clinical studies. The ALADIN score reduces the complexity of the transcriptome into a tractable multivariate index that can be used to quantitate and monitor patients' disease status and progression. We deployed ALADIN in our studies of mouse AA, and demonstrated that ALADIN scores provide a quantitative measure of observed disease reversal, and 'molecular distance to skin homeostasis' during the course of successful prevention and treatment of disease. ALADIN is currently being assessed in cross-sectional studies to validate its utility in human AA. We anticipate that ALADIN will be useful as a dynamic functional biomarker to stratify and longitudinally track patients enrolled in observational and interventional clinical studies.

1043F

Psoriasis susceptibility genes in patient with severe compared with mild phenotype. *P. Nikamo, M. Ståhle.* Department of Medicine, Dermatology Unit, Karolinska Institutet, Stockholm, Sweden.

Psoriasis is usually graded as mild (affecting less than 3% of the body), moderate (affecting 3-10% of the body) or severe (affecting more than 10% of the body). According to National Psoriasis Foundation the distribution of psoriasis severity are 67% mild, 25% moderate and 8% severe (http://www.psoriasis.org/). The degree of severity is generally based on the following factors: the proportion of body surface area affected; disease activity (degree of plaque redness, thickness and scaling); response to previous therapies; and the impact of the disease on the person. Several psoriasis GWAS have been published, but phenotypic stratifications are lacking. Our large patient material with detailed clinical and genotype information allows for different stratification procedures. Healthy controls (n=1750) are compared with patients with mild psoriasis (only topical therapy n=712) and severe psoriasis (systemic therapy n=543). The analysis involves genes known to associate with psoriasis. Major differences are detected, such as only patients with severe psoriasis associate with the IL23R (p=0.0007, OR=1.31, CI 1.12-1.53) and IL23A (p=0.0001, OR=2.04, CI 1.40-2.93) - genes in a key psoriasis pathway (the effective biologic ustekinumab is directed against IL23/IL12). In contrast, HLA-C and its putative functional partner ERAP1 (peptidase involved in trimming MHC Class I peptides) associates more with the mild phenotype. These are novel and rather dramatic data and additional genotypes are now being explored to dissect the differences between these phenotypes. Altogether, these findings indicate that genes participating in IL23 signaling play a significant role in the severity of psoriasis. Conversely, the characterization of non-MHC disease loci has been problematic, owing to the small effects of the underlying genetic variants. This can be overcome by careful phenotyping and characteristics of the material.

1044W

Validating trans-eQTLs using evidence of cis-mediation: a genome-wide analysis among 1,800 South Asians. *B. Pierce[1], L. Tong[1], R. Rahaman[1], L. Chen[1], M. Kibriya[1], M. Argos[1], J. Farzana[1], R. Shantanu[1], R. Paul-Brutus[1], R. Zaman[2], M. Rahman[2], J. Baron[3], H. Ahsan[1].* 1) Health Studies, Univ Chicago, Chicago, IL; 2) UChicago Research Bangladesh, Dhaka, Bangladesh; 3) University of North Carolina, Lineberger Comprehensive Cancer Center, Chapel Hill, NC.

Genome-wide studies of expression quantitative trait loci (eQTLs) indicate that the majority genes in the human genome show evidence of regulation by a cis-eQTL (i.e., cis-eSNP). However, less is known regarding the effects of genetic variants on expression of distant genes (trans-eQTLs), which are difficult to detect in genome-wide analyses for two primary reasons: (1) trans effects, on average, are weaker than cis effects and (2) a huge number of tests must be conducted to comprehensively assess the evidence for trans eQTLs, resulting in the use of stringent significance thresholds. Several large eQTL studies (>1,000 participants) have attempted to identify trans signals, typically using leukocytes as source of RNA. These studies have reported hundreds of 'significant' trans-eQTL signals, but only a fraction of these signals have been identified across multiple studies. While there are many explanations for this lack of consistency for trans findings, this observation highlights the difficulty of identifying true trans signals with confidence. In this work, we use data on genome-wide SNP and mononuclear cell array-based expression data from ~1,800 Bangladeshi individuals to (1) characterize the cis- and trans-eQTL patterns among South Asians and (2) demonstrate how identified trans signals can be validated by assessing evidence for cis-mediation. We first used standard eQTL methods to identify cis and trans signals using a false-discovery rate (FDR) of 0.05. Among the 448 independent putative trans-eSNPs we identified, we could classify 52 of these as cis-eSNPs (defined as r2 >0.8 with the lead eSNP for a cis-eQTL signal). For each of these 52 SNPs showing both cis and trans associations, we estimated the trans association both with and without adjustment for the cis-transcript, a potential mediator of the trans effect. For approximately half of these signals, the magnitude of the trans-eQTL association was drastically reduced (fold change >60%) after adjustment, implying that the trans effect is mediated by a measured cis-transcript. For weaker trans signals that did not surpass the FDR threshold, we show that cis-mediation analysis can identify a subset of these signals that are likely to be true effects. In conclusion, assessing evidence for cis-mediation is a promising and straightforward approach for validating trans signals and enhancing trans-eQTL discovery. This approach may be especially useful when ideal replication data is not available.

1045T

Transcriptome analysis of CD4+ lymphocytes in asthmatics with or without depression. *J. He, T. Wang, Z. Liang, X. Xiong, Y. Yang, Y. Ji.* Department of Respiratory, Sichuan University, Chengdu, Sichuan, China.

Rationale: Cumulative studies have shown that asthma is positively associated with depression. The underlying mechanisms are poorly understood. Assuming common pathophysiological mechanisms especially immune mechanisms exist, the aim of this study was to determine the common pathways between asthma and depression using the global gene expression pattern of CD4+ lymphocytes from asthmatics with or without depression. Methods: Four groups of subjects (Non-depressive asthmatic, NDA; Depressive asthmatic, DA; Depression, DE; and Healthy control, HC) consisted of 6 participants in each group were studied. Peripheral CD4+ lymphocytes were isolated and the global transcriptome profile was performed using the Agilent SurePrint G3 Human GE 8x60K Microarray. Experimental data analyses were performed using the software R and Bioconductor. Genes with expression level of at least ± 2.0-fold change and a false discovery rate ≤ 0.05 were defined as statistically significant changes. Functional analysis and pathway enrichment analyses of differentially expressed genes were performed using the Ingenuity Pathway Analysis (IPA). Results: A total of 1145 differentially expressed genes were identified in any of the AS vs. HC, DA vs. HC, or DE vs. HC comparisons after corrected for multiple comparisons. Among these genes, 148 genes were demonstrated as differentially expressed genes only in DA vs. HC. Gene pathway analysis using these 148 genes identified six significant biological processes: Acute Phase Response Signaling, Nicotine Degradation III, Melatonin Degradation I, Superpathway of Melatonin Degradation, Serotonin Degradation and Nicotine Degradation II. Conclusion: The use of microarray analysis to identify genes commonly differentially expressed in CD8 cells in non-depressive asthmatic; depressive asthmatic and depression groups may provide a foundation for further studies on the pathogenesis of co-exist of asthma and depression. The results of this study shed new light on the common molecular mechanisms of asthma and depression and may provide theoretical clues for further pathophysiological and clinical therapeutic studies.

1046F

Genetic variation in the ER stress response network in the mouse. *C.Y. Chow, D.D. Riccardi, X. Wang, M.F. Wolfner, A.G. Clark.* Dept Mol Bio & Gen, Cornell Univ, Ithaca, NY.

The endoplasmic reticulum (ER) is responsible for synthesis and maturation of many proteins essential for cellular function. ER dysfunction can have devastating consequences when misfolded proteins accumulate in the ER, resulting in "ER stress" associated with pathogenesis of diseases from diabetes to neurodegeneration. To understand the contribution of ER stress resistance to disease, it is important to determine the extent of natural variation in ER stress resistance. Previous studies in human cell lines and Drosophila demonstrate that there is extensive variation in ER stress response, but comparable data in mouse is conspicuously absent. To characterize the variation in ER stress transcriptional response in mouse, we studied the eight genetically diverse inbred laboratory mouse strains that had been employed in the Collaborative Cross. Mouse embryonic fibroblasts (MEFs) from each strain were exposed to tunicamycin, an ER stress inducing drug. RNA-seq was performed to identify transcriptional changes in response to ER stress by contrasts with RNA-seq runs of cells under drug-free control conditions. Many canonical ER stress genes showed inter-line differences in response, indicating that the core ER stress pathways may be more plastic than previously thought. To partition this regulatory variation into cis and trans effects, we measured ER-stress-induced gene expression by RNA-seq in MEFs from F1 hybrids of the original eight inbred lines, and quantified allele-specific expression across nearly every expressed gene. The results reveal a surprising degree of trans regulatory variation, consistent with the variability lying close to the head of the regulatory cascade. Whether the most important ER stress genes modulating variability in disease risk play these key regulatory roles remains to be demonstrated. But differences in cis vs. trans regulatory architecture of genes in a network such as ER stress may have important implications for effect sizes of allelic variants, the nature of mutation-selection balances, and the prospects for therapies that target single gene products.

1047W

Global Metabolite profiling and the Risk of Osteoporotic Fractures: A Systems Biology Approach by Integrating Genomics and Tissue-Specific Gene Expression Profiling, Proteomics and Metabolomics in Postmenopausal Caucasian Women. *YH. Hsu[1,2], S. Reppe[3], CL. Chi[4], D. Karasik[1], J. Brain[5], DP. Kiel[1], K. Gautvik[6].* 1) Hebrew SeniorLife Inst Aging Res and Harvard Medical School, Boston, MA; 2) Molecular and Integrative Physiological Sciences Program, Harvard School of Public Health, Boston, MA; 3) Oslo University Hospital, Ulleval, Norway; 4) Medical Informatics Center, Harvard Medical School, Boston, MA; 5) Dept. Environmental Health, Harvard School of Public Health, Boston, MA; 6) Department of Medical Biochemistry, Oslo University Hospital, Ulleval, Norway.

Metabolomics is the quantitative measurement of dynamic metabolic response to pathophysiologic stimuli or genetic modification making metabolome quintessence of biological processes. To identify metabolite signatures for osteoporotic fractures and to understand its underlying biological implication, we measured ~500 metabolites in bone tissue and serum from osteoporotic hip fracture patients (OP) and age-matched normal-bone mineral density controls. Bone tissues were obtained from trans-iliac biopsies in 84 post-menopausal Caucasian women. Transcriptome profiling of bone biopsies was measured by Affymetrix microchips. GWAS genotyping was done in all samples to identify variants associated with metabolites and gene expression. Metabolites were measured by three types of mass spectrometers (MS). MS signatures of metabolites were matched to a small-molecule library. Age, BMI, estrogen use, cigarette smoking, menopause age and medication were collected and used in statistical analyses. To find metabolite signature, we applied a support vector machine algorithm with a nonlinear classification and identified 12 endogenous metabolites significantly differed between OP and controls. These 12 metabolites are involved in fatty acid amides; lipid oxidation; bile acid metabolism; tryptophan metabolism and redox homeostasis. Among them, significant accumulation of 5 dicarboxylic fatty acids suggests potentially suppressed mitochondrial β-oxidation in OP. Higher levels of the cannabinoid (CB) receptors agonist, oleamide, were observed in OP, indicating a disruption of CB signaling. Serotonin (5HT) was also higher in OP. To further characterize underlying mechanisms, we performed regression analysis between transcriptome and these 12 metabolites, adjusted for covariates described above. The expression of 71 genes was associated with these 12 metabolites at $p < 5\times10^{-5}$. AR, FGF2, PRKAA2 and several G-protein coupled receptors expressions associated with oleamide, providing supportive evidence of the involvement of CB receptors in OP. A pathway enrichment analysis pointed out three distinct cellular functional pathways. In summary, we identified a 'signature' of 12 endogenous metabolites associated with osteoporosis in postmenopausal Caucasian women. We used the transcriptome to further characterize their functional implications. A replication study of metabolite signatures in ~1,000 Framingham Study participants is underway to further validate these findings.

1048T

Integrating Genome-Wide Association Data with a Protein-Protein Interaction Network to Define Chronic Obstructive Pulmonary Disease Gene Modules. *M-L.N. McDonald[1], M. Cho[1,2], M. Matthiesen[3], B. Harshfield[1], T.H. Beaty[4], C.P. Hersh[1,2], C. Lange[5], E.K. Silverman[1,2] on behalf of the COPDGene investigators.* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Aarhus University, Denmark; 4) Johns Hopkins School of Public Health, Baltimore, MD; 5) Harvard School of Public Health, Boston, MA.

Chronic obstructive pulmonary disease (COPD) is currently the third leading cause of death in the United States. Genetic association studies have identified several key genetic loci associated with this complex trait. The majority of SNPs tested in genome-wide association studies (GWAS) do not achieve conventional genome-wide significance but still may provide valuable information. Using systems biology approaches integrating genotype and protein-protein interaction (PPI) data, genes and proteins involved in the same disease have been shown to cluster together when centrality is measured by protein physical interactions. Therefore, we applied dmGWAS (Jia, P et. al. 2011, Bioinformatics, 27:94-102), a method that integrates GWAS results with PPI, to identify COPD modules related to case-control status in the COPDGene Study.

Non-Hispanic White (NHW) (Ncases=2,812, Ncontrols=2,534) and African American (AA) (Ncases=821, Ncontrols=1,749) participants were genotyped on the Illumina Human Omni Express chip. GWAS were performed separately for the NHW and AA participants adjusting for population substructure, smoking history and age of enrollment. Using the smallest P-value per gene (+/- 50kb), dmGWAS identified 11,815 modules for the NHW group and 11,841 modules for AA. The module score is based on the GWAS test statistics for all genes in the module and then normalized by randomly selecting modules with the same number of genes 100,000 times. These normalized module scores ranged from 4.7 (P=1.5E-6) to 11.8 (P<1E-16) in NHW and from for 2.6 (P=0.005) to 8.1 (P=2.2E-16) in AA. The average module size was 10.6 +/- 1.3 genes for the NHW and 9.1 +/- 1.0 for the AA. A total of 110 modules were significant in both NHW and AA analyses when the NHW served as a discovery group. These modules were comprised of 239 distinct genes that were input to Ingenuity Pathway Analysis (IPA®). The top highlighted molecular and cellular function, represented by 111 genes, was cell death and survival (P-value range: 2.1E-16 to 2.8E-4). IPA highlighted 90 different genes for their involvement with neurological disease (P-value range: 1.5E-17 to 2.7E-4), the top disease and disorder category. Future directions for this analysis include additional validation of the top modules using permutation and comparing these results to traditional pathway based analysis. Funding: P01HL105339, R01HL089856, and R01HL089897.

1049F

Hypothesis independent pathway analysis identifies biologic pathways influencing susceptibility to glaucoma. *J.L. Wiggs[1], J.N. Cooke Bailey[2], L.R. Pasquale[1,3], S.J. Loomis[1], J.H. Kang[3], B. Yaspan[4], M. Brilliant[5], W. Christen[3], J.H. Fingert[6], D. Gaasterland[7], T. Gaasterland[8], R.K. Lee[9], P.R. Lichter[10], Y. Liu[11], S.E. Moroi[10], L.M. Olson[2], J.E. Richards[10], J.S. Schuman[12], W.K. Scott[13], K. Singh[14], A. Sit[15], D. Vollrath[16], G. Wollstein[12], D.J. Zack[17], K. Zhang[18], R.R. Alllingham[11], M.A. Pericak-Vance[13], R.N. Weinreb[18], M.A. Hauser[19], J.L. Haines[2], NEIGHBORHOOD Consortium.* 1) Dept Ophthalmology, Harvard Med Sch, MEEI, Boston, MA; 2) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Genentech Inc, San Francisco, CA; 5) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 6) Departments of Ophthalmology and Anatomy/Cell Biology, University of Iowa, College of Medicine, Iowa City, IA; 7) Eye Doctors of Washington, Chevy Chase, MD; 8) Scripps Genome Center, University of California at San Diego, San Diego, CA; 9) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 10) Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI; 11) Department of Ophthalmology, Duke University Medical Center, Durham, NC; 12) Department of Ophthalmology, UPMC Eye Center, University of Pittsburgh, Pittsburgh, PA; 13) Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 14) Department of Ophthalmology, Stanford University, Palo Alto, CA; 15) Department of Ophthalmology, Mayo Clinic, Rochester, MN; 16) Department of Genetics, Stanford University, Palo Alto, CA; 17) Wilmer Eye Institute, Johns Hopkins University Hospital, Baltimore, MD; 18) Department of Ophthalmology, University of California, San Diego, SD, CA; 19) Department of Medicine, Duke University Medical Center, Durham, NC.

Using genome-wide SNP data, we assessed biologic pathways as annotated in the KEGG database for association with risk of primary open-angle glaucoma (POAG) overall as well as with the normal-tension POAG subgroup with increased susceptibility to optic nerve degeneration. Our pathway analysis included all SNPs in the GWAS to identify meaningful associations that did not meet the genome-wide significance threshold. The PARIS (Pathway Analysis by Randomization Incorporating Structure) algorithm used in this study creates a null distribution of random pathways to mimic the size and structure of the actual pathways. The algorithm uses this null distribution to determine whether statistical associations are due to the biologic properties of the pathway, or instead, its size and structure. We conducted a POAG case-control analysis of 3108 cases and 3430 controls from the GLAUGEN and NEIGHBOR studies genotyped on the Illumina 660W quad platform. We first analyzed each study individually for single-SNP associations using logistic regression models to test for association. A subsequent meta-analysis of the two studies was performed. After multiple testing correction, we found 14 pathways associated with risk of POAG overall (permuted $p<0.001$ for all pathways). These pathways were loosely grouped into three categories: metabolism, cellular adhesion and signaling, and autoimmune disorders. Twelve of these pathways contained the same HLA genes and after removal of these genes, 11 of the pathways were no longer significant, while the HLA gene set was highly significant ($p<0.001$). In the normal tension subgroup analysis 7 pathways were associated with disease risk ($p<0.001$) falling into 5 general categories: metabolism (butanoate and ketone bodies), amino acid synthesis, MAPK signaling; hedgehog signaling and glycosaminoglycan biosynthesis. Only the butanoate metabolism pathway was significant for both POAG overall and the normal-tension subgroup. The butanoate pathway includes enzymes responsible for GABA metabolism, which can influence ganglion cell function in the retina. Overall, these data provide insight into the complex genetic etiology of glaucoma-related optic nerve degeneration and provide hypotheses for future functional studies.

1050W

Gene Expression Profile of Synovial Fluid following Meniscal Injury; Osteoarthritis (OA) Markers Found. *D.D. Vance[1,3], L. Wang[2], E. Rampersaud[2], T. Guettouche[2], B.P. Lesniak[3], J.M. Vance[2], M.A. Pericak-Vance[2], L.D. Kaplan[3].* 1) Univ Miami Miller Sch Med, Miami, FL; 2) Hussman Institute for Human Genomics, Miami, FL; 3) UHealth Sports Performance and Wellness Institute, University of Miami, Miller School of Medicine, Miami, FL.

BACKGROUND: Meniscal tears predispose individuals to the development of osteoarthritis (OA). This is due in part to the mechanical changes that occur within the knee joint following meniscal injury. Additionally, molecular differences between arthritic and normal knee joints have been shown. It is, however, not clear when these molecular changes start to develop. We hypothesize that, in addition to the mechanical changes important molecular changes occur within the knee joint soon after the meniscal tear contributing to eventual development of OA. The purpose of this study is to characterize the gene expression profile of synovial fluid (SF) following a meniscal tear; looking for early expression of inflammatory and arthritis-related genetic markers. **METHODS:** SF was collected from 11 patients (<48yrs) undergoing an arthroscopic partial meniscectomy at the time of surgery. RNA was extracted from cell pellets precipitated from the SF. Gene expression analysis was completed using Illumina HumanHT-12 array and Lumi R code (Bioconductor) software. A Heat map identified patterns among gene expression profiles and pathway analysis was performed by Metacore software. T-tests with Bonferroni corrections compared subsets of genes between individuals with recent injury (< 2 months; short duration) versus longer time lapse (long duration) until meniscectomy. **RESULTS:** The top 5% genes expressed across all samples mapped to inflammatory and cytoskeleton remodeling pathways. Included in the most abundant genes; *LAIR 1, TMSB4X, CCR6, IL18 and IL10* all have been associated arthritic conditions previously. Clustering of 3 samples with short duration was observed from the heat map and when compared to those samples with long duration across a subset of 340 genes with previous links to inflammatory or arthritic conditions, the following showed significant differential expression ($p < 1.47\times10^{-4}$); *IL1A, ILF2, NFKB1, IL10RB, IL18, ILF3, IL13RA, BMP2K and IL10*. These data support the inception of changes early in the disease process as well as indicated continued changes in expression profiles over time from injury onset. **CONCLUSION:** Our results support the hypothesis that molecular changes predisposing meniscal injury patients to OA may occur earlier after injury than previously reported. These findings suggest the possibility for early interventional treatments to halt these molecular changes and reduce the susceptibility of patients to OA following a meniscal tear.

1051T

Genetic association study of adaptor protein complex 4 with cerebral palsy in Han Chinese population. *Q. Xing[1,2], H. Wang[1,2], T. Li[3], M. Chen[1,2], Q. Shang[4], D. Zhu[3], L. Wang[1,2], Q. Li[1,2], L. He[1,2], C. Zhu[3,5].* 1) Children's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; 2) Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University; Shanghai, China; 3) Department of Pediatrics, the Third Affiliated Hospital of Zhengzhou University; Zhengzhou, China; 4) Department of Pediatrics, Zhengzhou Children's hospital; Zhengzhou, China; 5) Center for Brain Repair and Rehabilitation, University of Gothenburg; Sweden.

Cerebral palsy (CP) is caused by injury or developmental disturbances to the immature brain and leads to substantial motor impairments. Adaptor protein complex 4 (AP-4) plays a key role in vesicle formation, trafficking, and sorting processes, which are critical for brain development and function. AP-4 consists of four subunits encoded by AP4E1, AP4B1, AP4M1 and AP4S1 gene respectively. AP-4 is identified to involve in trafficking of the a-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptor. A number of studies have pointed to the involvement of AP-4-mediated vesicular trafficking pathway in the etiology of cerebral palsy, the most notable of which are the causative mutations identified in each of them that have recently been reported in different CP families. We therefore postulated that variations in AP-4 coding genes may exert an important role in the susceptibility to CP. In the present study, sixteen SNPs were geneotyped among 517 CP patients and 502 healthy controls from the Chinese Han population. We systematically analyzed the association of the AP4E1, AP4B1, AP4M1 and AP4S1 genes with CP on the basis of clinical characters. No significant associations were found between these variants and the risk of overall CP. Subgroup analysis showed that rs1217401 of AP4B1 was significantly associated with CP as a sequela of hypoxic-ischemic encephalopathy (HIE) (HIE + CP) (allele: $p=0.042151$; genotype: $p=4.46\times10^{-6}$). The current results indicate that 16 variants studied in four units of AP-4 have no detected effects on the susceptibility to overall CP, but AP4B1 is a susceptible gene for HIE +CP in the Chinese population.

1052F

Association analysis of candidate gene polymorphisms in Asthma, Rhinitis and Chronic Bronchitis: preliminary results from the GEIRD study. *C. Bombieri[1], F. Belpinati[1], A. Baldan[1], A.R. Lo Presti[1], G. Malerba[1], S. Accordini[2], L. Calciano[2], M. Ferrari[3], I. Perbellini[4], P.F. Pignatti[1], R. De Marco[2].* 1) Dpt. of Life and Reproduction Sciences, Section of Biology and Genetics, University of Verona, Verona, Italy; 2) Dpt. of Medicine and Public Health, Section of Epidemiology and Medical Statistics, University of Verona, Verona, Italy; 3) Unit of Internal Medicine, University of Verona, Verona, Italy; 4) Unit of Occupational Medicine, University of Verona, Verona, Italy.

Within The Gene Environment Interactions in Respiratory Diseases (GEIRD) study (Int Arch Allergy Immunol 2010;152:255-263), we performed an association analysis, in a large and accurately defined series of Italian subjects, to investigate the genetic involvement in the susceptibility to respiratory diseases. The study population included 1000 subjects (aged 20-66 years) from the general population, enrolled between 2007 and 2010, through a 2-stage screening process (postal questionnaire and clinical examination). Cases and controls were diagnosed on the basis of the answers to a detailed interview, to collect informations about family history for respiratory diseases, individual and ecological exposures, diet, smoking habits, medication, life-styles, and of pre/post bronchodilator spirometry, methacoline challenge, and skin prick tests. According to collected data, subjects were hierarchically classified as follows: 342 asthma, 7 COPD (with no asthma), 112 chronic bronchitis (with neither asthma nor COPD), 205 rhinitis (without asthma, COPD or chronic bronchitis) patients, and 312 controls (subjects without respiratory diseases). Classification was not possible for 22 subjects which did not completely meet diagnostic criteria of neither cases nor controls. A total of 971 subjects were genotyped, by a customized version of the GoldenGate Genotyping assay (Illumina), for a panel of 384 SNPs (Haplotype-Tagging-SNPs), representative of 63 candidate genes with a previous indication of possible association to the studied diseases. COPD subjects (due to the limited number) and unclassified subjects were not considered in the present analysis. A preliminary association analysis based on allele frequency comparison was performed. Presence of association (unadjusted $p<0.001$) was observed between polymorphisms of RAD50-IL13-IL4 region and past-asthma, PDE4D and ever asthma with atopy, PDE4D and chronic bronchitis with atopy, SPINK5 and rhinitis, GPR154 and chronic bronchitis. Moreover, considering only atopic subjects, ADAM33 polymorphisms showed association with ever asthma, total asthma, and rhinitis. Haplotype analysis is going on in order to confirm the association of these genes with the studied diseases.

1053W

Extending the population spectrum for nonsyndromic orofacial clefting: Recruitment and genetic analyses in an Arabian population from Yemen. *K.U. Ludwig[1,2], K. Aldurai[3], A.C. Boehmer[1,2], B. Lippke[1,2], N. Daratsianos[4], M.M. Noethen[1,2], A. Jaeger[4], M. Knapp[5], E. Mangold[1].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life and Brain Center, Bonn, Germany; 3) Faculty of Dentistry, University of Science and Technology, Sana'a, Yemen; 4) Department of Orthodontics, University of Bonn, Bonn, Germany; 5) Institute of Medical Biometry Informatics and Epidemiology, University of Bonn, Germany.

Nonsyndromic orofacial clefting (nsOFC) is a frequent congenital disorder, and has a genetically complex etiology. There are two common subtypes of nsOFC, namely nonsyndromic cleft lip with or without cleft palate (nsCL/P) and nonsyndromic cleft palate only (nsCPO). Research into the genetics of nsCL/P and nsCPO has identified numerous genetic susceptibility loci. The past genetic analyses have predominantly included diverse European and Asian populations. The present study now investigated a large and phenotypically well-characterized sample of Arabian origin from the Yemen population. Individuals included nsOFC patients (n=310), their relatives (n= 107) and unaffected controls (n=423). Patients were recruited within the context of surgical outreach programs that took place in Yemen between 2010 and 2012. Controls were recruited from blood donors of the same ethnic origin in the same area. Blood samples of patients/controls were transferred to the Institute of Human Genetics, Bonn where DNA was extracted. As research into nsCL/P has seen major advances by the recent application of genome-wide association studies (GWAS), we analyzed the genome-wide significant variants that have been previously been identified in European and Asian samples in the nsCL/P (n=252) and the nsCPO group (n=45) of the Yemeni patients. Twenty-four markers (SNPs, single-nucleotide polymorphisms) representing fifteen loci were genotyped using MassArray spectroscopy (n=23) or TaqMan assay (n=1). Loci included the European high-risk locus at 8q24.21 (rs987525) and the IRF6 locus. Statistical analyses using Armitage trend test revealed SNPs at nine of these loci to show nominal significant P-values in nsCL/P, the lowest P-value observed being 2.46×10^{-05} for rs987525, with an relative risk of 1.80 (95% confidence interval (CI): 1.27-2.55) for the heterozygous and 2.45 (95% CI: 1.54-3.90) for the homozygous genotype. In total, markers at three of the loci (8q24.21, 9q22 and 13q31) withstood correction for multiple testing. In the nsCPO group, no nominal significant associations were identified. This observation is consistent with results from other populations and might reflect the fact that there is little genetic overlap between nsCL/P and nsCPO. Our results reveal that a majority of nsCL/P risk loci identified in European and Asian ethnicities are also conferring risk for nsCL/P in the Arabian population.

1054T

Further evidence suggesting a role for variation in *ARHGAP29* in nonsyndromic cleft lip/palate. *L. Maili[1], A. Letra[1,2,3], J.B. Mulliken[4], S. Slifer[5], S.H. Blanton[5], J.T. Hecht[1,3].* 1) Department of Pediatrics, Pediatric Research Center, University of Texas Health Science Center Medical School at Houston, Houston, TX; 2) Department of Endodontics, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 3) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 4) Children's Hospital of Boston, Boston, MA; 5) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a common birth defect occurring in approximately 1 in 700 births. The complex etiology of NSCL/P reflects multiple genetic and environmental factors acting individually or in concert. A variety of research approaches, including candidate gene, genome-wide linkage, and genome-wide association studies (GWAS), have been used to identify the etiologic genes contributing to NSCL/P. While numerous genes have been implicated, only a few have been replicated across datasets. *ARHGAP29* was suggested as a candidate gene for NSCL/P for being located in close proximity to *ABCA4* (1p22), a gene previously identified in a GWAS of NSCL/P. Rare, potentially damaging, coding variants in ARHGAP29 were found in NSCL/P cases, and its expression was detected during craniofacial development in mice. Taken together, these findings suggest a role for this gene in NSCL/P. In this study, we investigated if variations in *ARHGAP29* were associated with NSCL/P. Five SNPs flanking and within *ARHGAP29* were genotyped in our NSCL/P datasets consisting of simplex and multiplex families of white NonHispanic (WNH, primarily western European) and Hispanic race/ethnicity. Family-based association tests were performed, stratified by ethnicity and family history of NSCL/P. P-values ≤ 0.01 were considered statistically significant. Results showed strong association of three *ARHGAP29* SNPs with NSCL/P in the WNH families. Two intronic SNPs (rs1541098 and rs3789688) showed association with NSCL/P in all WNH families (p=0.0005 and p=0.0002, respectively), and in the simplex WNH families (p=0.003 for both SNPs). A SNP in the 3' UTR (rs1576593) also showed an association with NSCL/P, in all WNH families (p=0.002) and the multiplex subset (p=0.002). No associations were found between variants in *ARHGAP29* and NSCL/P in the Hispanic datasets. This study further supports *ARHGAP29* as a candidate gene for human NSCL/P in families of Caucasian descent.

1055F

Association of GABRG2 rs211307 polymorphism with susceptibility to epilepsy in Asians: a multicentre case control study and meta-analysis.

B.S. Haerian[1], L. Baum[2], P. Kwan[3], S.S. Cherney[4], H.J. Tan[5], A.A. Raymond[5], Z. Mohamed[1]. 1) Pharmacology, University of Malaya, Kuala Lumpur, Malaysia; 2) School of Pharmacy, The Chinese University of Hong Kong, Shatin, Hong Kong, China; 3) Division of Neurology, Department of Medicine and Therapeutics, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, Hong Kong, China; 4) Department of Psychiatry and The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China; 5) Department of Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

The gamma-aminobutyric acid receptor subunit gamma-2 (GABRG2) gene encodes GABR-gamma2 protein, which has been implicated in susceptibility to epilepsy. Several studies have investigated whether the synonymous GABRG2 rs211037 polymorphism is a risk factor for various epilepsy types including febrile seizure (FS), idiopathic generalized epilepsy (IGE) and symptomatic epilepsy (SE), however results have been inconclusive. Therefore, we examined association of this polymorphism in FS, IGE, SE, and cryptogenic epilepsy (CE), through both a multicenter case control study and a meta-analysis. rs211037 was genotyped in Hong Kong Chinese and in Malaysian Chinese, Indian, and Malay participants. Genotypes of 5101 participants, of which 1769 were from Malaysia and 3332 from Hong Kong, were included in this case-control study. Of the 1179 patients, 66 percent; 13 percent;, and 21 percent; were Chinese, Indians, and Malays, respectively. Of the Chinese patients, 37 percent; were from Malaysia and 63 percent; were from Hong Kong. Significant association was observed between rs211037 polymorphism and susceptibility to SE (T vs. C p= 0.000003 and TT vs. CC, p=0.00001) or to epilepsy with FS (T vs. C p= 0.02 and CT vs. CC, p=0.02) in overall Chinese and to IGE in Chinese from Malaysia (TT vs. CC, p=0.01), but not to CE. Meta-analysis revealed a strong association between rs211037 with FS and SE in Asians for alleles (p=0.02, p=0.002 and p<0.00001, respectively) and for all genotype models. Our data suggests that GABRG2 rs211037 polymorphism is a risk factor for susceptibility to SE and FS in Asians, particularly in Chinese.

1056W

Myosin light chain kinase gene associates with asthma in Spanish and Latino populations.

M. Acosta-Herrera[1,2], M. Pino-Yanes[1,3], A. Corrales[1,4], A. Barreto-Luis[1,4], J. Cumplido[5], E. Perez-Rodriguez[6], P. Campo[7], C. Eng[3], JC. Robaina[6], I. Machin[8], I. Quintela[9], J. Villar[1,2], M. Blanca[7], A. Carracedo[10], T. Carrillo[5], JG. Garcia[11], EG. Burchard[3], SF. Ma[12], C. Flores[1,4]. 1) CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; 2) Multidisciplinary Organ Dysfunction Evaluation Research Network (MODERN), Research Unit, Hospital Universitario Dr. Negrin, Gran Canaria, Spain; 3) Department of Medicine, University of California, San Francisco, CA; 4) Research Unit, Hospital Universitario N.S. de Candelaria, Tenerife, Spain; 5) Allergy Unit, Hospital Universitario Dr. Negrín, Gran Canaria, Spain; 6) Allergy Unit, Hospital Universitario N.S. de Candelaria, Tenerife, Spain; 7) Allergy Service, Carlos Haya Hospital, Malaga, Spain; 8) Allergy Unit, Hospital del Tórax, Complejo Hospitalario Universitario NS Candelaria, Tenerife, Spain; 9) Grupo de Medicina Xenómica, CEGEN- Universidad de Santiago de Compostela, Santiago de Compostela, Spain; 10) Grupo de Medicina Xenómica, CIBERER- Universidad de Santiago de Compostela-Fundación Galega de Medicina Xenómica (SERGAS), Santiago de Compostela, Spain; 11) Institute for Personalized Respiratory Medicine, Section of Pulmonary and Critical Care, Sleep and Allergy, Department of Medicine, University of Illinois at Chicago, Chicago, IL; 12) Section of Pulmonary and Critical Care Medicine, University of Chicago, Chicago, IL.

Asthma is a complex disease characterized by airway inflammation and bronchial smooth muscle contraction. About a dozen loci for asthma susceptibility have been identified by genome-wide association studies (GWAS) but more loci are likely contributing to disease risk. The myosin light chain kinase (*MYLK*) gene, which has been associated with asthma albeit not constituting a GWAS hit, encodes a key cytoskeleton effector that is highly expressed in asthmatic airway smooth muscle cells. Here we aimed to perform a replication of *MYLK* gene association with asthma in unrelated subjects from Spanish (n=3249) and Latino (n=3774) populations. A total of 30 tSNPs of *MYLK* were genotyped using iPLEX Gold assays (Sequenom) and used for imputation in the 1000 Genomes Project, in a sample of 606 asthma cases and 1258 controls from the Genetics of Asthma (GOA) study in the Spanish population (Sample 1). The top ancestry-adjusted associated SNP was followed up for replication in an independent Spanish sample of 320 cases and 250 controls (Sample 2) using TaqMan assays (Life Technologies). This SNP was also tested for replication in two additional study samples with genome-wide genotypes obtained from population-optimized SNP content Axiom arrays (Affymetrix). These studies included 264 cases and 551 controls of Spanish descent (Sample 3) and 1893 cases and 1881 controls of Latino ancestry from the Genes-environments & Admixture in Latino Americans (GALA II) study (Sample 4). Two non-coding SNPs in perfect linkage disequilibrium (r^2=1) were associated in Sample 1 after multiple testing adjustments (rs78442149 and rs77820417, $p=1\times10^{-6}$). The SNP rs77820417 was followed up, which replicated both in Sample 2 (p= 0.004) and Sample 4 (p=0.019), but not in Sample 3 (p=0.598). A random effects meta-analysis of the 7,023 individuals confirmed the strong association of rs77820417 with asthma (OR=1.64, 95% CI=1.01-2.69, $p=3.6\times10^{-7}$). These results support the importance of the *MYLK* gene in asthma susceptibility in Spanish and Latino populations. Funded by the Health Institute 'Carlos III' grants FIS PI11/00623 to CF; FI11/00074 to MAH; and FI12/00493 to ABL, co-financed by the European Regional Development Funds, 'A way of making Europe'. SFM was supported by the University of Chicago Core Subsidy Mini Award (ITM/CTSA UL1 RR024999) and MPY by Fundación Ramón Areces. We thank Servicio de Apoyo Informatico a la Investigación (University of La Laguna) for the HPC support.

1057T

APOE e2 homozygous individuals are underrepresented among elderly Brazilian population. _M.S. Naslavsky[1], M.L. Lebrao[2], Y.A.O. Duarte[2], E. Amaro Jr[3,4], T.A.B. Mendes[4], A.S. Rodrigues[4], G. Bandeira[1], D. Schlesinger[5], L.T. Grinberg[3,6], C.K. Suemoto[3], R.E.P. Leite[3], R.E.L. Ferretti[3], C.A. Pasqualucci[3], J.M. Farfel[3], R. Nitrini[3], W. Jacob Filho[3], M. Zatz[1]._ 1) Human Genome and Stem Cell Center, University of Sao Paulo (USP), Sao Paulo, SP, Brazil; 2) Public Health Faculty, USP, Sao Paulo, SP, Brazil; 3) Medical School, USP, Sao Paulo, SP, Brazil; 4) Instituto do Cerebro, Hospital Israelita Albert Einstein, Sao Paulo, SP, Brazil; 5) Mendelics Analise Genomica, Sao Paulo, SP, Brazil; 6) Department of Neurology, University of California San Francisco, San Francisco, CA, USA.

The genome variation of healthy individuals undergoing aging has been the focus of worldwide interest and research. In order to enhance our understanding on human genetic variability our group is building a comprehensive collection of data from a cohort of elderly individuals from São Paulo (Brazil) , which has a mixed population of about 12 million inhabitants characterized by a great ethnic variability. Up to now, the cohort is composed by more than 1400 individuals who will be submitted to whole genome analysis. However, a preliminary analysis of apoliprotein E, a key protein in the metabolism and transport of lipoproteins and cholesterol and coded by chromosome 19 _APOE_ gene, has shown unexpected results and called our attention. _APOE_ variants have been highly associated with the physiopathology of late onset Alzheimer's disease (LOAD) and thus genotyping the different alleles became a standard procedure within studies of aging-related cognitive impairment. The main isoforms are _APOE_ e2, _APOE_ e3 and _APOE_ e4. While _APOE_ e4 is thought to increase AD risk, _APOE_ e2 has been considered to be a protective variant. DNA samples from 1447 individuals, 60 years old or older were collected, including cognitively healthy octogenarians. Up to now, 1000 _APOE_ genotypes were obtained by allele-specific amplification. Genotypic and allelic frequencies were derived by direct count. Surprisingly, although allelic frequencies were within the expected frequency according to other population studies, the _APOE_ e2 homozygous genotype was underrepresented and shifted from Hardy-Weinberg Equilibrium (HWE). The same analysis in a subset of 519 samples of individuals deceased from natural causes (aged 50 to 102) obtained from the largest brain bank of São Paulo showed no deviation from HWE. It is suggested that, even considering _APOE_ e2 an LOAD protective allele, carriers of _APOE_ e2 isoform may be at risk for early vascular disorders such as hyperlipoproteinemia. We hypothesize that early vascular disorders may increase the mortality rate of _APOE_ e2 homozygotes and, therefore, be responsible for the decreased frequency of individuals homozygous among older healthy subjects. We are currently increasing the sample size and analyzing the correlation between vascular causes of mortality and expected _APOE_ e2 homozygous frequency. We believe this may bring important contribution to our understanding on the role of _APOE_ e2 polymorphism in LOAD and vascular disorders susceptibility.

1058F

GWAS-nominated variants in homogeneous asthma sub-phenotypes. _E. Lavoie-Charland[1], J.-C. Bérubé[1], M. Laviolette[1], L.-P. Boulet[1], Y. Bossé[1,2]._ 1) Centre de recherche Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada; 2) Department of Molecular Medicine, Laval University, Québec, Canada.

Purpose: Asthma is a heterogeneous disease and subgrouping patients with similar clinical characteristics is likely to be important to improve our molecular understanding of this disease. Large-scale genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) robustly associated with asthma. The goal of this study is to test GWAS-nominated SNPs in phenotypically similar asthmatic patients.

Methods: The number of clinical variables was reduced by factor analysis. With selected variables, k-means clustering was applied on 523 asthma patients to produce more homogenous subgroups of patients with asthma. Genotypes from 49 independent SNPs selected from previous GWAS on asthma were then compared between these phenotypically similar subgroups.

Results: Five variables were selected from the factor analysis. From these, four clinical sub-groups were produced: smoking history, low atopy and low lung function, high atopy, and young non-smoking with average atopy. Genotype frequencies of four SNPs were significantly different between sub-groups (p-value<0.05). One is located near IKZF4 and the others within RCBTB1, DPP10 and IL33. The frequency of the asthma risk allele for SNP on DPP10 (rs1435879) was less frequent in the highly allergic sub-group compared with the other sub-groups.

Conclusion: The genotype frequencies of four SNPs were different across homogeneous sub-groups of asthmatic patients. DPP10-rs1435879 was particularly different in a subgroup characterized as highly allergic. Whether SNPs in DPP10 can be used to modify the current asthma classification remains to be validated. Additional studies are needed to identify SNPs differentially distributed within asthma cohorts in order to improve current asthma classification, and ultimately offer targeted therapy.

1059W

Genetic profiles of lipid-associated polymorphisms for lipid-related traits for addressing constitution-based discrepancy in Koreans. _S. Cha, S.K. Chung, H. Yu, A.Y. Park, J.Y. Kim._ Med Res, KIOM, Daejeon, South Korea.

Abnormal lipid concentrations are risk factors for atherosclerosis and cardiovascular disease. Each constitutional type classified by Sasang constitutional medicine, a Korean conventional type of personalized medicine, has shown to have a distinctive pathological characteristic for cardiovascular disease risk. Here, we performed an association analysis between lipid-related traits and genetic variants from several genome-wide association studies according to Sasang constitutional types classified using an integrated diagnostic model consisting of face, body shape, voice, and questionnaire information. We estimated the associations of 26 lipid-associated variants in 20 loci using multiple regression analyses in 2 Korean populations, i.e., 8,597 individuals. Of the 26 variants, 12 were significantly associated with lipid levels, including LDL cholesterol, HDL cholesterol, and triglycerides. In subgroup analysis of Tae-Eum (TE) and non-Tae-Eum (NTE) types (each 2,664 individuals) on the basis of tertiles of probability values from the diagnostic model for TE type harboring predominant cardiovascular risk, the associations of 3 variants near APOA5-APOA4-APOC3-APOA1 (rs6589566 on triglyceride: p = 8.90E-11), APOE-APOC1-APOC4 (rs4420638 on triglyceride: p = 1.63E-05), and LIPG (rs2156552 on HDL cholesterol: p = 0.00428) remained significant in the TE type, while those of 3 variants near ANGPTL3 (rs10889353 on triglyceride: p = 0.00233), APOA5-APOA4-APOC3-APOA1 (rs6589566 on triglyceride and HDL cholesterol: p = 3.09E-06 and 0.000555, respectively), and LPL (rs6586891 on HDL cholesterol: p= 0.000200) remained significant in NTE type. Interestingly, the minor allele effects of the lipid-associated variants in TE type had a harmful influence on lipid risk, whereas those of the variants in NTE type had neutral influences due to a compensating effect among the associated variants. These findings supported that the genetic susceptibility to lipid risk presented by the minor allele effects of the variants may predispose TE type subjects to a high cardiovascular disease risk.

1060T

Pleiotropic effects of three SLE associated functional variants within IFIH1 linked to several autoimmune diseases. _J.E. Molineros, S.K. Nath._ Oklahoma Medical Research Foundation,Arthritis & Clinical Immunology Research Program, 825 Ne 13th, Oklahoma City, OK 73104.

Background Using admixture mapping, we recently identified three functional variants within IFIH1 (Interferon induced helicase 1) associated with Systemic lupus erythematosus (SLE)[Molineros et al. 2013. PLoS Genet 9, e1003222]. These three variants are rs1990760 (A960T), rs10930046 (H460R) and intronic SNP rs13023380. IFIH1 has been related to viral response and apoptosis functions. There is a close correlation between viral response and the development of autoimmune disease (AD). Therefore, ADs are likely to share susceptibility variants/genes, especially those related to viral immune response. In fact, the pleiotropic effect has been reported for IFIH1 missense variants rs1990760 with some ADs. Objective The objective was to extend the analysis of pleiotropic effects to six ADs and 2 non ADs, and to identify other possible novel AD associated variants within IFIH1. **Approach** We performed an imputation based association analysis using publicly available GWAS data from cohorts of European descent with six ADs: Rheumatoid Arthritis (RA, N=1999), Scleroderma(N=833), Psoriasis(N=919), Celiac Disease(CD, N=1716), Vitiligo(N=1089), Type1 diabetes (T1D, N=823), and two non-ADs: Myocardial Infarction (N=2967), Parkinson's disease(N=2000). All were compared with a common set of controls from the HRS(N=8833). Missing genotypes were imputed from 1000 Genomes variants using MACH. Only SNPs with imputation quality (Rsq >0.7) were used for this analysis. **Results** We identified significant associations with all three independent SNPs in all ADs. SNPs were not significant (P>0.05) in non-ADs. The rs1990760 was associated with Vitiligo (P = 8.16×10-5, OR = 0.83) and Psoriasis (P = 1.12×10-4, OR = 0.82); rs10930046 was associated with CD (P = 1.95×10-2, OR=1.40) and T1D (P =9.94×10-3, OR = 0.43); and rs13023380 was associated with Vitiligo (P = 6.17×10-5, OR= 0.83), Psoriasis (P = 7.15×10-3, OR=0.88), and T1D (P = 1.20×10-2, OR=0.88). Nine additional SNPs were identified but were correlated (r2>70%) with either one of these three SNPs. There was no haplotype effect of any combination of these SNPs with any other AD. **Conclusion** IFIH1 plays an important role in autoimmunity through multiple causal variants. In this study we confirmed association of three functional variants common to four ADs. Replication of the same multiple variants association with multiple ADs highlights the importance of IFIH1 for our understanding of the development of autoimmunity.

1061F

Are genes previously associated with schizophrenia also predictive of dimension-specific psychotic experiences in adolescence?
D. Sieradzka[1], R.A. Power[2], F. Dudbridge[3], E.L. Meaburn[1], R. Plomin[2], A. Ronald[1]. 1) Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck, University of London, London, UK; 2) MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK; 3) Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK.

Evidence suggests that a considerable amount of variance in schizophrenia liability can be accounted for by common genetic variants. As psychosis can be viewed as a continuum, here we test the hypothesis that these common variants associated with schizophrenia also influence sub-clinical psychotic experiences in adolescence. Our aim was to test whether polygenic risk scores from schizophrenia GWAS and specific single nucleotide polymorphisms (SNPs) previously identified as risk variants for schizophrenia (Bergen & Petryshen, 2012), were associated with dimension-specific psychotic experiences. Six dimension-specific psychotic experiences (self-reported: paranoia, hallucinations, cognitive disorganisation, grandiosity, anhedonia, and parent-rated negative symptoms), as measured by the Specific Psychotic Experiences Questionnaire (SPEQ), were assessed in a community sample of 2,130 16-year-olds. First, polygenic risk scores were calculated using estimates of the logs of odds ratios from the Psychiatric Genomics Consortium (PGC) GWAS mega-analysis of schizophrenia and polygenic risk analyses were conducted. Schizophrenia polygenic risk scores were a significant predictor of anhedonia at pT=.40, pT=.50 and pT=1 (all p-values<.008; corrected) and parent-rated negative symptoms at pT=.01 only (p-value=.004; corrected); however, not in the expected direction. The signal accounted for less than 0.5% of the variance in both dimensions. Second, individual SNP analyses were performed to test for associations between dimension-specific psychotic experiences and 28 SNPs previously associated with schizophrenia. A significant association was found between the SPEQ paranoia dimension and rs17512836 (TCF4; p=.02, post correction). Replication of these individual SNP analyses in an independent sample of 16-year-olds (N=3,427) drawn from the Avon Longitudinal Study of Parents and Children (ALSPAC) and assessed using the Psychotic-Like Symptoms Questionnaire (PLIKS-Q) failed to yield significant results. This study presents the first empirical test of whether aggregated common variants and single SNPs associated with schizophrenia are also predictive of dimension-specific psychotic experiences in adolescence. The results do not provide support for the hypothesis, but due to the limited power for identifying very small genetic effects, further research in larger samples sizes is required.

1062W

Association of *GALNT10* genetic variants with adiposity in African Americans.
M.E. Stromberg[1, 2, 3], J.M. Hester[1, 2, 3], P. Mudgal[1, 2], J. Li[1, 2], P.J. Hicks[1, 2, 4], B.I. Freedman[5], D.W. Bowden[1, 2, 4, 5], M.C.Y. Ng[1, 2]. 1) Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC; 3) Molecular Genetics and Genomics Program, Wake Forest University School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC; 5) Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC.

Despite the higher prevalence of obesity (35.7%) in African Americans (AAs) compared to non-Hispanic whites (23.7%), genetic contributors to obesity in AAs remain poorly understood. Recent genome-wide association studies identified a region near the *GALNT10* gene as strongly associated with body mass index (BMI) in AAs. *GALNT10* is a member of the N-acetylgalactosaminyltransferase family and is located in the Golgi apparatus. The 20 known members of this family have been reported to have varying expression patterns in different tissues and multiple associations with metabolic traits. *GALNT10* is expressed in the hypothalamus, which regulates hunger. We examined multiple SNPs in the associated linkage disequilibrium (LD) region for association with BMI and other adiposity measures including waist circumference to hip ratio (WHR), visceral adipose tissue (VAT), subcutaneous adipose tissue (SAT) and VAT:SAT ratio (VSR). The 9 common variants tagging the LD block surrounding previous GWAS top hits and 12 exonic variants from the NHLBI ESP GO database were genotyped. Coding SNPs were prioritized for minor allele frequency (MAF) $\leq$0.05 with predicted functionality using the PolyPhen program. Variants were tested for association with BMI, WAIST, WHR, VAT, SAT and VSR in a meta-analysis of African American cohorts totaling up to 4,992 subjects. The most significantly BMI-associated SNPs were common (MAF between 0.32-0.44), noncoding SNPs; the most significantly associated was rs6890277 (β=-0.096(SE=0.02) p=3.64E-06), followed by rs7708584 (p=3.71E-05), rs815611 (p=4.65E-06), rs7719067 (p=6.99E-06), rs2033195 (p=7.26E-06), rs1366219 (p=9.91E-06), and rs4958361 (p=0.001). Associations were seen for BMI, not other adiposity traits. There was minimal evidence for association of coding or rare variants (individually or jointly with sequence kernel association test analysis). Common variants near *GALNT10* are associated with BMI (overall body size), but not other measures of adiposity, and coding variants in the gene do not contribute measurably to the association. Fine mapping of the *GALNT10* gene region will assist in detecting causal variants for BMI in the region.

1063T

Investigation of the rs2157719 SNP in the CDKN2B-AS1 gene in a Primary Open-Angle Glaucoma Brazilian population.
H.F. Nunes[1], J.P.C. Vasconcellos[2], V.P. Costa[2], N.I.T. Zanchin[3], M.B. Melo[1]. 1) CBMEG, University of Campinas, Campinas - SP, Brazil; 2) Department of Ophthalmology, Faculty of Medical Sciences, University of Campinas, Campinas - SP, Brazil; 3) Fiocruz, Curitiba - PR, Brazil.

Primary open-angle glaucoma (POAG) is a chronic neurodegenerative disease that leads to progressive damage of retinal ganglion cells resulting in visual field loss. Glaucoma is recognized as the main cause of irreversible blindness worldwide. Although the pathophysiology of glaucoma is not well understood, positive family history is one of the most important risk factors for glaucoma development. Single nucleotide polymorphisms (SNPs) have been widely associated with glaucoma, although their frequencies vary among different populations. CDKN2B-AS1 gene encodes a long noncoding RNA, and it is likely to play a role in regulating the expression of genes at the 9p21 locus through epigenetic mechanisms. CDKN2A encodes 2 proteins, p16INK4A and p14ARF, in different reading frames from alternative splicing of exon 1, whereas CDKN2B encodes p15INK4B. The proteins are cyclin-dependent kinase inhibitors and are involved in cell cycle regulation through the Rb and p53 tumor suppressor pathways. They also seem to be involved in inducing apoptosis in response to stress in terminally differentiated neurons, which is of particular relevance to glaucoma, a disease characterized pathologically by apoptosis of retinal ganglion cells and loss of their axons. The goal of this study was to investigate the distribution of the rs2157719A>G SNP, located in the CDKN2B-AS1 gene as a glaucoma risk or protective factor in a sample of Brazilian POAG patients. A case control study involving the genotyping of 147 POAG patients and 127 control subjects was performed by direct sequencing. The frequency of each of the three possible genotypes of rs2157719 was compared between POAG patients and controls and a significant difference was detected (p=0.00019). The minor allele frequency of rs2157719 (G) in our study group from Brazil was 32% in the POAG patients and 51% in control subjects, suggesting a protective effect previously described in other studies. Significant difference was detected in the allele frequencies between the POAG patients and controls (p=0.0019). However, the mechanism through which alterations in these gene might cause or prevent glaucoma is unknown. Overall, these data and observations imply that the G allele at this locus is possibly associated with protection against glaucoma in this sample of the Brazilian population.

1064F

Genetic Testing for Age-Related Macular Degeneration in an Armenian Population.
K.W.S. Small[1 2 3], A.A. Abraamyan[1 2 3], B.Z. Zanke[4], P.R. Ramamoorthy[4]. 1) Molecular Insight Research Foundation, Los Angeles, CA; 2) Cedars-Sinai Regenerative Medicine Institite, Los Angeles, CA; 3) Macula & Retina Institute, Los Angeles, CA; 4) Arctic Dx - Macula Risk; Bonita Springs, FL.

Purpose Age-related macular degeneration (ARMD) has a significant genetic influence, especially in Caucasian groups. The Armenian ethnogenesis dates back at least 3,000 years and is considered to be a genetically isolated population. Additionally, the Armenian genocide created a genetic bottleneck. Because of these demographics, we hypothesized that the genetic involvement in ARMD may be different than that of other Caucasian populations. To our knowledge, this is the first reported evaluation of the genetic contribution of AMRD susceptibility genes in an Armenian population. Methods A retrospective review was performed of 38 Armenian patients with wet (exudative) ARMD who had genetic analysis using the commercially available Macula Risk genotyping method. We obtained buccal cheek swabs which were sent to Advanced Diagnostic Laboratories for routine clinical testing of previously documented ARMD risk alleles: Complement Factor H (CFH) with 5 Single Nucleotide Polymorphisms, Complement Component 3 (C3), Age-Related Maculopathy Susceptibility 2 (ARMS2) and mitochondrially encoded NADH dehydrogenase 2 (MT-ND2). Each genetic component adds to a specific portion of a Macula Risk score, which predicts risk of developing advanced macular degeneration. This was compared to the Macula Risk genetic database of a 786 person Caucasian population with wet ARMD. The data was analyzed using Mann-Whitney U test to show differences in age and Macula Risk score. Chi squared test with two degrees of freedom was used to show differences in the genotypes (except MT-ND2, which only required a 2x2 table). Results The Mann-Whitney U tests showed no statistically significant difference between the Armenian and Caucasian data sets for Macula Risk score. The average age for Armenians (83.6) was higher than Caucasians (79.8), p=0.04. The Chi squared tests show no statistically significant difference between the genotypes. Conclusion In Armenian and Caucasian patients with wet ARMD, there is no difference between genotypes and Macula Risk score. Although we found no statistical significance, our data set is relatively small and may not be sufficiently powered. Additional analysis is necessary to evaluate the reasons contributing to the age difference within these populations. In patients who present with wet ARMD, we can find no genetic difference in the ARMD risk alleles in the Armenian population compared to the Caucasian population.

1065W

Associations of Age-Related Macular Degeneration Susceptibility Genes to Drusen.
K. Yamashiro[1], M. Yoshikawa[1], M. Miyake[1,2], H. Nakanishi[1,2], I. Nakata[1,2], N. Gotoh[1], Y. Kurashige[1,2], K. Kumagai[1], M. Oishi[1], A. Tsujikawa[1], R. Yamada[2], F. Matsuda[2], N. Yoshimura[1]. 1) Ophthalmology, Kyoto Univ Graduate Sch of Med, Kyoto, Japan; 2) Center for Genomic Medicine/Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Large drusen, typical characteristics of age-related macular degeneration (AMD), is a great risk for developing vision-threatening choroidal neovascularization in the later stage of AMD. Although susceptibility genes for AMD have been indentified in previous studies, genetic background of large drusen is still unclear. We evaluated associations between large drusen and known 19 regions associated with AMD, using Japanese individuals recruited from the Nagahama Prospective Genome Cohort for the Comprehensive Human Bioscience (The Nagahama Study). Among the participants with age ≥50 years old, 1224 were analyzed for their drusen and genotyped using Illumina HumanOmni2.5M or HumanHap610K. After imputation based on 1000 Genomes Project data and our standard quality control, 3,139,805 single nucleotide polymorphisms (SNPs) were included for the analysis. Large drusen was observed in 377 participants, while 837 did not have large drusen in their both eyes. We screened 19 AMD susceptibility regions reported by the AMD Gene Consortium with an adjustment for age and sex. P-value of ≤0.05 was observed in CFH, COL8A, CFI, IER/DDR1, VEGFA, TGFBR1, B3GALTL, RAD51B, LIPC, C3, APOE, TIMP3, and SLC16A8. In contrast to the strongest susceptibility effect of ARMS2/HTRA1 to AMD, SNPs in ARMS2/HTRA1 did not show significant associations to large drusen. Furthermore, SNPs in ADAMTS9, C2/CFB/SKIV2L, COL10A1, TNFRSF10A, and CETP did not show significant associations to large drusen. Stepwise analysis revealed 9 genes associated with large drusen; C3, CFI, CFH, TIMP3, LIPC, APOE, TGFBR1, B3GALTL, and IER3/DDR1. Our findings that some strong susceptibility genes for AMD did not show significant associations to large drusen suggest that roles of AMD susceptibility genes might be different in the early stage of AMD and in the late stage of AMD.

1066T

Positive Associations of ZIC2, RASGRF1, and SHISA6 Gene with High Myopia in Japanese.
M. Yoshikawa[1,2], K. Yamashiro[1], M. Oishi[1,2], M. Miyake[1,2], Y. Akagi-Kurashige[1,2], K. Kumagai[1], I. Nakata[1,2], H. Nakanishi[1,2], A. Oishi[1], N. Gotoh[1], A. Tsujikawa[1], R. Yamada[1], F. Matsuda[2], N. Yoshimura[1], the Nagahama Study Group. 1) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Center for Genomic Medicine Inserm U.852, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Introduction: Myopia is the most common ocular disorder in the world. High myopia, with a prevalence of 1%-2% in the general population, is distinguished from common myopia by excessive increase in axial length of the eye ball and is associated with various ocular complications leading to blindness. Recently, 24 new chromosomal loci associated with refractive error and common myopia were reported in a large scale multi-ethnic genome wide association study (GWAS) and several loci were confirmed by another GWAS dealing myopic patients in Caucasian population. Therefore, we conducted a case-control study to investigate whether these genetic variations, which confer risk of common myopia in Caucasians, were associated with high myopia in Japanese or not. Because choroidal neovascularization (CNV) is the most common cause of visual loss related to high myopia, we also evaluated the contribution of those genetic variations to the occurrence of CNV among the high myopic patients group. Methods: 5 single nucleotide polymorphisms (SNPs) from 5 candidate genes; TOX, RDH5, ZIC2, RASGRF1 and SHISA6 reported in both of the two previous GWASs were genotyped using Taqman assay in a total of 1339 unrelated highly myopic Japanese patients. As a control, the genotype data of 3248 healthy Japanese individuals were obtained from the Nagahama Prospective Genome Cohort for the Comprehensive Human Bioscience dataset. Results: Significant associations of rs8000973 near ZIC2 (P=8.80E-10), rs4778879 in RASGRF1 (P=3.85E-07), and rs2969180 in SHISA6 (P=0.028) with high myopia were observed in this study. The odds ratios (95% confidence intervals) were 1.39 (1.25-1.54) for rs8000973 C allele, 0.78 (0.71-0.86) for rs4778879 A allele, and 1.11 (1.01-1.22) for rs2969180 G allele, respectively. The effect of rs2969180 allele G was in the opposite direction to the previous report, while the other 2 SNPs showed same direction to the original report. The SNPs in RDH5 (rs3138144) nor near TOX (rs7837791) didn't show association with high myopia. A total of 516 high myopic patients with CNV and 823 high myopic patients without CNV were evaluated, but none of 5 SNPs showed significant associations with the presence of CNV among them. Conclusion: This study suggests that ZIC2, RASGRF, and SHISA6 would be susceptibility genes for not only common myopia but also for high myopia.

1067F

A large clinical biorepository linked to de-identified electronic medical records mimics a random sample from the general population. *D.C. Crawford[1], M.D. Ritchie[2], L. Dumitrescu[1], S.A. Pendergrass[2], R. Goodloe[1], J. Boston[1], E. Farber-Eger[1], H.H. Dilks[1], J.L. Haines[1], W.S. Bush[1].* 1) Ctr Human Gen Res, Vanderbilt Univ, Nashville, TN; 2) Ctr Systems Genomics, Depart Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.

Present day limited resources demand DNA and phenotyping alternatives to the traditional prospective population-based epidemiologic collections. To accelerate genomic discovery with an emphasis on diverse populations, we as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed all non-European American samples available in BioVU, the Vanderbilt University biorepository linked to de-identified medical records, for Metabochip genotyping. BioVU began DNA collection in the Davidson County area in 2007; as of 2011, a total of 15,863 samples were identified for Metabochip genotyping including African Americans (n= 11,521), Hispanics (n=1,714), and Asians (n=1,122). Overall, more than half of the non-European descent patients are female (63.35%) and are young (mean age 37 years). The average number of clinical visits and billing (ICD-9) codes per patient was 81.8 (1 to 1,456) and 147.3 (1 to 3,617), respectively. The most frequent billing codes (collectively >50%) observed among African American adults >18 years reflected known population differences in prevalence and incidence of specific conditions such as hypertension, type 2 diabetes, and end stage renal disease. Despite the clinical origins of these data, comparisons with the population-based National Health and Nutrition Examination Surveys (NHANES) III and 1999-2002 (n=3,458 and 3,950 non-Hispanic blacks and Mexican Americans, respectively) suggest that common health metrics such as body mass index or BMI [African American mean = 27 kg/m2 (EAGLE BioVU) and 27 kg/m2 (NHANES)] and HDL-C [African American mean =50 mg/dL (EAGLE BioVU) and 51 mg/dL (NHANES)] exhibit little bias. Similar observations were made for BMI and HDL-C in EAGLE BioVU Hispanics versus NHANES Mexican Americans: 25.2 versus 27.1 kg/m2 and 46 versus 47 mg/dL, respectively. Differences observed for mean LDL-C between BioVU (African Americans = 99 mg/dL) and NHANES (non-Hispanic blacks = 116 mg/dL) were attributed to well-documented lipid lowering medication usage in clinical populations. Collectively, these data suggest this clinical collection is essentially a random sample of the general population, making this resource more than comparable to traditional epidemiologic collections in the context of genetic association studies.

1068W

Characterization of a Hispanic population from a biorepository linked to de-identified electronic medical records for genetic association and gene-environment studies. *E. Farber-Eger, J. Boston, R. Goodloe, S. Wilson, W.S. Bush, D.C. Crawford.* Ctr Human Genet Res, Vanderbilt University, Nashville, TN.

The rapid accrual of DNA samples offered by biorespositories linked to de-identified electronic medical records (EMRs) such as BioVU in Nashville, Tennessee has the potential to accelerate genomic discovery, particularly in diverse populations such as Hispanics. A major challenge for these clinical-based collections is the extraction of environmental variables suitable for gene-environment (GxE) studies. To contribute towards this massive challenge, we as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed ~15,000 non-European descent samples in BioVU for Metabochip genotyping including 1,714 Hispanics. More than half of EAGLE BioVU Hispanics are female (>63%) and most are young (mean age ~26 years). Despite the clinical origins of these data, the two most frequent billing (ICD9) codes for this population include normal pregnancy for adults (>18 years) and routine infant child health check for children (≤ 18 years). Given that race/ethnicity is administratively assigned in the clinic, we compared the Hispanic label in EAGLE BioVU to genetic ancestry from variants targeted on the Metabochip using STRUCTURE and EIGENSTRAT. Overall, the observed patterns of genetic ancestry among EAGLE BioVU Hispanics were consistent with a complex three-way admixture event. To further characterize this Hispanic sample for GxE studies, we performed automated searches of the clinic notes for acculturation variables 'country of origin' and 'primary language spoken.' A preliminary search of Spanish-speaking countries and their common misspellings was performed in approximately half of the sample; of these, 322 (19%) had a Spanish-speaking country mentioned in the clinical notes. The most frequent country mentioned was Mexico (37%) followed by Puerto Rico (12%), El Salvador (11%), and Guatemala (11%). String searches for 'needs a translator', 'Spanish speaking', and 'does not speak English' among the first half of this Hispanic sample suggest Spanish is the primary language for ~27% of the patients. Further data mining is being conducted to fully characterize the diverse population in EAGLE BioVU for genetic association and GxE studies.

1069T

Mitochondrial genetic effects on proliferative diabetic retinopathy. *D.C. Samuels[1], M.A. Brantley, Jr[2].* 1) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Eye Institute, Vanderbilt University Medical Center, Nashville, TN, USA.

Diabetic retinopathy is the leading cause of vision loss in working-age adults in the US. The most severe form of this disease is Proliferative Diabetic Retinopathy (PDR), which is defined by aberrant blood vessel growth in the retina. The increased glucose level in diabetes causes increased flow of metabolites into the glycolytic pathway, while reactive oxygen species produced by mitochondria modulate the flow of metabolites out of the glycolytic pathway. The combination of these two effects, acting at opposite ends of glycolysis, determines the level of intermediate glycolytic metabolites and the rate at which these metabolites are diverted into alternative cytotoxic pathways, such as the polyol, hexosamine, protein kinase C, and the advanced glycation end products (AGE) pathways. Increased activation of these pathways in diabetes has been suggested to cause cellular damage leading to diabetic retinopathy and other diabetic complications. We hypothesized that mitochondrial genetic variation alters reactive oxygen species production, thus modulating the flow of metabolites through glycolysis and affecting the development and progression of diabetic retinopathy. Using BioVU, the de-identified version of the Vanderbilt Electronic Medical Record, we found that the common European mitochondrial haplogroup H was significantly overrepresented in the PDR cases compared to the Non-Proliferative Diabetic Retinopathy (NPDR) controls (p=0.038, OR= 3.4 [95% CI 1.1-10.7], 21 cases & 54 controls). This result was replicated in a clinical cohort (p = 0.0024, OR = 3.8 [1.6-8.8], 58 cases & 44 controls). In addition, the common haplogroup U(k) was significantly protective against PDR in the combined BioVU and Clinical cohort group (p=0.031, OR = 0.43 [0.20-0.90]), though the association was not significant in the individual groups.

1070F

The association of 9p21.3 with acute myocardial infarction in managed care populations is independent of statin therapy. *P. Erlich[1], D. Carey[1,2], S. Steinhubl[1], C. McCarty[3], D. Cross[4,5].* 1) Center for Health Research, Geisinger Health System, Danville, PA; 2) Weis Center for Research, Geisinger Health System, Danville, PA; 3) Division of Research, Essentia Institute of Rural Health, Duluth, MN; 4) Marshfield Clinic Research Foundation, Marshfield, Wisconsin; 5) Department of Forensic and Investigative Genetics, School of Biomedical Sciences, University of North Texas Health Science Center, Fort Worth, Texas.

Introduction: A locus on 9p21.3 is associated with risk for acute myocardial infarction (AMI) in multiple GWAS and replication studies. Current evidence suggests that 9p21.3 represents a previously-unknown pathway in AMI etiology, which is independent of traditional cardiovascular risk factors. If this hypothesis is true, it follows that the effect of 9p21.3 should not be influenced by the use of statins and that it should add a significant increment of performance to risk factor-based predictive models. **Methods:** This nested case-control study used electronic health records (EHR) linked to population-based biobanking initiatives of two integrated healthcare delivery systems in the US - Geisinger Health System in PA and Marshfield Clinic in WI. A cohort of 18,329 individuals (57% female; 108,400 person-years) aged 40 to 80 was constructed. EHR were scanned to ascertain incident AMI status and covariates. Subjects with prevalent cardiovascular disease at baseline were excluded. A nested sample of 721 incident AMI cases and 722 matched controls was genotyped for a SNP (rs2383206) located in an intron toward the 3' end of the CDKN2B-AS1 transcript. Bivariate and multivariate logistic regression was used to test for association of rs2383206 with incident AMI and examine hypotheses related to effect independence. Receiver operating characteristic-area under the curve (ROC-AUC) was used to discern the net predictive performance of rs2383206 within a model with traditional risk factors. **Results:** The crude AMI incidence in the cohort was 10.8 and 5.0 per 1,000 person-years among males and females, respectively. In the nested sample, rs2383206 was associated with AMI in bivariate analysis (OR per A allele=0.71; 95%CI 0.62, 0.83; p<0.0001) and, with no notable difference in effect magnitude, in multivariate analysis adjusted for traditional risk factors and statin therapy (OR per A allele=0.72; 95%CI 0.61, 0.85; p<0.0001), suggesting an independent effect. The ROC-AUC of the parsimonious predictive model excluding rs2383206 was 70% (95%CI 67%, 73%), and increased by 1.2% upon addition of rs2383206 (95%CI for the increment 1.6%, 2.2%; p=0.02). **Conclusions** A SNP in 9p21.3 is associated with incident AMI in managed care populations independently of traditional risk factors and statin treatment. When added to a risk factor model, the SNP enhanced predictive performance by a small, statistically significant increment.

1071W

Computational resources required to transform bedside data to base-pair research. *W. Bush, J. Boston, E. Farber-Eger, R. Goodloe, D. Crawford.* Ctr Human Gen Research, Vanderbilt Univ, Nashville, TN.

Biorepositories linked to de-identified electronic medical records (EMRs) have enormous potential for disease gene mapping and pharmacogenomics. Tapping this potential, however, requires a range of specialized skills, software, and hardware. To illustrate these needs, we describe here the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) BioVU, a subset of the Vanderbilt University biorepository of ~15,000 DNA samples from non-European Americans representing African Americans (n=11,521), Hispanics (n=1,714), and Asians (n=1,122). EMR data for each patient, including billing (ICD9) codes, procedure codes, medications, labs, problem lists, and clinical notes have been scrubbed of identifiers and are accessible for research in the Synthetic Derivative (SD), the de-identified mirror image of the EMR updated monthly to capture new clinic visits and associated clinical data. To create the EAGLE BioVU resource for genomic analysis, we first captured a static view of the SD records to create a stable data structure or 'data freeze' for subsequent electronic phenotyping. This capture included 2.3 million billing codes, 4.6 million procedure codes, 13 million medication records, 11 million laboratory values, and over 13 billion characters of clinical text, amounting to more than 14 gigabytes of data. Electronic ascertainment for a given study requires a collaborative multidisciplinary team. Genetic epidemiologists and content experts define criteria for electronic phenotyping. Database queries are performed over the entire collection by software, and database developers to define phenotype groups. Laboratory values and longitudinal data are processed by statisticians to assess distribution assumptions and outliers. The final result is a set of samples with well-defined phenotypes for subsequent study. To date, this process has been completed for body mass index, type 2 diabetes and associated traits and conditions, cardiovascular disease and related traits, reproductive traits (ages at menarche and menopause), ocular diseases, peripheral artery disease, and kidney disease and associated labs. While new advances improve the ability to manipulate large datasets, these processes are computationally and labor intensive necessitating the creation of specialized workflows to transform clinical data into usable phenotypes for genomic discovery.

1072T

Hirschsprung Disease Research Collaborative (HDRC): A multidisciplinary partnership to advance Hirschsprung disease research. *C. Berrios[1], F. Abdullah[1], P.K. Frykman[2], R. Kapur[3], J.C. Langer[4], A. Chakravarti[1].* 1) Johns Hopkins University, Baltimore, MD; 2) Cedars-Sinai Medical Center, Los Angeles, CA; 3) Seattle Children's Hospital, Seattle, WA; 4) Hospital for Sick Children, Toronto, Ontario, Canada.

Hirschsprung disease (HSCR) or congenital aganglionosis, a functional intestinal obstruction affecting 1 in 5,000 neonates, is a multifactorial neurodevelopmental genetic disorder with >15 susceptibility genes identified and numerous more becoming evident through exome sequencing. Importantly, rare coding variants, common polymorphisms and chromosomal anomalies, involving both coding and non-coding sequences, are all implicated in HSCR. Demonstrated correlation of variant types with clinical presentation offers promise for translation of genomic findings into prognostic markers. However, these results have been obtained from referrals of exceptional, not random, cases. HSCR is treated surgically, and has a 30% incidence of suboptimal post-operative outcomes. A lack of standardized pathology and surgical approaches has prevented a full understanding of the variability in outcomes, and research is therefore needed using standardized phenotypic data from randomly ascertained patients. The HDRC was formed in 2011, as a partnership of geneticists, pediatric surgeons, pathologists and gastroenterologists, to fill this need. Interest in joining the HDRC has been expressed by clinicians and researchers at 60 sites in 9 countries, with 35 commitments to join. IRB approval for enrollment has been obtained at 10 sites and 2 additional sites refer patients to the Coordinating Center for enrollment. Parent-child trios are enrolled, with collection of detailed, standardized medical/family history from questionnaires and medical records. Blood samples are collected, with plans to include fresh and fixed pathology samples for genomic and histopathological analyses. Over the past year, a growing number of sites have submitted 74 patient and 84 parental samples. The HDRC has already embarked on a pilot retrospective study to correlate genetic and histopathology markers with post-operative outcomes. Each study site contributes a few samples with clinical data, and subsequently benefits from the opportunity to access the pool of HDRC samples and data for IRB-approved studies. While challenges have included coordination of a large number of sites and agreement among IRB approvals and materials transfer contracts, the HDRC is making steady progress toward building a large international biorepository of samples, accompanied by clinical and genomic data, within a multidisciplinary collaboration to facilitate research and improve patient care for children with HSCR.

1073F

Risk assessment of *APOL1* genetic variants on renal function and lipid profile in multiethnic hospital-based population. *M. Udler[1, 2], V. Lotay[3], G. Belbin[2,3], C. Wyatt[1], O. Gottesman[3], E. Bottinger[3], E.E. Kenny[2,3,4,5], I. Peter[2,5].* 1) Dept of Medicine; 2) Dept of Genetics and Genomic Sciences; 3) The Charles Bronfman Institute for Personalized Medicine; 4) The Center for Statistical Genetics; 5) The Institute for Genomics and Multiscale Biology, all at Icahn School of Medicine at Mount Sinai, NY, NY.

Variants in *APOL1* are strongly associated with non-diabetic end-stage renal disease (ESRD) in African Americans (AA), however, the mechanism by which these variants affect renal function is unknown. *APOL1* encodes apolipoprotein L1, a component of high density lipoprotein (HDL) cholesterol. Given that patients with chronic kidney disease are prone to developing dyslipidemia, we investigated whether *APOL1* variants were also associated with lipid levels. Additionally, it was of interest to obtain genetic risk estimates in a diverse, hospital-based population. The Mount Sinai Biobank is a repository of genetic data collected from patients and anonymously linked to their electronic medical records (EMRs). Three coding variants in *APOL1* previously shown to be associated with ESRD in AA (Genovese *et al* 2010) were genotyped in 13,527 Biobank participants (3706 AA, 4465 European Americans (EA), and 5356 Hispanic/Latinos (HL) by self-reported ancestry). MDRD estimated glomerular filtration rate (eGFR) and lipid levels were obtained from the patients' medical records. Associations were analyzed using regression analysis with recessive model and adjustment for global proportion European, African, or AmerIndian ancestry per individual, using Illumina 770K chip data and the ADMIXTURE algorithm. The frequency of homozygous risk alleles at any of the three variants was 7.6% in AAs, 0.7% in HLs, and 0% in EAs. Among non-diabetic patients, harboring two *APOL1* risk alleles at any variant was significantly associated with ESRD in both AA (OR 5.5, 95% CI 2.4-12.5, $P=5\times10^{-5}$) and HL (OR 21.9, 95% CI 2.3-207.2, $P=0.007$). Associations in AA, but not in HL, remained significant with adjustment for age, sex, blood pressure, BMI, and smoking status. The variants were not significantly associated with HDL, low density lipoprotein, triglyceride levels, non-HDL cholesterol, or total cholesterol in patients who were not prescribed lipid-lowering medication (N=2448 AA, 3032 EA, 3099 HL). Nor were the variants associated with being prescribed lipid-lowering medication. Our findings suggest that *APOL1*'s effect on ESRD risk is not mediated through altering lipid levels. This is the largest and most ethnically diverse study to-date to characterize variation at *APOL1* with regard to both eGFR and lipid profile, and demonstrates how Biobanks linked to EMRs can facilitate investigation of genetic associations with multiple phenotypes in diverse ethnic populations.

1074W

Towards a phenome-wide catalog of human clinical traits impacted by genetic ancestry. *L. Dumitrescu[1], R. Goodloe[1], J. Boston[1], E. Farber-Eger[1], W.S. Bush[1,2], D.C. Crawford[1,3].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 3) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

Racial/ethnic differences for commonly measured clinical variables, such as cholesterol and blood pressure (BP), are well documented. Although the causes of these observed differences are unclear, it has been postulated that population-specific genetic factors may play a role. The genetic heterogeneity of admixed populations, such as African Americans (AA), provides a unique opportunity to identify genomic regions and variants associated with the clinical variability observed for diseases and traits across populations. To begin a systematic search for these population-specific genomic regions at the phenome-wide scale, we determined the relationship between global genetic ancestry, specifically European ancestry, and clinical variables measured in a population of AA from BioVU, Vanderbilt University's biorepository linked to de-identified electronic medical records (EMR), as part of the Epidemiologic Architecture using Genomics and Epidemiology (EAGLE) study. Through billing (ICD9) codes, procedure codes, labs, and clinical notes, 37 common clinical and laboratory variables were mined from the EMR, including body mass index (BMI), kidney traits, lipid levels, BP, and electrocardiographic measurements. A total of 11,167 DNA samples from AA were genotyped on the Illumina Metabochip containing ~200,000 variants. A subset of Metabochip SNPs (n=75,000) were chosen as ancestry informative markers and percent global European ancestry was estimated using STRUCTURE version 2.3.4 (K=3), with data from 395 HapMap samples (CEU, YRI, and CHB/JPN) included as predefined parental clusters. Linear regression, sex-stratified and sex-combined, was used to examine associations between global ancestry and the phenotype of interest. EAGLE/BioVU AA had a mean percent European ancestry = 17.2±15%. Increased European ancestry was most strongly correlated with an increase in QRS duration (p=6.7E-05, beta(SE)=8.0(2.0), n=837), consistent with previous observations that AA tend to have shorter a QRS duration than European Americans. A nominally significant association was also observed for BMI among AA females (p=3.9E-02, beta(SE)=1.3(0.6), n=5959). However, despite known racial/ethnic disparities in BP, European ancestry was neither associated with diastolic nor systolic BP measurements. Collectively, these results suggest that this clinical population can be used to identify traits in which population differences may be due, in part, to population-specific genetics.

1075T

Practical implementation of polygenic risk scores in personalized risk assessment for common cardio-metabolic traits.
K. Fischer[1], K. Läll[2, 1], R. Mägi[1], T. Haller[1], L. Leitsalu[1], T. Esko[1], A. Metspalu[1]. 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Institute of Mathematical Statistics, University of Tartu, Estonia.

Large-scale genome-wide association studies (GWAS) identify an increasing number of genetic variants associated with common complex diseases. Such markers (SNPs) are used to form polygenic risk scores. We will discuss the essential steps needed to translate GWAS results to personalized risk estimates. The process is illustrated using the data of Estonian population-based biobank and polygenic risk scores for Type 2 Diabetes (T2D), Coronary Artery Disease (CAD) and primary hypertension (HTN). The following questions will be addressed: 1)Validity of the genetic risk score in the target population. We will propose some graphical tools that address the question on consistency of published effect estimates (based on meta-analysis of a large number of cohorts) and estimates within the cohort of interest. That is done for each of the individual SNPs, as well as summary risk scores. 2)Conventional (non-genetic) risk factors and adjustment. Typically the GWAS results are obtained using minimal adjustment for other covariates. In practice, one is mainly interested in the added predictive value of a genetic risk score conditional on conventional risk factors. We will study the differences of conditional and unconditional effects, as well as possible heterogeneity across subsets with different baseline risk levels. 3)Number of markers to be included in the risk score. Increasing the number of SNPs does not necessarily lead to improved predictive accuracy in a particular cohort and there is an optimal number of SNPs producing the best results. A graphical tool is proposed to aid marker selection. 4)Population risk estimates from available genotyped samples. Often the genotyped (subsets of) biobank cohorts include a larger proportion of prevalent cases than the underlying population. We will show how the risk estimates can be corrected using external data on population prevalence. 5)Communication of the risk estimates in practice. We will propose some useful graphs that help to interpret the actual meaning of the risk estimate. For the Estonian Biobank cohort we show that the risk score for T2D has a very good discriminatory ability between high-risk and low-risk individuals, especially in those who are overweight, but not extremely obese (BMI 25-35). We will also show that the risk score for CAD should be mainly used to identify individuals at the highest percentiles of risk, whereas the risk score for HTN may be useful in some well-defined subsets of the cohort.

1076F

Optimization of Genetic Coverage and Biomedically Relevant Content of Microarrays for Genotyping in Biobank Cohorts.
J. Schmidt, Y. Zhan, J. Gollub, Y. Lu, G. Hsiao, M. Nitzberg, E. Schell, L. Bellon, T. Webster. Informatics, Affymetrix, Santa Clara, CA.

Growing numbers of biobanks hold large collections of clinically interesting samples. Common trends have emerged in needs for genotyping these large cohorts for biomedical studies, primarily the desire to simultaneously assay rare, potentially causal variants and large panels of common markers for genome wide association studies (GWAS). However, the detailed needs for each biobank or study depend on the ancestry of the cohort as well as the specific study aims. We present a system for optimizing an Affymetrix® Axiom® genotyping microarray for a specific study or biobank, including: (1) a computationally tractable method for selecting space-efficient probes to maximize genetic coverage in the population and allele frequencies of interest for GWAS; and (2) modular sets of markers relevant to specific biological and biomedical questions, which can be further customized for the specific phenotypes and populations of interest. We demonstrate the selection of only 246k markers, with minimal space requirements on the array, to provide 87% imputation-based coverage of the CEU population across 6.7M markers with MAF ≥ 5% in the 1000 Genomes Phase 1 (March, 2012) genotype data. We also describe the selection of smaller sets of markers ('booster panels') to augment the coverage provided by this core 246k set in non-target populations, such as African and Asian populations, or to increase coverage of lower frequency markers. E.g., the addition of only 50k markers can be used to significantly raise imputation coverage of common Yoruban variants (MAF ≥ 5%), or increase coverage at lower minor allele frequencies in the primary target population, or raise coverage in any particular region of interest. Other sets of markers can be added to suit specific study needs, such as ADME or HLA markers, known GWAS hits from previous studies, coding, non-synonymous SNPs and indels or markers with predicted loss-of-function impact, etc. This system for marker selection can improve the efficiency of genotyping in a biobank setting, potentially allowing more and more effective genetic studies in very large cohorts.

1077W

ChIP-seq in alcoholic steatohepatitis and normal liver tissue identifies candidate disease mechanisms suggesting progression to cancer.
C. Wadelius[1], MB. Bysani[1], O. Wallerman[1,5], S. Bornelöv[2], K. Zatloukal[3], J. Komorowski[2,4]. 1) Dept Immun, Gen & Pathol, Uppsala University, Uppsala, Sweden; 2) Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; 3) Institute of Pathology, Medical University of Graz, Austria; 4) Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warszawa, Poland; 5) Current address: Department of Medical Biochemistry and Microbiology, BMC, Uppsala.

Background: Excessive consumption of alcohol may induce accumulation of fat in the liver concurrent with inflammation in a disease called steatohepatitis, which may progress to liver cirrhosis and hepatocellular carcinoma. Its molecular pathogenesis is to a large degree unknown. Histone modifications play a key role in transcriptional regulations as marks for silencing and activation of gene expression and for location of functional elements. Many transcription factors (TFs) are crucial for the control of gene activity and abnormality in their function may lead to disease. **Methods**: We performed ChIP-seq of histone modifications associated with active promoters and enhancers and a candidate transcription factor, in liver tissue from patients with steatohepatitis and normal livers and correlated results to mRNA-expression and genotypes. **Results**: We performed ChIP-seq of the three histone modifications H3K4me1, H3K4me3 and H3K27ac and the TF upstream stimulatory factor 1 (USF1) in tissue from normal liver and from alcoholic steatohepatitis patients. We found many peaks that are differentially enriched between disease and normal tissue, and qRT-PCR results indicated that the expression of the tested genes strongly correlated with ChIP-seq of histone modifications but is independent of USF1 enrichment. By gene ontology analysis of differentially modified genes we found some which had previously been implicated in the etiology of steatohepatitis and others that are candidates to contribute to the disease. Importantly, the genes associated to the strongest histone peaks in the patient were over-represented in cancer specific pathways suggesting that the tissue was on a path to develop to cancer, a common complication to the disease. We also found several novel SNPs and GWAS catalogue SNPs that are candidates to be functional and therefore needs further study. In summary we find that analysis of chromatin features in tissue samples provides insight into disease mechanisms.

1078T

National Biological Sample And Data Repository For WHO Group 1 Pulmonary Arterial Hypertension.
M. Pauciulo[1], A. Reponen[1], K. Lutz[1], C. Winslow[1], A. Walsworth[1], J. Harley[2], M. Barnes[2], L. Martin[1], K. Marsolo[3], W.C. Nichols[1]. 1) Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH; 2) Division of Rheumatology, Cincinnati Children's Hospital, Cincinnati, OH; 3) Division of Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH.

We have established the National Biological Sample and Data Repository for PAH with funding from the National Heart, Lung, & Blood Institute of the National Institutes of Health. Biological samples, clinical data, and genetic data are being collected on approximately 3,000 WHO Group 1 PAH patients from 3/2012 to 2/2017 to create this bio repository. Twenty four pulmonary hypertension centers across the United States have been enlisted to enroll patients, collect peripheral blood samples, and enter clinical data into a web-based eCRF. Many of the patients recruited for the bio repository are already enrolled in a multicenter, observational, U.S.-based Registry of PAH. Efforts are also being made to enroll treatment naïve patients to obtain both pre and post-treatment samples. Serum and plasma are being isolated and banked using the patient blood samples. Additionally, both DNA and RNA are being isolated from a portion of the obtained lymphocytes. Immortalized lymphocyte cell lines are also being established for each patient. Genetic data are being generated for each patient including genotypes for genome wide SNPs (5M) and coding sequence and/or MLPA data for BMPR2, ALK1, ENG, CAV1 and SMAD9. Additional genes can be added for screening as identified. All biological samples, clinical data, as well as the SNP genotype and sequencing/MLPA data for the patients will be made available to the scientific community using a web-based application process. The samples and data will become available for request after the first 500 patients have been enrolled and the data are completed. The National Biological Sample and Data Repository for PAH represents an unparalleled collaboration between pulmonary hypertension centers in the United States to enable the collection of the largest cohort of PAH patients' biological samples in the United States. To date, 600 patients have been enrolled and sampled. This endeavor will provide the PAH research community an opportunity for here to now unprecedented hypothesis-driven studies. Funded by NHLBI HL105333.

1079F

Exome sequencing and genome-wide copy number variant mapping reveal novel associations with sensorineural hereditary hearing loss.
R.R. Haraksingh[1], F. Jahaniani Kenari[2], J. Rodriguez-Paris[3], J. Gelernter[4], K. Nadeau[5], J. Oghalai[6], I. Schrijver[3], M. Snyder[2]. 1) Psychiatry, Stanford University, Palo Alto, CA; 2) Genetics, Stanford University, Palo Alto, CA; 3) Department of Pathology, Stanford University School of Medicine, Stanford, California, USA; 4) Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 5) Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA; 6) Department of Otolaryngology - Head and Neck Surgery, Stanford University School of Medicine, Stanford, CA 94305, USA.

The genetic basis of hearing loss is not fully understood in terms of the diversity of loci and types of mutations that are responsible. We used multiple approaches, including exome sequencing of families and of probands with hearing loss, as well as copy number variation mapping in a case-control cohort, to identify loci associated with non-syndromic sensorineural hearing loss. Analysis of three distinct families revealed a novel gene candidate, MYH7B, associated with hearing loss in one family, and candidate loci in two other families. MYH7B encodes a Type II myosin, extending the role of cytoskeletal proteins in hearing. High-resolution genome-wide copy number variation analysis of 151 cases and 157 controls further revealed deletions in known hearing genes (e.g. GJB6, OTOA, and STRC, encoding connexin 30, otoancorin, and stereocillin, respectively), indicating that CNVs may be responsible for hearing loss. In addition, we found a novel association with hearing loss of a deletion on chromosome 16 containing the gene PDXDC1 (OR = 3.85, p = 1.45 × 10-7). Overall, our results indicate that a large number of loci, some of which are novel, and distinct types of mutations not typically tested for, may contribute to the etiology of hearing loss in humans.

1080W

Mutations in the BMP genetic network in patients with congenital GnRH deficiency.
D. Cassatella[1, 2], J. Liang[3], A. Dwyer[1], G. Sykiotis[1], H. Miraoui[1], C. Xu[1], S. Santini[1], V.A. Hughes[1], X.Z. Liu[3], P.M. Bouloux[5], M. Lang-Muritano[6], R. Quinton[4], J.G. Zhang[3], B.J. Stevenson[2], Y. Sidis[1], N. Pitteloud[1]. 1) Service of Endocrinology, Diabetology and Metabolism - Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 2) Vital-IT - Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) BGI-Shenzen, Shenzen, China; 4) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK; 5) Centre for Neuroendocrinology, Royal Free and University College School of Medicine, London, UK; 6) Department of Endocrinology and Diabetology, University Children's Hospital, Zurich, Switzerland.

The hypothalamic secretion of gonadotropin-releasing hormone (GnRH) is well-known as the "pilot light" of reproduction in all mammals. GnRH neurons arise in the olfactory placode during embryonic development and migrate to the preoptic area of the hypothalamus. Mutations in genes involved in GnRH neuron ontogeny, survival and/or function result in GnRH deficiency (congenital hypogonadotropic hypogonadism, CHH), characterized by absent puberty and infertility. The mechanisms controlling GnRH neuron specification are only partially known. Through human genetic studies, Fibroblast growth factor 8 (Fgf8) was identified as a strong morphogen for GnRH neuron specification. Bone morphogenetic protein 4 (Bmp4), which antagonizes Fgf8 signaling in various tissues, is expressed in the olfactory placode at E10.5. Heterozygous BMP4 mutations have been identified in severe myopia, polydactyly, cleft lip/palate, and renal dysplasia, phenotypes that are all associated with CHH. In addition, we identified a duplication of gremlin 2, an inhibitor of BMPs, in a CHH patient. Therefore, we hypothesized that mutations within the BMP genetic network could cause CHH by impairing GnRH neuron specification. After performing whole-exome sequencing in 36 CHH patients, we employed a multi-step bioinformatics strategy incorporating computational prioritization of the candidate genes, protein-protein interaction databases, and in silico functional predictions of variants. We found variants in 17 BMP pathway genes including ligands (BMP2 and BMP4), receptors (BMPR1B and BMPR2), inhibitors (NOG and FST), and enhancers (ENG and FBN1), and selected BMP4 as the best candidate to further study. In vitro studies revealed detrimental effects of BMP4 mutants on cell signaling/protein maturation. To strengthen the evidence that BMP4 is a novel locus for CHH, we are currently examining GnRH neuronal development in Bmp4 deficient mice. Studies of mutations in other BMP pathway genes are ongoing. In conclusion, we have identified mutations in the BMP genetic network among patients with CHH. Impaired BMP signaling may disrupt the cross-talk between FGFs and BMPs in the olfactory placode during a critical stage for GnRH neuron development.

1081T

Identifying rare, non-coding DNA variants in Systemic Lupus Erythematosus.
S.J. White[1], S. Cantsilieris[1], E.F. Morand[2]. 1) Monash Institute of Medical Research, Monash University, Australia; 2) Southern Clinical School, Monash University, Australia.

Systemic lupus erythematosus (SLE) is a debilitating multisystem autoimmune disease. Patients with SLE suffer a marked loss of life expectancy, and severe morbidity, due to autoimmune-mediated inflammation of multiple organs. A classic hallmark of SLE is the presence of autoantibodies. This is a clear biological link between SLE and antibody-producing B-lymphocytes (B-cells), strengthened by the clinical efficacy of anti-B cell therapies in certain patients. Genetic studies have associated SLE susceptibility with a number of Single Nucleotide Polymorphisms (SNPs), many of which are found in or near genes with B-cell functions. However, the known SLE SNPs together only account for a small percentage of the genetic contribution to SLE, and as most SNPs are located in non-coding DNA, the mechanism of their effect is often unclear. Several SLE-associated SNPs are located in such DNA elements that control gene regulation, with sequence changes disrupting the binding of specific transcription factors. In 2012 the ENCODE consortium published papers outlining genome-wide studies of open chromatin in a range of cell types. These studies demonstrated that DNaseI hypersensitive sites strongly correlate with known and predicted regulatory sequences. In addition, they are enriched for known and predicted transcription factor binding sites (TFBS). This provides a clear link between sequence variants and functional activity, and rare variants in B-cell regulatory loci (defined by open chromatin in B-cells) would explain genetic susceptibility to SLE in a way not detectable by GWAS. We undertook a pilot study to screen three B-cell regulatory elements previously associated with SLE. We identified a previously undescribed sequence variant in DNA from an SLE patient, predicted to have a significant effect on transcription factor (TF) binding. This finding validates our hypothesis that non-coding regulatory elements will contain private sequence variants affecting TF binding sites and chromatin structure, and may explain part of the 'missing heritability' issue in the post-GWAS era. We are currently performing additional mutation screening of putative B-cell regulatory loci in SLE patients, along with functional validation using reporter constructs and changes in DNaseI-sensitivity.

1082F

MTHFS Mutation May Contribute to Cerebral Folate Deficiency Syndrome.
H. Zhu[1], Y. Lei[1], R.H. Finnell[1, 2]. 1) Dell Pediatric Research Instutite, Department of Nutritional Sciences, UT Austin, Austin, TX; 2) Department of Chemistry and Biochemistry, College of Natural Sciences, UT Austin, Austin, Texas.

Cerebral Folate Deficiency (CFD) syndrome is characterized by very low concentration of 5-methyltetrahydrofolate (5-MTHF) in the patient's cerebrospinal fluid (CSF), while folate levels in plasma and red blood cells are within normal limits. Major clinical features of CFD syndrome include unrest, irritability, and insomnia, decelerating head growth, neurodevelopmental delay, regression, hypotonia, ataxia, dyskinesias, spasticity, speech difficulties, and seizures. It has been speculated that the etiology of CFD syndrome may be explained by: disrupted folate transport through the blood brain barrier (BBB), autoantibodies against folate receptor alpha (FOLR1) impeding transport, mutations in folate transport and metabolism genes, as well as mitochondrial deficiency. Folate genes that have been implicated in CFD syndrome include: *FOLR1*, *DHFR* (dihydrofolate reductase) and *PCFT* (proton coupled folate transporter). We performed exome sequencing analysis of the family (proband, parents and a healthy sibling) of a CFD patient, hoping to identify causal mutation(s) that contribute to the patient's conditions. Genomic DNA was extracted from whole blood using the Gentra Puregene Blood Kit (Qiagen). Whole exome libraries were established using Agilent SureSelect Kit v4.0 (3'UTR included) and submitted to sequencing on Illumina HiSeq 2000. Sequencing data was analyzed using NEXTGENe software (Softgenetics). Ninety-nine recessive novel mutations were identified through the exome sequencing, of which 21 were predicted to be damaging by the PolyPhen program. No de novo mutations were detected in the patient. The patient carries a homozygous mutation, c.101G>T (p.R34L, rs200058464), in a folate pathway gene, 5,10-methenyl-tetrahydrofolate synthetase (MTHFS). Both parents are heterozygotes, and the healthy sibling of the patient does not carry the mutant allele. After confirming the mutation using Sanger sequencing, we performed in vitro functional assays by transfecting the wild type and mutant alleles into HEK293T cells. The mutant allele (T) showed more than 50% reduction in luciferase activity, indicating the functional effect of this MTHFS mutation. Our study suggested that exome sequencing analysis of the CFD family is an efficient way to identify potential candidate genes and may lead to the discovery of causal mutation(s) responsible for this clinical disorder. (Supported in part by NIH grants HD067244 and NS076465).

1083W

Pilot whole genome sequencing of germline DNA from 186 breast cancer cases. *P. Kraft[1], J. Allen[2], C. Chen[1], B. Decker[2], J. Figueroa[3], S. Hart[4], S. Lindstrom[1], J. Long[5], M. Yeager[3], S. Chanock[3], F. Couch[6], D. Easton[2], C. Haiman[7], W. Zheng[5], D. Hunter[1].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK; 3) Division of Cancer Epidemiology and Genetics, NCI, Rockville, MD; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 5) Division of Epidemiology, Vanderbilt University School of Medicine, Nashville, TN; 6) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 7) Department of Preventive Medicine, USC Keck School of Medicine, Los Angeles, CA.

High coverage whole genome sequencing (WGS) has been proposed to identify rare germline mutations associated with complex traits in both coding and regulatory regions. However, WGS studies present many logistical and analytic challenges, including coordinating variant calling and cleaning across multiple sites, accruing sufficient sample size, and designing and interpreting appropriate statistical analyses. We illustrate some of these challenges using a pilot whole genome sequencing study of 143 European-ancestry (EA), early-onset, family-history-positive breast cancer cases, 21 Asian cases and 25 African-American cases from six studies participating in the DRIVE consortium, an NCI-sponsored post-GWAS initiative. Samples from each study were sequenced separately using Illumina HiSeq to an average depth of 30x and called individually using CASAVA. Preliminary analyses identified a missense *BRCA1* mutation present in three EA cases that was absent in 4,300 EA subjects in the Exome Sequencing Project ($p=6.8\times10^{-5}$). A nonsense variant in another gene was present in 11 cases and 0 ESP subjects ($p=5\times10^{-16}$), but this apparent association was driven by false negative calls in the ESP--the average read depth at this position in the ESP was 1. We describe and present results from a rare-variant burden that uses summary data from individual studies. We discuss future plans, including combined variant calling and analysis of non-coding variants.

1084T

Exome Sequencing in Hodgkin Lymphoma Families. *M. Rotunno, M. McMaster, L. Goldin.* NIH/NCI/DCEG, Bethesda, MD.

Hodgkin lymphoma (HL) shows strong familial aggregation but no major susceptibility genes for HL have been identified to date. Studies based on exome sequencing are promising for identifying disease susceptibility rare genetic variants. The goal of this study was to identify high-penetrance variants in HL-prone families. Using Nimblegen v2.0 and v3.0 exome capture array and the Illumina HiSeq2000 sequencer, we exome sequenced 45 HL cases or obligate carriers from 11 HL-prone families with three or more affected members and 2 families with two HL cases. Reads were aligned using Novoalign v.2.07.14 and variants were called using GATK software. Possible technical sequencing artifacts were eliminated by filtering out variants found in more than 1% of samples from other studies identically processed and sequenced in our laboratory. The Ingenuity Variant Analysis software was used to analyze, annotate and prioritize the resulting exome data. Only rare (<1%; frequency in European populations), non-synonymous variants were kept. According to a dominant segregation model, between 2 and 59 variants were shared by all cases or obligate carriers within each family with 3 or more members. No variants were shared by more than one family. Six variants occurred in 3 genes shared by two families each: FSTL5, SMTNL2, and ZFYVE28. We plan to perform targeted sequencing of these genes in additional HL families. In addition, genes from individual families were also prioritized for further follow-up based on damage prediction and conservation of the shared variant, and literature link to cancer or immune related processed. Further study is needed to determine whether rare variants in the identified genes increase risk for familial HL.

1085F

The burden of coding, non-coding and chromosomal mutations in Hirschsprung disease. *T. Turner[1], K.D. Nguyen[1], N. Krumm[2], S. Chatterjee[1], A. Kapoor[1], Q. Jiang[1], A.Y. Ling[1], M.X. Sosa[1], N. Gupta[3], E.E. Eichler[2], S. Gabriel[3], C. Berrios[1], A. Chakravarti[1].* 1) Center for Complex Disease Genetics, McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Genome Sequencing and Analysis Program, Broad Institute of MIT and Harvard, Cambridge, MA.

Hirschsprung Disease (HSCR) affects ~1/5000 live births and is characterized by the failure of gut innervation. This multifactorial neurodevelopmental disorder has variable recurrence risks depending on the proband's sex, familiality and segment length of aganglionosis. 15 genes with rare, deleterious mutations in HSCR, most commonly at the receptor tyrosine kinase gene *RET*, are known. Although common functional variants in two enhancers of *RET* exist in ~90% of HSCR cases and all known rare coding functional variants in ~10% of patients, they cumulatively explain <10% of the risk variation. To understand the total burden of genomic variation, we performed exome capture (Agilent 44Mb) and sequencing on an Illumina platform in 304 patients representing the gamut of HSCR variability. We identified 90,818 high quality coding and splicing single nucleotide variants (SNVs) using GATK, and 146 rare (<1% in ESP) copy number variants (CNVs) using CoNIFER. Our major analyses have focused on 189 unrelated, European ancestry, non-admixed individuals and >361 analogous controls from dbGAP, ESP and 1000 Genomes. We focused on genes with 3 or more putative highly deleterious coding variants (NMD nonsense, donor/acceptor intronic splice, conserved (phyloP≥4) missense) in cases and identified genes, expressed in the gut, for which this distribution was significantly different ($P<10^{-4}$) than in controls. This analysis identified 64 genes with deleterious alleles at significantly higher frequencies as compared to the EVS. Among these, three genes are known to underlie HSCR in humans (*RET*, *EDNRB*, *SEMA3D*), one has dosage effects in patients (*NAV2*), and one causes aganglionosis in the mouse (*PAX3*). The novel genes show significant enrichment of genes in neuronal and/or ubiquitination pathways (22 or 37%) adding these to known HSCR defects in enteric neural crest cell differentiation, mesenchymal function and axonal guidance. We found 11 CNVs, at >10-fold increased frequency, at loci previously identified with recurrent deletion/duplication syndromes, such as 16p11.2, 1q21.1 and 17p11.2, suggesting a broader role for these loci in neuronal disease. Further analysis of these data show that HSCR can be sub-divided into four types arising from the effects of single genes, recurrent deletion/duplication loci, multi-organ syndromes of 'multifactorial' etiology and 'multifactorial' isolated cases, with effects, on average, from ~5 deleterious variants.

1086W

Analyses of WES data in multiplex Syrian non-syndromic oral clefts families. *J. Bailey-Wilson[1], E. Holzinger[1], M. Parker[2], S. Szymczak[1], Q. Li[1], C. Cropp[1], M. Nöthen[3], J. Hetmanski[2], H. Ling[4], E. Pugh[4], P. Duggal[2], M. Taub[5], I. Ruczinski[5], A. Scott[4], M. Marazita[6], H. Albacha-Hejazi[7], E. Mangold[3], T. Beaty[2].* 1) Inherited Disease Res Branch, NIH/NHGRI, Baltimore, MD, USA; 2) Department of Epidemiology, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA; 3) Institute of Human Genetics, University of Bonn, Bonn, Germany; 4) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD, USA; 5) Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA; 6) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 7) Ibn Al-Nafees Hospital, Damascus, Syrian Arab Republic.

Oral clefts (cleft lip, cleft palate and cleft lip & palate) are common birth defects with a complex and heterogeneous etiology. Some genes and chromosomal regions have been associated with risk in genome wide association and linkage studies. This whole exome sequencing (WES) study used 22 affected 2nd degree or more distant relatives drawn from 10 multiplex inbred families (2 families with 3 relatives and 8 families with 2 relatives) initially ascertained in the Syrian Arab Republic for linkage studies. WES was done by the Center for Inherited Disease Research using the Agilent SureSelect v.4 capture reagents & Illumina HiSeq 2000 sequencers. Variants were called for all samples together within this project using Unified Genotyper (2.3-9). Variants were flagged by VQSR annotation using a Gaussian Mixture model for both SNVs and INDELs. Random Forests was used to estimate probability of high versus low quality calls. Results of analyses using the Ingenuity software on these newly recalled data will be presented. Additional sequencing studies of more families and more affected individuals in these families are ongoing to determine which genes segregate with oral clefts in these Syrian families.

1087T

Multigenic inheritance as a cause of familial congenital diaphragmatic hernia.
T.F. Beck[1], P.M. Campeau[1], J.T. Lu[2,3], C. Gonzaga-Jauregi[1], J.R. Lupski[1,4], R.A. Gibbs[2], B.H. Lee[1,5], W. Reardon[6], D.A. Scott[1,7], Centers for Mendelian Genomics. 1) Molec & Human Gen, Baylor College Med, Houston, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Structural and Computational Biology & Molecular Biophysics, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Howard Hughes Medical Institute, Houston, TX; 6) Our Lady's Hospital for Sick Children, Crumlin, Dublin 12, Ireland; 7) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

Congenital diaphragmatic hernia (CDH) is a common life-threatening birth defect which can present either alone or as part of a genetic syndrome. Isolated CDH has a low recurrence risk (1 to 2%) with most cases being caused by a combination of genetic and environmental factors. We report a male patient with isolated CDH whose father was also diagnosed with CDH at age 8 and exhibited severe scoliosis, low intellectual achievement, and unusual windswept hands. Both father and son had normal G-banded chromosome analyses. In an effort to identify the cause of CDH in this family, genomic DNA samples from the patient and his father were subjected to whole-exome sequencing analysis. This analysis revealed a frameshift mutation (c.4970_4971insA → p.Cys1658Leufs*29) in the fibrilin 1 gene (*FBN1*) which is associated with Marfan syndrome~a connective tissue disorder with a range of associated phenotypes, including lens dislocation, aortic dilatation, and skeletal anomalies. Sanger sequencing confirmed this mutation in father and son, as well as an asymptomatic male sibling who also shares his father's unusual windswept hands. An ophthalmological evaluation revealed iridodonesis in both father and proband confirming the diagnosis of Marfan syndrome. The father also exhibited phacodonesis. A cardiovascular evaluation has been scheduled. Although CDH has been reported in a handful of Marfan cases, it is unusual for CDH to be the presenting symptom and to reoccur in more than one family member. With this in mind, we looked for mutations in other CDH-related genes which might account for the high prevalence of CDH in this family. Putatively damaging sequence changes were identified in the *FRAS1*-associated extracellular matrix protein 1 (*FREM1*), desmin (*DES*), and hepatocyte growth factor receptor (*HGFR*) genes in both the father and his affected son. These findings underscore the importance of considering the potential influence of deleterious changes in other genes as an explanation for rare and/or unusual phenotypes associated with common genetic syndromes.

1088F

Exome seqencing identified novel genetic mutations in the patients with congenital vertebral anomalies.
Y. Nakamura[1], S. Kikugawa[2], S. Seki[3], M. Takahata[4], H. Terai[5], Y. Akaoka[6], M. Matsubara[6], F. Fujioka[6], H. Inagaki[7], H. Kurahashi[7], T. Kobayashi[8], H. Kato[1]. 1) Shinshu University School of Medicine, Matsumoto, Japan; 2) DNA chip institute, Kanagwa, Japan; 3) Toyama Medical University, Toyama, Japan; 4) Hokkaido Medical Univesiry, Sapporo, Japan; 5) Osaka City Medical University, Osaka, Japan; 6) Nagano Prefectual Chilcren's Hospital, Azumino, Japan; 7) Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan; 8) Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

[Purpose] Congenital Vertebral Anomalies (cVA) affect 0.5-1/1000 children and show a significant high mortality. Causes of most cVA remain largely unknown. The purpose of this study is to identify genetic abnormalities in Japanese cVA patients. [Methods] We focus on primary cases with cVA that affect more than 1 vertebral body with malformation. Cases with congenital scoliosis with rib malformation, Klippel Feil syndrome, or VACTERL association were included in the patient cohort. We first performed exome sequencing on 8 patients and their 9 unaffected family members using the Agilent SureSelect Human All Exon 50Mb Kit and the illumina HiSeq Sequencer. Reads were mapped to the hg19 reference genome with Burrows-Wheeler Aligner and variant calling was performed by The Genome Analysis Toolkit. We excluded mismatches that did not change amino-acid sequences and known SNPs whose allele frequencies were greater than 1% according to the dbSNP JPT and 1000 Genompe project. Approximately 700 such mismatches were found in each patient. We then confirmed the data by Sanger sequencing. [A summary of results] We have found de novo mutations in the DCLRE1B, AGBL5, PDE2A, OLFML1, and ASB16 genes. All of these mutations are non-synonymous, and there has been no report of any mutations based on dbSNP and 1000 genome databases. DCLRE1B is involved in DNA interstrand cross-link repair; the major role of AGBL5 is tubulin deglutamylation; PDE2A is a phosphodiesterase involved in many signal transduction pathways; the function of OLFML1 is unknown. Although the function of ASB16 is mostly unclear, based on the sequence similarities and its domain structure, ASB16 may be a component of an E3 ubiquitin-protein ligase complex. Interestingly, a possible de novo non-sense mutation was found at a different position of the ASB16 gene in another patient. We are currently performing exome sequencing analyses on additional 4 families. We are also planning to create genetically-modified zebrafish models to understand the in vivo functions of those genes. [Conclusion] These findings suggest that cVA are genetically heterogenous. Our study identified several potentially disease-causing genes of cVA.

1089W

Whole-genome sequencing of individuals from a founder population identifies candidate genes for asthma. *C.D. Campbell[1], K. Mohajeri[1], M. Malig[1], F. Hormozdiari[1], B. Nelson[1], G. Du[2], K. Patterson[2], C. Eng[3], D.G. Torgerson[3], J.X. Chong[2], A. Ko[1], L. Vives[1], B.J. O'Roak[1], M. Abney[2], E.G. Burchard[3], C. Ober[2], E.E. Eichler[1,4].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Genetics, University of Chicago, Chicago, IL; 3) Department of Medicine, University of California San Francisco, San Francisco, CA; 4) Howard Hughes Medical Institute, Seattle, WA.

Asthma is a complex genetic disease caused by a combination of genetic and environmental risk factors. We sought to test classes of genetic variants largely missed by whole-genome association studies (GWAS) by performing whole-genome sequencing of 16 individuals from asthma-enriched and depleted families from an extended 13-generation Hutterite pedigree with reduced genetic heterogeneity due to a small founding gene pool and reduced environmental heterogeneity as a result of a communal lifestyle. We sequenced each individual to an average depth of 13-fold (Illumina HiSeq), generated a comprehensive catalog of genetic variants, and tested for association with asthma in 1200 individuals from the same population. Specifically, we identified and validated 2881 copy number variants (CNVs), 19 gene-disruptive single nucleotide variants (SNVs), and 18 frameshifting insertions or deletions ('indels'). We genotyped 593 of the CNVs in 1199 individuals, including 164 individuals with asthma, 488 controls, and 547 with intermediate phenotypes, from the extended pedigree. We identified a nominally significant association (p=0.03; Odds ratio=3.13) between a 6 kbp deletion in an intron of *NEDD4L* and increased risk of asthma. We genotyped this deletion in 742 cases and 755 controls of Puerto Rican ancestry and observed this variant in two cases and zero controls. Interestingly, *NEDD4L* is an ubiquitin ligase expressed in bronchial epithelial cells, and conditional knockout of this gene in the lung in mice leads to severe inflammation and mucus accumulation. This gene resides under a linkage peak for asthma in the Hutterites making it a plausible candidate. Additionally, this deletion is not in strong linkage disequilibrium with any particular SNP genotyped by the HapMap or 1000 Genomes Projects and would have been missed by GWAS. Of the 37 gene-disruptive SNVs and indels, we observed a nominally significant association with asthma for six variants in the larger cohort, which have allele frequencies between 1-6% in the Hutterites. We have extended these results by resequencing 16 genes in 837 cases and 540 controls of Puerto Rican ancestry in order to test for a burden of rare protein-altering mutations. Our study represents one of the early instances of applying whole-genome sequencing to complex disease with a large environmental component, and we have used these data to assess genetic variation largely untested in GWAS, including CNVs and low-frequency variants.

1090T

Cluster detection approaches to identify disease genes in CNVs implicated in psychiatric disorders: applications to whole-exome sequencing studies on autism and schizophrenia. *I. Ionita-Laza[1], B. Xu[2], V. Makarov[1], J. Buxbaum[3], J. Louw Roos[4], J. Gogos[5,6], M. Karayiorgou[2].* 1) Biostatistics, Columbia University, New York, NY; 2) Department of Psychiatry, Columbia University, New York, NY; 3) Department of Psychiatry, Mount Sinai School of Medicine, New York, NY; 4) Weskoppies Hospital, Pretoria, South Africa; 5) Department of Physiology & Cellular Biophysics, Columbia University, New York, New York, USA; 6) Department of Neuroscience, Columbia University, New York, New York, USA.

Advances in next-generation sequencing technologies make possible to exhaustively explore the genetic influence of rare variants, both point and structural, on the risk of developing complex psychiatric and neurodevelopmental disorders. A number of studies have highlighted a significant role for CNVs (both de novo and inherited) as a constant source of pathogenic variation in psychiatric disorders. Most of the identified CNVs span multiple genes and highlight regions in the human genome likely containing susceptibility genes. It is thought that it is the imbalance of several affected genes that determines the overall phenotype associated with a given CNV although one or a few genes may have a greater phenotypic impact. We propose new cluster detection methods based on scan statistics, for both family-based and case-control designs, to identify significant clusters of rare disease risk variants, corresponding to disease susceptibility genes. Unlike conventional gene-based association tests, these cluster detection tests take into consideration the spatial aggregation of risk variants into a small region, e.g. a gene, within a larger CNV. Using simulation studies based on exome-sequencing data, we show that the proposed cluster-detection methods are more powerful than conventional gene-based tests when variants cluster significantly within a gene in a larger CVN. We present results from several exome sequencing studies on autism and schizophrenia, using both family-based and case-control designs, and identify in each case statistically significant associations with rare nonsynonymous variants within genes for several well-known CNVs, including 15q13.3. Notably, the genes involved highlight new mechanisms in these disorders. Furthermore, on these data, only the cluster detection tests allow for robust statistical findings, while the conventional tests result in findings that are not statistically significant after multiple testing adjustment, highlighting the utility of the cluster detection approaches for the identification of susceptibility genes in larger CNVs.

1091F

Identification and characterization of the first OSBPL1A mutations in individuals with low plasma HDL-C levels. *M.M. Motazacker[1], H. Kentala[2], J.A. Kuivenhoven[3], A.W. Schimmel[1], Y. Zhou[2], J. Pirhonen[2,4], E. Ikonen[2,4], G.M. Dallinga-Thie[1], G.K. Hovingh[1], M. Jauhiainen[5], V.M. Olkkonen[2,4].* 1) Department of Vascular Medicine, Academic Medical Center, Amsterdam, Netherlands; 2) Minerva Foundation Institute for Medical Research, , Helsinki, Finland; 3) University Medical Center Groningen, Department of Molecular Genetics; 4) University of Helsinki, Institute of Biomedicine and Anatomy, Helsinki, Finland; 5) National Institute for Health and Welfare, Public Health Genomics Unit, Helsinki, Finland.

Oxysterol-binding protein-like proteins (OSBPLs) have regulatory or sterol-transfer functions in eukaryotic cells from yeast to human. However, the detailed mechanisms of their action and their contribution in human phenotypes have remained elusive thus far. In search of variants involved in extreme HDL-C phenotypes, we sequenced ~200 lipid associated genes including several members of OSBPL gene family in individuals with low or high plasma HDL-C levels. Results: We identified 2 missense (p.R186W and p. Q459P) and 1 indel (p.C39X) variant in OSBPL1A in 3 unrelated individuals with low HDL-C levels. All mutations involved highly conserved nucleotides (PhastCons score =1) and were predicted to be deleterious to protein structure/function. Families of the index patients were expanded and even after excluding index patients, plasma HDL-C levels were significantly lower in carriers of the variants compared to non-carrier relatives ($p<0.05$) while there was no difference in LDL-C, TG or Total cholesterol levels between the two groups. Immunoprecipitation experiments in HuH7 cells showed a lack of binding to Rab7 -GTPase controlling membrane trafficking between early/late endosomes and lysosomes- for the p.C39X variant. We then studied subcellular localization of GFP-tagged OSBPL1A by overexpressing wild-type and mutant proteins in HuH7 cells. Unlike WT protein which was mainly detected on late endosomal compartments, p.C39X variant showed a cytosolic distribution and p.R186W variant showed enhanced association with endoplasmic reticulum and nuclear membrane. The p.Q459P variant was localized on late endosomes however these compartments were smaller in size and more frequent in number compared to WT protein. Conclusions: We have identified the first OSBPL1A mutations in individuals with low plasma HDL-C levels. The variants have deleterious effect on localization of the protein and can potentially influence endosome motility and function. An anticipated relation between disturbed endocytic pathway function, cholesterol transport and low plasma HDL-C levels is currently being investigated.

1092W

Whole exome sequencing to identify variants influencing both pre-diabetic traits and type 2 diabetes mellitus in Pima Indians. *L.J. Baier, K. Huang, A. Nair, Y.L. Muller, M. del Rosario, S. Kobes, R.L. Hanson, W.C. Knowler, C. Bogardus.* PECRB, NIDDK/NIH, Phoenix, AZ.

Genes associated with type 2 diabetes (T2D) have been previously identified, but the causative variant underlying the association and the physiologic mechanism whereby the gene influences T2D risk are often unknown. The goal of this study is to identify coding variation that increases susceptibility to T2D via its effect on a pre-diabetic trait. Exome sequencing was done on 177 Pima Indians and 148,616 variants (11.1% novel) were detected that met our quality control criteria. Of these variants, 15,799 and 15,642 were non-synonymous and synonymous SNPs, respectively. Selected variants (N=360) were genotyped in 555 non-diabetic Pima Indians characterized as inpatients in our Clinical Research Center for % body fat (PFAT), central obesity (waist/thigh), glucose disposal rates (M) during a hyperinsulinemic-euglycemic clamp, acute insulin response (AIR) to intravenous glucose, and 2-hour plasma glucose concentrations (2-hr glucose) during an OGTT. Variants were further assessed for association with BMI and T2D in up to 7,667 subjects who had participated in a longitudinal study of T2D in the Gila River Indian Community. Two SNPs (in the genes *ASXL3* and *CYB5A*) were significantly associated with a pre-diabetic trait after correction for multiple testing (360 SNPs analyzed for 5 traits required a $p<2.7\times10^{-5}$). The SNP in *ASXL3* (rs2282632) was associated with insulin sensitivity (M; $p = 2.0\times10^{-5}$, adjusted for age, sex, PFAT and nuclear family membership) but was not associated with T2D. In contrast, the SNP in *CYB5A* (rs7238987) was associated with PFAT ($p=6.7\times10^{-6}$) in 555 non-diabetic subjects, BMI measured at a non-diabetic exam in 5880 subjects ($p=6.2\times10^{-7}$) and T2D in 7667 subjects (OR=1.14 [1.03-1.14]; $p=9.2\times10^{-3}$). Variants in 5 other genes (*RNF10*, *BBS12*, *ACACB*, *ZNF530* and *HNF4A*) had comparable associations with T2D ($p<9.0\times10^{-3}$), but their associations with pre-diabetic traits were weaker (not significant after adjustment for multiple testing). In summary, *CYB5A* provided the strongest evidence for a pre-diabetic loci. It encodes a membrane bound microsomal hemoprotein which acts as an electron carrier for the stearoyl-CoA-desaturase (SCD) complex facilitating the conversion of saturated fatty acid to mono unsaturated fatty acid. Prior studies have reported that SCD activity is associated with obesity. We conclude that exome sequencing is a viable technique for identifying new loci that increase risk for T2D via an influence on a pre-diabetic trait.

1093T

Whole-exome sequencing of 4,000 samples identifies rare variants strongly associated with type 2 diabetes risk in Mexicans and Latinos. *K. Estrada[1,2,3] for the SIGMA T2D Consortium.* 1) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 3) Harvard Medical School, Boston, MA. USA.

Type 2 diabetes (T2D) is more prevalent in Latino populations than in people of European descent. We investigated the hypothesis that this increased T2D risk in the Latino population may be driven by rare, and possibly population-restricted, protein-coding variants that have remained undetected by previous studies focusing on common genetic variation.

We performed the largest ever whole-exome sequencing project of Mexicans and Latinos totaling 3,946 samples equally distributed between T2D cases and controls.. Samples were selected based on high Native American ancestry from a pool of 9,225 well-characterized individuals who had been genotyped with OMNI 2.5 and ExomeChip arrays. After stringent quality control 3,792 samples were available for analysis and included ~1.2M variants.

The most significant locus identified pointed to a novel common variant in the 5'UTR of SLC16A11 (minor allele frequency [MAF]=36.7%, $P_{EMMAX}=1.7\times10^{-10}$, OR=1.3). Our prior GWAS identified missense variants in this gene associated with T2D (in review). The 5'UTR variant is rare in European and African populations (MAF <2%) but common (MAF = 28%) in Mexicans of the 1000G project.

The second most significant signal pointed to a novel non-synonymous rare variant (control MAF = 0.2%, case MAF = 1%, $P_{EMMAX}=4.4\times10^{-7}$, OR=3.8). This variant seems to be largely private to the Latino population: we only observed 2 carriers in a collection of 26,000 non-Latino exomes of mainly Caucasian ancestry. This variant has previously been reported as dominant causal for a mild version of *maturity-onset diabetes of the young* (MODY) in two cases. However, our data indicate that this variant confers a large-effect (but not fully penetrant) risk for T2D present in 0.2% of our controls with age range of 45-73.

Ongoing work includes 1) imputation of our non-sequenced samples using a reference panel based on those samples with both array and sequence data 2) collapsing and other burden tests and 3) exome analysis of severe Mendelian forms of T2D such as MODY and neonatal diabetes. Our results suggest that previously uncharacterized rare population-specific genetic variants contribute to the high burden of T2D in the Mexican population.

1094F

A test of association of genome-wide coding variation with type 2 diabetes in 13,000 individuals from five ancestry groups. *P. Fontanillas[1], N. Burtt[1], P. Cingolani[2], J. Flannick[1], K. Gaulton[3], H. Highland[4], A. Mahajan[3], A. Morris[3], M. Rivas[3], X. Sim[5], T. Teslovich[5] on behalf of T2D-GENES and GO-T2D Consortia.* 1) Prog Medical & Population Genetics, Broad Institute, Cambridge, MA; 2) McGill University and Génome Québec Innovation Centre, Montréal, Canada; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 4) Human Genetics Center, University of Texas Health Science Center, Texas, USA; 5) Department of Biostatistics and Center for Human Genetics, University of Michigan, Ann Arbor, MI.

We sequenced the coding regions of 18,162 genes in ~13,000 individuals (6,562 cases and 6,459 controls) to evaluate the contribution of coding variation to risk of type 2 diabetes (T2D). To maximize genetic diversity, samples were chosen from five ancestry groups: African American, East Asian, European, Mexican American, and South Asian. Of 3,473,514 SNPs and short INDELs identified, 2,038,377 were in exons and the remainder within 50bp of exons in introns or UTRs. Most variants had low minor allele frequency (MAF) (96%<1% and 88%<0.1%) and were specific to one ancestry group (79%).

We tested for single-variant association under an additive genetic model using meta-analysis. For common variants (MAF>5%), sequencing revealed no new coding variant association signals that were exome-wide significant ($p<2.5\times10^{-6}$). Eight variants in seven genes showed association with T2D at $p<10^{-4}$ [odds ratios (OR) <1.26]. The four known GWAS T2D lead variants located in coding regions (in *MACF1*, *PPARG*, *SLC30A8*, and *KCNJ11*) all had association p>10-4. For low-frequency variants (1% <MAF<5%), we demonstrated T2D association for an ancestry-specific coding variant in *PAX4*, a transcriptional regulator of β-cell development ($p=6.4\times10^{-9}$, OR=1.81 [1.48-2.17]). Three others variants (in genes *ATAT1*, *NT5E*, and *ZNF283*) showed p<10-4 and OR<1.57. For rare variants (0.1%<MAF<1%), 17 variants were observed with $p<10^{-4}$ and 2.16<OR<5.26. A variant in *FES* ($p=1.3\times10^{-5}$, OR=4.24 [2.09-12.04]) was of particular interest because of its location in a known T2D GWAS region (near *FURIN-PCR1*); it may be the transcript mediating the association signal at this locus. Gene-based burden analyses on rare variants (MAF<1%) using collapsing approaches identified few genes with T2D association signals ($p<10^{-4}$), in particular *ASCM3*, an acyl-Co A synthetase, which showed an enrichment of coding deleterious variants in T2D cases.

Our 13,000 sample multi-ancestry design was well-powered (80%) to detect exome-wide significant single variant association for low-frequency (1%<MAF<5%), and medium-large effects (OR>2.5). While some rare variant T2D association signals may emerge from our analyses, the hypothesis that low-frequency variants with large effects could be a main feature of T2D architecture is not supported by our results.

1095W

Whole Genome Sequencing to Identify Variants that Influence Pre-diabetic Traits in American Indians. *K. Huang, P. Piaggi, S. Kobes, R. Hanson, C. Bogardus, L. Baier.* Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, AZ.

To identify genetic variation that influences pre-diabetic traits and thereby increases risk for type 2 diabetes (T2D), we obtained whole genome sequence data on 235 Pima Indians (127 men, age: 24.8±5.4 years, BMI: 33.4±7.4 kg/m2). These individuals had been metabolically characterized for the following traits when they were non-diabetic: BMI, % body fat, central obesity (waist/thigh), insulin-stimulated glucose disposal rate during a hyperinsulinemic-euglycemic clamp, acute insulin response to 25 g intravenous glucose bolus, and 2-hour plasma glucose concentrations during an oral glucose tolerance test. A subset of these subjects had also been profiled for gene expression levels (69 had expression data from a subcutaneous adipose biopsy and 90 had expression data from a skeletal muscle biopsy). Sequencing was performed by Illumina (N=200) and Complete Genomics, Inc (N=35). ~12 million variants were found, including ~10.2 million SNPs, ~1.3 million Indels and 255,802 substitutions. Association analysis between variants and pre-diabetic traits is ongoing, as are association analyses between variants and gene expression data to identify cis-acting variants. Variants with the strongest associations, as well as variants predicted to be damaging that map near candidate genes, will be genotyped in a sample of 555 Pima Indians who also have phenotypic information on pre-diabetic traits. Association with T2D will be assessed by genotyping in up to 7,667 American Indians who are part of a longitudinal study and have data on T2D status and BMI.

1096T

Whole exome sequencing identifies PAX4 nonsynonymous variant as susceptibility loci for type 2 diabetes in Koreans. *S.H. Kwak[1], J.I. Kim[2], K. Kim[3], Y.M. Cho[1], H.S. Jung[1], Y.J. Park[1], K.S. Park[1].* 1) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea; 2) Department of Biochemistry and Molecular Biology Seoul National University College of Medicine, Seoul, Korea; 3) Department of Statistics, Sookmyung Women's University, Seoul, Korea.

Type 2 diabetes is a common complex disorder with strong genetic predisposition. Although more than 65 common genetic variants of diabetes have been identified so far, their effect sizes are small and they explain only a limited part of the heritability. In this study we used whole exome sequencing to identify low frequency, functional variants of type 2 diabetes. This was a case-control analysis using 324 confirmed type 2 diabetes patients and 101 carefully selected normal glucose tolerant elderly subjects without family history of diabetes. We limited our diabetes subjects to those who had at least one first degree relative of diabetes in their family. Whole exome capture was prepared using Agilent SureSelect version 4 + UTR and sequencing was performed by Illumina HiSeq 2000. Sequence alignment was done using BWA and Picard software and variant identification was done using GATK software. EPACTS software was used for rare variant association testing. A nonsynonymous variant in PAX4 (rs2233580, R192H) was associated with risk of diabetes in near exome-wide significance of P= 9.66E-05 with minor allele frequency (MAF) in diabetes subjects 0.122 and controls 0.030. We then compared diabetes subjects having the rs2233580T variant of PAX4 gene (N=77) and diabetes subjects not having this variant (N=247). Those with the variant had lower body mass index (23.3±2.5 vs. 24.0±2.5, P=0.037) and waist circumference (83.0±7.3 vs. 84.7±6.7, P= 0.047) with nominal significance. Another variant located in intron of SORBS3 (rs12680280) was significantly associated with risk of diabetes with MAF 0.017 in diabetes subjects compared to 0.099 in control subjects (P=2.15E-08). However, these variants require further replication and we are currently genotyping these variant in another set of 1,387 diabetes cases and controls and 821 gestational diabetes women. In conclusion, we have identified low frequency nonsynonymous variant in PAX4 to be associated with risk of diabetes in Koreans.

1097F

Large-scale exome chip association analysis identifies rare and low-frequency coding variants associated with glycemic traits. *A. Mahajan[1], X. Sim[2], A.K. Manning[3], M.A. Rivas[1], N. Grarup[4], H.K. Im[5], H.M. Highland[6], A.E. Locke[2], P. Fontanillas[3], T.M. Teslovich[2], J. Flannick[3], C. Fuchsberger[2], K. Gaulton[1], H.M. Kang[2], A.P. Morris[1], J.B. Meigs[7], C.M. Lindgren[1] for T2D-GENES and GO-T2D Consortia.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 4) Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, DK; 5) Department of Health Studies, University of Chicago, Chicago, USA; 6) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 7) Massachusetts General Hospital, Boston, Massachusetts, USA.

The extent to which low-frequency (LF; minor allele frequency (MAF) 1-5%) and rare (MAF <1%) coding variants contribute to variability of glycemic traits has not been systematically evaluated to date. To discover novel coding variants associated with fasting plasma glucose (FG) and fasting insulin (FI), and examine whether LF and rare coding alleles could explain established genome-wide association (GWA) signals for these traits, we studied up to 33,553 and 31,149 non-diabetic individuals of European ancestry for association with FG and FI respectively. Samples were genotyped using the Illumina HumanExome Beadchip array, which collectively represents >80% coding variation with >0.5% MAF in European ancestry populations. We tested variants for association with FG and FI using a linear mixed model to account for relatedness. We also carried out gene-based tests using the sequence kernel association optimal (SKAT-o) test. We then combined summary statistics at up to 135,904 high-quality autosomal variants and 14,667 genes across studies by meta-analysis. We identified one rare non-synonymous variant associated with FI at chip-wide significance (P= 3.7×10^{-7}) in a gene not mapping to an established locus: *URB2* (P= 2.75×10^{-7}; 0.1% MAF; E594V). Two novel non-synonymous variants (one rare and one LF) were also associated with FG, approaching chip-wide significance in *TP53BP1* (P=6.08×10^{-7}; 0.2% MAF; T1278I) and *GLP1R* (P=7.64×10^{-7}; 1.5% MAF; A316T). Of these, *GLP1R* encodes glucagon-like-peptide-1 receptor, and therefore transduces incretin signaling in pancreatic beta cells. We also identified multiple rare coding variants in *G6PC2*, that in aggregate were associated with FG (SKAT-o P=5.31×10^{-13}), which reside within an established GWA locus. Furthermore, conditional single variant analysis confirmed the presence of two independent association signals in this gene, one common (P_{COND}=2.28×10^{-15}; 48.2% MAF; V219L) and one rare (P_{COND}=1.23×10^{-9}; 0.8% MAF; H177Y), neither of which can explain the GWA signal at this locus (lead SNP rs560887). They are also independent of nearby promoter variants (r^2<0.2) identified in a recent study to be potentially functional. In conclusion, these results provide evidence that LF and rare coding variants contribute to variability in glycemic traits, and support *G6PC2* as the causal gene at the established FG GWA locus.

1098W

Exome Chip genotyping in 9,000 individuals (type 2 diabetes and controls) in Mexican and Latinos. *J.M. Mercader[1,2], H. Moreno[3], A. Huerta[3], M.J. Gomez[3] for the SIGMA T2D Genetics Consortium.* 1) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA., United States of America; 2) Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, Massachusetts, United States of America; 3) Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran, Mexico City, Mexico.

Type 2 diabetes (T2D) has become a mounting health problem in Mexico, with an estimated nationwide 25% increment in prevalence over a 7 years period. To date, most genome-wide association studies (GWAS) to identify susceptibility genes for T2D have been carried through common variant genotyping arrays with very comprehensive coverage of European variability, but with lower coverage in other populations. Under the hypothesis that population-specific common and rare coding variants, in Mexicans and Latino-Americans may influence risk of T2D, we genotyped 4,210 T2D cases and 4,786 controls with the Exome Chip. The Exome Chip was designed to capture 235,933 protein-coding variants (mostly non-synonymous, splice-site and stop altering SNPs) that were selected based on ~12,000 sequenced exomes (including hundreds of Mexican American samples), as well as some common variants to control for population stratification and to tag already known GWAS hits. After quality control, based on several metrics and on genotype concordance with OMNI2.5M array in the same samples as well as a subset of 4,000 samples with exome-sequence, 161,676 variants in 8,621 individuals were tested for association using logistic regression adjusting for body mass index, age and ancestry via 10 principal component analysis. Among the top hits, we replicated the previously reported associations of SNPs in or near TCF7L2, and KCNQ1, with similar effect sizes and direction of effects shown in other populations. We also validated those variants identified last year in *SLC16A1* (SIGMA T2D Genetics Consortium, submitted) as well as an association in a splice-site acceptor variant in the *INS-IGF2* gene (OR=0.79; $p=7.0\times10^{-8}$) found through imputation in the same dataset. This variant, not yet replicated, has only been reported in Latin Americans, with its highest frequency in Mexico (MAF=0.17), and East Asian populations (MAF=0.01), according to 1000 Genomes. This study represents a proof of concept of how analyses of rare and common coding variants using a cost-effective methodology, such as the Exome Chip, in diverse populations can allow the identification of novel genes and functional variants involved in the susceptibility of complex diseases, such as T2D. Nonetheless, despite having systematically tested low frequency variants (MAF range of 0.1-5%) in 9,000 samples, we did not observe low frequency variants of larger effect as significantly meeting genome-wide significance for T2D risk.

1099T

A low frequency coding variant (A316T) in the glucagon-like protein receptor 1 (GLP1R) is associated with fasting glucose levels. *D. Waterworth[1], R. Scott[2], L. Li[3], C. Gillson[2], J. Aponte[4], L. Warren[3], S. Chissoe[4], M. Ehm[4], N. Wareham[2].* 1) Genetics, GlaxoSmithKline, King of Prussia, PA; 2) Department of Public Health, University of Cambridge, UK; 3) Statistical Genetics, GlaxoSmithKline, RTP, NC; 4) Genetics, GlaxoSmithKline, RTP, NC.

It has been anticipated that mining of the low frequency spectrum of variation will identify meaningful associations that were not observed with common variation (>5% MAF). With a focus on type 2 diabetes and related traits as well as obesity, we attempted to replicate and extend findings from a large sequencing study of 202 drug-target genes in 4065 subjects phenotyped for cardiovascular and metabolic traits (Nelson 2012). We selected 30 variants to be assessed in up to 46,037 subjects of European origin (cohorts including CoLaus, LOLIPOP, EPIC-Norfolk, Norfolk Diabetes Study, Ely and Fenland). All variants identified in the initial study were also imputed into a subset of these studies where genome-wide data were available (up to 10K individuals). We performed meta-analyses to test for association between these 30 variants and diabetes, insulin, glucose and obesity-related traits. A p value of <3.6e-5 was considered significant (independent tests for 202 genes for a given trait). Imputed and genotyped results of interest included SNPs within the GLP1R gene associated with fasting glucose (p=4e-4), HRH1 gene variants associated with insulin (p=2.7e-5), IL1R1 (p=2e-4) variants associated with 2-hr glucose, CNR2 variants associated with HOMAB in obese subjects (p=3e-4) and NTRKB variants associated with waist (4e-4). These results were not previously identified by genome-wide studies at genome-wide significant levels. Additional results from recently published studies were added to increase the level of evidence. The most compelling result emerging from 11,497 subjects from the CoLaus and LOLIPOP studies was an association between rs10305492 coding for A316T (MAF 0.02) in GLP1R with fasting glucose (adjusted for BMI), the target of the GLP1 agonists. This variant was included in a recent meta-analysis of 38,000 individuals (Manning 2012) , including individuals from the Ely and Fenland studies, and also showed an association with fasting glucose (p=1e-4). We have identified associations with a number of existing drug target genes for diabetes including GLP1R and ILIR1 which can assist in target validation or drug repositioning. Moreover, we provided information on novel targets such as CNR2 and identified a common variant in NTRKB which had previously been identified only as a rare cause of extreme obesity gene in children. Nelson et al, 2012, Science 337;100-4 Manning et al, 2012, Nat Genet 44:659-69.

1100F

Whole genome sequencing identifies novel low frequency variant associations in liver function traits. *L. Quaye on behalf of the UK10K Consortium Cohorts Group.* Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom.

1,754 individuals from the TwinsUK cohort were whole genome sequenced (WGS) to average 6.5x coverage using next-generation sequencing technology, as part of UK10K Consortium Cohorts Group. Variants from 1000 Genomes and those discovered through WGS conducted on 3,621 individuals within UK10K were imputed into individuals in the TwinsUK cohort who had been previously genotyped, thereby increasing the sample size for analysis. In a meta-analysis of directly sequenced and imputed variants, we evaluated variants for association with liver function traits (albumin, alkaline phosphatase, total bilirubin and gamma-glutamyl transpeptidase (GGT)) of individuals from the TwinsUK cohort. Association analysis was performed on quantile-normalised abundance of the liver function traits, with adjustments for age and body mass index. 22 previously reported genome-wide significant associations between total bilirubin and the *UGT1A* cluster on chromosome 2 were replicated (P<5E-08). In addition, an association was identified between levels of bilirubin and a variant of *KCNB1* (20q13.2) with a minor allele frequency (MAF) of 0.026 (beta= 0.61, standard error (SE)=0.11, P=6.02E=08). This variant, also associated with height, is a missense mutation within the coding region, resulting in a proline to serine amino acid substitution. This association is being replicated with exome sequencing data and an independent cohort. A further genome-wide significant association was found between GGT and a rare variant not available in HapMap2 reference panels), near *TEX14*, on chromosome 17p13 (MAF= 0.019; beta beta=-0.82842, SE=0.17, P=8.99E-7), which is not available on Hapmap2. A SKAT analysis of 26,367 gene regions using 42,799 non-overlapping windows with maximal 50 coding variants with MAF<5% showed a suggestive association (P=3.1E-5) between KCNB1 and bilirubin. In conclusion, we found novel associations have been found between bilirubin and GGT and variants with MAF<0.05. These promising early results from the UK10K project highlight the value of WGS studies in identifying novel associations and indicate that low frequency variants are likely to contribute to the genetic variance of liver function traits.

1101W

Autozygosity Mapping in Pakistani Intellectual Disability Families. *M. Rafiq[1], K. Mittal[1], I.A. Balouch[1,2], A. Noor[1,3], C. Windpassinger[4], A. Mikhailov[1], M. Aslam[5], M. Ayaz[5], A. Mir[6], M. Ansar[7], P. John[2], M. Ayub[5,8], J.B. Vincent[1].* 1) Molecular Neuropsychiatry, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad-Pakistan; 3) Dept. of Pathology & Laboratory Medicine, Hospital for Sick Children, Toronto, Canada; 4) Institute of Human Genetics, Medical University of Graz, Austria; 5) Lahore Institute of Research & Development, Lahore, Pakistan; 6) International Islamic University, Islamabad, Pakistan; 7) Dept. of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan; 8) Division of Developmental Disabilities, Dept. of Psychiatry, Queen's University, Kingston, ON, Canada.

Autosomal recessive causes of intellectual disability (ARID) have, until very recently, been under-researched due to the high degree of genetic heterogeneity. However, now that genome-wide approaches can be applied to single multiplex consanguineous families, identification of genes harboring disease-causing mutations by autozygosity mapping is expanding rapidly. We have ascertained more than 165 multiplex, consanguineous ARID families from Pakistan. These families are selected for lack of obvious syndromic features. Our strategy includes genotyping family members on genome-wide single nucleotide polymorphism microarrays, looking for large regions of shared homozygosity (and haploidentity) between affected individuals (homozygosity-by-descent, or autozygosity). We also screen for potential disease-related CNVs- either as a shared homozygous genotype, or heterozygous as a potential cause of phenocopy. We firstly exclude any known ARID genes in HBD regions, then either select candidates from within the HBD region for mutation screening by Sanger sequencing, or we embark on whole exome sequencing to identify disease mutations. Our successes include a number of new genes for apparent non-syndromic ARID, such as MAN1B1, TRAPPC9, NSUN2, as well as new genes for syndromic forms or ARID, such as Joubert syndrome (CC2D2A and TCTN2), and many known ARID genes (TUSC3, TPO, VPS13B, PEX1, PSPH, PMM2). Here we describe the use of autozygosity mapping and whole exome sequencing to identify an additional 6 new genes for NS-ARID. As more and more genes for ID are identified, using these and other strategies, we are building a picture of the biological pathways that, when perturbed, may lead to intellectual disability.

1102T

Identifying Genetic Variants Associated with Anorexia Nervosa via Exome Sequencing. *S. Yu, E. Pruett, R. Cone, B. Li.* Vanderbilt University, Nashville, TN.

Anorexia nervosa (AN) is a highly heritable psychiatric disease characterized by inability to maintain a minimal normal weight, persistent fear of gaining weight and preoccupation about body shape. It affects 0.5-1% of the population with females 10 times more likely to be affected than males. The estimated heritability of AN is 56-75%; however the genetic basis of AN is largely unknown. Results from previous candidate gene studies are often not replicated and genome wide association studies focusing on common variants (e.g. minor allele frequency >5%) do not reveal convincing association signals. We hypothesize that low frequency and rare variants with relatively higher genetic effects play an important role in the genetic etiology of AN. To identify genes and pathways harboring such high-risk rare variants, we carried out exome sequencing on ~70 AN patients and their relatives selected from multiple-generation pedigrees with high loads of AN and other comorbid psychiatric traits. We followed the best-practice procedure for variant calling and annotated all variants using ANNOVAR. We performed single variant analyses using MQLS to identify variants enriched in AN cases compared to their relatives. Although no variants reached exome-wide significance, we observed 14 genes harboring deleterious variants with p value < 10-3. We further carried out gene- and pathway-based analysis by collapsing multiple rare functional variants (i.e. nonsynonymous, splice and stop variants) predicted to be deleterious by bioinformatics algorithms. Using the KEGG knowledge base we identified pathways that are enriched for genes harboring deleterious functional variants, with the most common pathways identified being involved in neurodevelopment. We will sequence additional samples from an independent cohort to further replicate our initial findings.

1103F

Identifying the genetic architecture of neural tube defects by exome sequencing a multiplex anencephaly family. *K. Soldano, D. Krupp, H. Cope, M. Garrett, A. Ashley-Koch, S. Gregory.* Center for Human Genetics, Duke University, Durham, NC.

Neural tube defects (NTDs) are among the most common of birth defects, occurring in approximately 1 in 1000 live births in the US. Caused by improper closure of the developing neural tube during development, NTDs encompass a broad range of phenotypes that are likely to have a complex etiology with both genetic and environmental factors. The most severe NTD phenotype is anencephaly, an open NTD that results in the partial or complete loss of the brain and skull, and is therefore incompatible with life. While most cases are sporadic, some families have multiple affected offspring and are more likely to harbor a highly penetrant genetic risk factor. We performed exome sequencing on one such family who had three consecutive anencephalic pregnancies and a fourth unaffected pregnancy. We have completed the sequencing and analysis of the parents and the first two affecteds in this multiplex family. While we observed that the second affected sample was contaminated with maternal tissue (likely taken at the time of collection), we were able to correct for this using parental genotypes and allele proportions at each site. Exonic regions were enriched using the Illumina TruSeq Exome Enrichment kit, sequenced 2×100nt on an Illumina HiSeq2000 instrument, and aligned against NCBI build 37. Variants were called using the Genome Analysis Toolkit, and only sites with a minimum read depth of 3, quality score of 50, mapping quality of 45 and quality by depth of 7 were included in analysis. The initial set of 115,066 variants was further filtered by requiring concordant genotypes and the presence of at least one minor allele in both affecteds; 34,701 variants met these criteria. We identified 10 nonsynonymous variants in genes belonging to the Gene Ontology term cilium morphogenesis. All of these variants had deleterious effects as predicted by PolyPhen-2, SIFT, and/or PROVEAN, including a pair of compound heterozygous sites in the gene KIF24. These two variants had minor allele frequencies of 0.8% and 19.1% and were inherited from both the mother and father. Exome sequencing of the third affected and the unaffected sibling are currently underway to refine the putative causal genetic risk factors in this family. We are encouraged that the initial findings have implicated molecular pathways previously associated with NTD risk. Downstream functional analyses will be necessary to fully delineate the role of these polymorphisms in the development of anencephaly.

1104W

Whole exome sequencing case-control using 1,000 severe obesity cases identifies putative new loci and replicates previously established loci. *A. Hendricks on behalf of the UK10K Consortium: Obesity.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The UK10K project (www.uk10k.org) has sequenced 10,000 individuals: 4,000 cohort participants using low depth (6x) whole-genome sequencing (WGS) and 6,000 disease cases using high depth (50x) whole-exome sequencing (WES) with the aim of providing a variant resource to the community and identifying new disease loci. Of the 6,000 WES cases, 1,000 are patients from the Severe Childhood Onset Obesity Project (SCOOP). SCOOP consists of UK white patients, with an age of onset below 10, and with a body mass index (BMI) Standard Deviation Score (SDS) > 3, a design that is enriched for rare, highly penetrant causes of obesity. Single-variant and gene-region based case-control analyses, using two different control sets, has confirmed previously known obesity signals as well as identified potentially new variants and gene-regions associated with severe childhood obesity. The single-variant case-control analyses were performed using the low depth WGS UK10K population cohort, and a subset of the other WES UK10K disease cases that had been consented for use as controls. After thorough pre- and post- analysis quality control, the results do not show any apparent systematic bias and produce 27 promising single variant associations (p-value < 1e-5), which we are following-up in additional samples. Amongst these is a single variant association within the ADCY3 region (p=3.6e-06), a gene with an established association with BMI. Finally, we performed gene-region based case-control analysis using SKAT-O using the subset of other WES UK10K disease cases as controls. We find several potentially novel gene-regions associated with obesity and are currently performing targeted resequencing of nearly 3,000 independent SCOOP samples to replicate and add further evidence to these findings. The case-control analyses and replication efforts described here highlight the potential value of case-control using WES data to identify disease-associated loci and will help gain further insight into severe childhood obesity.

1105T

Rare variant association analysis reveals novel associations with lipids in genes within established loci via imputation up to the 1000 Genomes Project reference panel.
R. Magi[1], M. Horikoshi[2,3], I. Surakka[4,5], S. Wiltshire[2,3], A.-P. Sarin[4,5], T. Esko[1], A. Mahajan[2], T. Ferreira[2], M. Beekman[6,7], S. Gustafsson[8], S. Hägg[8], C.I. Ladenval[9], L. Marullo[2,10], C.P. Nelson[11,12], J.S. Ried[13], G. Thorleifsson[14], N. Tsernikova[1], S.M. Willems[15], C. Willenborg[16], T. Winkler[17], C.M. Lindgren[2,18], M.I. McCarthy[2,3], S. Ripatti[4,5], I. Prokopenko[2,19], A.P. Morris[2], ENGAGE consortium. 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 4) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland; 5) National Institute for Health and Welfare, Helsinki, Finland; 6) Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, Netherlands; 7) Netherlands Consortium for Healthy Ageing, Leiden, Netherlands; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 9) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University and Lund University Diabetes Centre, CRC at Skåne University Hospital, Malmö, Sweden; 10) Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy; 11) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 12) National Institute for Health Research (NIHR) Leicester Cardiovascular Disease Biomedical Research Unit, Glenfield Hospital, Leicester, UK; 13) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 14) deCODE Genetics, Reykjavik, Iceland; 15) Department of Genetic Epidemiology, Erasmus MC, Rotterdam, The Netherlands; 16) AG Kardiovaskuläre Genomik, Medizinische Klinik II, Universität zu Lübeck, Lübeck, Germany; 17) Public Health and Gender Studies, Institute of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany; 18) Broad Institute, Boston, MA, USA; 19) Genomics of Common Disease, Imperial College London, London, UK.

Genome-wide association studies (GWAS) have been extremely successful in identifying novel loci contributing genetic effects to plasma concentrations of lipids. These loci are typically characterised by common variant associations, and together account for no more than ~25-30% of the genetic variance of total cholesterol (TC), low- and high-density lipoprotein cholesterol (LDL-C and HDL-C), and triglycerides (TG). The aim of this study was to assess the evidence for association of these traits with rare genetic variation (MAF<1%) within genes through imputation into GWAS genotype data from the ENGAGE Consortium.

We considered 31,853 individuals from 17 GWAS of lipid traits, each imputed up to the Phase I 'all ancestries' 1000 Genomes Project reference panel (November 2011 release). In each study, we tested for association of all traits with accumulations of minor alleles at high-quality rare variants (MAF<1% and imputation quality >0.4) within genes (boundaries defined from the UCSC Human Genome database) using GRANVIL (http://www.well.ox.ac.uk/GRANVIL/index.shtml). Association summary statistics were then combined across studies via fixed-effects Z-score weighted meta-analysis.

We observed genome-wide significant evidence of association (Bonferroni correction for 30,000 genes, $p<1.7\times10^{-6}$) for two lipid traits. Accumulations of minor alleles at rare variants were associated with TG within two genes at the *APOA5* locus: *ZNF259* ($p=1.5 \times 10^{-11}$) and *APOA5* ($p=5.0 \times 10^{-8}$). Common variation at this locus has been associated with lipid traits, and rare mutations in *APOA5* have been implicated in hypertriglyceridemia. By restricting our analysis to non-synonymous changes, accumulations of minor alleles at rare variants were associated with HDL-C within *LIPC* ($p=2.1\times10^{-7}$). This gene encodes hepatic TG lipase, an essential enzyme in HDL-C metabolism, and common variation at this locus has been associated with lipid traits.

We have identified novel rare variant associations with TG and HDL-C within genes in established lipid loci. Our results highlight the potential of imputation into existing GWAS data up to high-density reference panels as a tool for discovering rare variant associations with complex traits, and an alternative to costly re-sequencing experiments.

1106F

Identification of common and rare variants associated with trunk fat mass using whole-genome sequencing in the UK10K project.
L. Paternoster, UK10K Consortium Cohorts Group. MRC Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom.

As part of UK10K Consortium, 1867 individuals from the Avon Longitudinal Study of Parents & Children (ALSPAC) were whole genome sequenced (WGS) to ~6.5x coverage. Variants discovered through WGS, along with those from 1000 Genomes were imputed into the full cohort with genome-wide SNP data, increasing the sample size to 5498. Trunk fat mass (a highly accurate measure of central obesity) was derived at age 9yrs using dual energy x-ray absorptiometry (DXA). To investigate common variation, we tested 8.8 million variants (MAF>1%) in a meta-analysis of sequenced and imputed variants for association with trunk fat mass. We found three loci with $p<5\times10^{-8}$. Two of these are in known BMI regions (*ADCY3* & *FTO*). The other locus (7p21.1) has previously shown suggestive evidence for association with BMI z-score change in Hispanic children ($p=7\times10^{-6}$). We also found a modest association signal between this variant and BMI in our data ($p=7\times10^{-4}$), but no association with waist hip ratio (WHR). The GIANT GWAS of BMI showed no association with this variant (p=0.847) which indicates that this signal might be specific to childhood obesity. 25 loci were suggestively associated ($p<1\times10^{-5}$); 17 are novel for weight-related traits and 9 are variants of low frequency (MAF<5%). We are replicating these in the TwinsUK arm of the UK10K project (n=1754 sequenced) and additional child cohorts. For rare variants (in WGS data only) we used SKAT to analyse 26,367 gene regions using 41,692 non-overlapping windows with maximal 50 coding variants with MAF<5%. We found three regions (*IL17RC*, *PPP4R1* & *RP11-1152B5*) associated with trunk fat mass ($p<1\times10^{-5}$). Further investigation of the signals in these regions show that two of these (*IL17RC* & *RP11-1152B5*) are driven by only one variant, whereas the signal for *PPP4R1* appears to be driven by two variants. Two of these genes (*IL17RC* & *PPP4R1*) also showed suggestive evidence for association with BMI (p<0.01), but no association with WHR. We are following these up by genotyping in the whole of the ALSPAC cohort to allow validation and replication. We will also investigate enrichment of hits according to functional annotation. The UK10K large-scale WGS project has identified novel loci for trunk fat mass. Our findings suggest that both rare and common variants contribute to the variance of this trait.

1107W

Statistical dissection of genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1) using whole-genome sequence data. *R. Rubicz[1], M. Almeida[1], E. Drigalenko[1], T.D. Dyer[1], T.M. Teslovich[2], G. Jun[2], J.M. Peralta[1], C. Fuchsberger[2], A.R. Wood[3], A.R. Manning[4], T.M. Frayling[3], P. Cingolani[5], R. Sladek[6], D.M. Lehman[7], J.W. Kent Jr.[1], J.B. Harley[8], M.A. Carless[1], J.E. Curran[1], M.P. Johnson[1], S.A. Cole[1], L. Almasy[1], E. Kraig[7], G. Abecasis[2], R. Yolken[9], R. Duggirala[1], C.T. Leach[7], J. Blangero[1], H.H.H. Göring[1], T2D-GENES Consortium.* 1) Dept of Genetics, Texas Biomed, San Antonio, TX; 2) University of Michigan, Ann Arbor, MI; 3) University of Exeter, Exeter, United Kingdom; 4) Broad Institute, Boston, MA; 5) McGill University, Montreal, Canada; 6) Montreal Diabetes Research Institute, Montreal, Canada; 7) University of Texas Health Science Center at San Antonio, San Antonio, TX; 8) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 9) Johns Hopkins University School of Medicine, Baltimore, MD.

Infection with Epstein-Barr virus (EBV) is highly prevalent among populations around the world and can cause infectious mononucleosis and more severe diseases including Hodgkin lymphoma, nasopharyngeal lymphoma, and post-transplant lymphoproliferative disorders. We measured IgG antibodies against EBV nuclear antigen 1 (EBNA-1) in extended Mexican American families from San Antonio, Texas. We previously performed association analysis in 1,956 individuals using 1M SNPs and reported significant association in the human leukocyte antigen (HLA) region on chromosome 6. There were at least two independent loci in this region (lowest p-value of 1.4×10^{-15}), with HLA class II genes *HLA-DRB1* and *HLA-DQB1* identified as the best candidates. This region was not associated with antibody titer levels against 13 other pathogens, indicating that the loci are specific to EBV. Whole genome sequences (WGS) were recently generated on a subset of the study participants as part of the T2D Consortium. There were a total of >20M variants with more than 5 minor allele copies. Here we used the available cleaned WGS, including highly accurate imputed data for some relatives, in order to search for additional genomic loci associated with EBNA-1 level and to dissect the HLA region in detail. We did not identify novel loci elsewhere in the genome, aside from suggestive evidence of association for SNP rs16921084 on chromosome 8 ($p = 9.39\times10^{-8}$) near gene *IMPAD1*. Within the HLA region, the most significant result was obtained for SNP rs204999 ($p = 6.97\times10^{-9}$ and $\beta = -0.39$, A→G transition with minor allele frequency of 19.1%), which had previously been included in the 1M SNP panel. This SNP was previously associated with decreased risk of nodular sclerosing Hodgkin lymphoma (NSHL) and with schizophrenia. SNP rs204999 is intergenic between HLA class III genes *FKBPL*, which plays a role in intracellular signalling and has been identified as a potential inhibitor of tumor growth, and *PRRT1*. Other significant HLA SNPs, after correcting for >70k variants in the extended HLA region, included rs9276472 ($p = 4.05\times10^{-7}$; gene *MIR3135B*) and rs204995 ($p = 6.39\times10^{-7}$; gene *PBX2*), also in the HLA III region.

1108T

Whole-exome sequencing in age-related macular degeneration. *P. Whitehead[1], W.K. Scott[1], G. Wang[1], W. Cade[1], M.D. Courtenay[1], S.G. Schwartz[3], J.L. Kovach[3], A. Agarwal[4], J.L. Haines[2,4], M.A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 3) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 4) Ophthalmology and Visual Sciences, Vanderbilt University School of Medicine, Nashville, TN.

Genome-wide association meta analysis (GWAMA) has implicated common variations in 19 genes as AMD risk factors. Rare variants (RV) are not well captured by genome wide association studies and may represent some of the unexplained heritability in AMD risk. We used whole exome sequencing (WES) in phenotypically extreme individuals to identify RVs implicating novel AMD genes. The GWAS 19 loci were used to calculate a genetic risk score (summed number of risk alleles weighted by effect sizes) for each individual. We created an 'extreme' case/control sample with the following characteristics: 38 individuals with bilateral neovascular AMD and the lowest calculated genetic risk score, and youngest ages at examination; 37 unaffected controls with no drusen, the highest calculated genetic risk score, and oldest ages at examination. Sequencing Capture was by Agilent SureSelect. Alignment and base calling used the Illumina CASAVA 1.6 pipeline, aligned to hg19 using BWA. Single nucleotide variants (SNV) and insertion-deletion variants (indels) were called by GATK Unified Genotyper with VQSR recalibration. Variants with VQSLOD<-3 and variant genotype likelihoods <99 were excluded. All variants were annotated using SeattleSeq Annotation. Association of individual SNV with AMD was assessed by Fisher's exact test. RVASSOC was used for the gene-based tests considering all variants and just rare (minor allele frequency < 5%), damaging coding variants. The 19 genes used to define genotypic extremes were excluded from the gene-based analysis. No gene-based tests met Bonferroni-corrected significance (2.2×10^{-6}). When considering all variants, 19 genes generated gene based test $p<5\times10^{-4}$. When considering only rare damaging variants, six genes (*KRT26*, *KRT27*, *MOG*, *PLD5*, *CD276*, and *SYT9*) generated gene based test $p<5\times10^{-4}$. *MOG* is close to known AMD gene *IER3*, and the others represent novel loci (*KRT26* and *KRT27* are adjacent genes). No novel variants in the 19 known AMD loci were associated with the extreme phenotype. Initial results of this WES suggest additional genes with multiple rare SNVs may influence AMD.

1109F

Whole Genome Sequencing Association Study for Quantitative Ultrasound of the Calcaneus. *S.G. Wilson* [1,2,3], . on behalf of the UK10K cohorts[4].* 1) Endocrinology & Diabetes, Sir Charles Gairdner Hosp, Nedlands, Australia; 2) Department of Twin Research, King's College London, London, UK; 3) Medicine & Pharmacology, University of Western Australia, Australia; 4) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK.

Aim: To identify genetic variants associated with Quantitative Ultrasound (QUS) of the Calcaneus which is used for evaluation of osteoporosis, we performed an association study using whole genome sequence (WGS) data from the UK10K Consortium.

Methods: 1327 individuals from TwinsUK cohort with QUS phenotypes (broadband ultrasound attenuation (BUA) and velocity of sound (VOS)) and WGS were studied. The BUA and VOS data were adjusted for age, age^2, height and weight. We undertook single variant tests using SNPTEST. Variants with a MAF <0.005 or not in Hardy-Weinberg equilibrium were excluded. Significant and suggestive thresholds were set at 5E-8 and 5E-6 respectively.

Results: In the single variant analysis we identified 2 variants significantly associated with BUA at genome-wide significance and a further 76 showing suggestive association. No variants were associated with VOS at genome-wide significant level, however we observed 94 variants showing suggestive association with the phenotype. For VOS, 9 of the variants showing suggestive association were located in the FSHR gene region which has previously been associated with bone mineral density (BMD) and QUS parameters in postmenopausal women. For BUA, both variants associated at genome-wide significance and 28 of those showing suggestive association were located in the WNT16 gene region which has previously been associated with BMD, cortical bone thickness and osteoporotic fracture. Of the remaining loci identified in the study as provisionally associated, most were located in novel genomic regions not previously implicated in bone metabolism.

Conclusion: In this study of whole genome sequence data for association with QUS traits, 2 variants were identified as associated at genome-wide significance with a further 170 showing suggestive association. We were able to confirm the association of variants in the WNT16 gene with bone structural phenotypes and generated supporting evidence suggesting a role for the FSHR gene region in bone metabolism. We also highlighted many other novel loci and genes for further study. These variants will undergo additional replication through in silico analyses and de-novo genotyping.

*Benjamin H. Mullin contributed equally to this work.

1110W

Targeted Sequencing, Augmented with Public Resources, Identifies a Rare C3 Allele Associated with Large Risk of Agerelated Macular Degeneration. *X. Zhan[1,2], D. Larson[3,4], R. Fulton[3,4], C. Wang[5], D. Stambolian[6], E. Chew[7], E. Mardis[3,4], A. Swaroop[7], G. Abecasis[1,2].* 1) Biostatistics, University of Michigan, Ann Arbor, MI, United States; 2) Center of Statistical Genetics, University of Michigan, Ann Arbor, MI, United States; 3) The Genome Institute, Washington University, St. Louis, MO, United States; 4) Department of Genetics, Washington University, St. Louis, MO, United States; 5) Department of Biostatistics, Harvard University, Boston, MA, United States; 6) Department of Ophthalmology and Human Genetics, University of Pennsylvania, Philadelphia, PA, United States; 7) National Eye Institute, Bethesda, MD, United States.

Macular degeneration is one of the most common causes of incurable blindness. Common alleles in >19 loci have now been associated with disease. We set out to investigate whether rare variants in the same loci were also associated with disease risk and to compare the relative effect sizes of common and rare variants.

In collaboration with the Genome Institute at Washington University in St. Louis, we designed a sequencing study focused on 8 of the known AMD risk loci (CFH, ARMS2, C3, C2/CFB, CFI, CETP, LIPC and TIMP3/SYN3) and 2 other candidate regions (LPL and ABCA1). We resequenced these regions in 3,124 individuals (2,335 cases and 789 controls) to an average depth of 85x. To augment the number of controls available for analysis, we designed an algorithm to identify previously sequenced samples with good coverage of our regions of interest and similar genetic ancestry. Finally, we investigated association between genetic variation in each locus and risk of disease using both single variant and gene level burden tests.

Across 967 kb of examined sequence, we discovered 41,202 high quality SNP variants in the 3,124 sequenced individuals (34,346 of these were novel, and not previously described in dbSNP). Among these variants, we focused our attention on 1,800 nonsynonymous, stop and splice variants. We estimated the ancestry of our sequenced samples and of samples from the NHLBI exome sequencing project, identifying an ancestry matched set of 2,268 cases and 2,268 controls. Individuals in this matched set had deep (minimum 10x) coverage of coding regions in the 10 targeted loci. Association analysis identified two strongly associated variants, one in the CFH gene (control frequency = 0.02%, exact Pvalue = 2.91×10^{-6}, OR = 23.11) and another in the C3 gene (control frequency = 0.40%, exact Pvalue = 2.73×10^{-4}, OR = 2.68). Replication efforts for these findings are ongoing.

Through targeted sequencing efforts, augmented with publicly available control data, we replicated a previously reported rare variant association in the CFH gene and identified a new rare variant signal in the C3 gene. In both instances, these rare variants are associated with substantially larger odds ratios than common variants in the same regions.

1111T

DES is uniquely useful in the identification of multi-gene alterations: Oligogenic findings make up a significant portion of previously undiagnosed patients. *K. Gonzalez, L. Shahmirzadi, E. Chao, S. Gandomi, M. Parra, B. Tippin Davis, W. Zeng, S. Tang.* Ambry Genetics, Aliso Viejo, CA.

Diagnostic exome sequencing (DES) has been instrumental in discovering the pathogenic etiology in patients in whom traditional molecular methods were uninformative. In addition to establishing a molecular diagnosis, enhancing genetic counseling, and aiding in clinical management, DES is uniquely useful in overcoming limitations posed by traditional molecular diagnostic strategies in the identification of multi-gene findings. Often complicating the interpretation of genetic variants are factors such as reduced penetrance and variable expressivity, frequently attributed to the effects of genetic modifiers. Herein, we examined the first 100 reported DES cases with positive gene alteration findings. Among them, 17 (17%) contained mutations in more than one reported gene of either positive or uncertain significance. The majority of the multi-gene cases (15) reported two gene findings, while two cases each reported three gene findings. All reported genes were associated with high penetrance disease as DES does not include the interpretation or reporting of disease-association risk alleles with low penetrance. The clinical spectrum of the patients with multi-gene findings was wide with 23% presenting with multiple congenital anomalies, 19% metabolic disease, 19% musculoskeletal phenotypes, and 15% childhood onset neurological disease. A minority of patients also presented with hematologic, cancer predisposition, cardiologic, and immunologic/infectious phenotypes. Interestingly, oligogenic findings were not identified among the cohort of positive results from patients presenting with autism spectrum disorder, which is typically considered multi-factorial. Virtually all patients had not previously received a clinical diagnosis consistent with our DES results and less than half had been provided with a consistent differential diagnosis. Among the 17 multi-gene cases, 11 (65%) contained at least one alteration classified as a deleterious mutation. If single gene testing had been performed for these patients, the pursuit of further molecular testing in the context of a single deleterious alteration in a well described gene would have been unlikely. Overall, these results reveal that oligogenic findings make up a significant portion of previously undiagnosed patients highlighting the value of DES in providing the most comprehensive molecular diagnosis available. Moreover, these data have significant implications for genetic counseling and clinical management.

1112F

Meta-analysis of rare variant associations for lipids metabolism traits. *D. Liu on behalf of the Global Lipids Genetics Exome Array Consortium.* Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Plasma lipids including total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglycerides are biomarkers that predict the risk of coronary artery disease. Although many genetic loci have been identified for blood lipids using the genome-wide association approach, these variants collectively explain a modest proportion of overall heritability for these traits. In order to investigate the contribution of rare genetic variants to blood lipid levels, we assembled one of the largest exome-wide dataset to date. We examined array genotype data for ~201,000 protein coding variants in ~85,000 individuals from 34 studies with lipids measurements. Most variants were rare, with ~180,000 (83.5%) variants having a frequency of <1%. For analysis, we developed and implemented a method that uses summary statistics measuring association at each variant and their covariance matrix to reconstruct single variant and gene-level association statistics across the entire sample. At the time of submitting the abstract, we completed first round of association analysis. Overall, 154 coding variants (139 common with MAF>1%, 15 rare with MAF < 1%) were associated with one or more lipid traits at $p < 3\times10^{-7}$. In addition, 25 genes showed significant association with one or more lipid traits at $p < 2.5\times10^{-6}$ using at least one of the three gene level association tests we considered (a simple burden test for variants with frequency < 5%, a variable threshold association test, and sequence kernel association test for variants with frequency <5%).Loci with evidence for association include known and novel loci, such as *ABCA1*, *APOC3*, *APOE*, *LDLR*, *PCSK9*, *CD300LG*, *CELSR2*, *APOH*, and *NPC1L1*. Using this dataset, we are investigating the genetic architecture of lipids traits, to estimate the number and characteristics of rare variants contributing to plasma lipid levels, and quantify the impact of performing deep sequencing and enlarging sample sizes on the power for detecting rare variant associations.

1113W

Exome-wide association study of fetal hemoglobin levels in African Americans with sickle cell disease. *S. Lessard[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montréal, Québec, Canada; 2) Université de Montréal, Montréal, Québec, Canada.

INTRODUCTION: Reactivation of fetal hemoglobin (HbF) production in patients suffering from sickle cell disease (SCD) is a promising treatment. To develop such therapy, it is however necessary to understand the genetic mechanisms underlying the regulation of HbF levels. Genome-wide association studies have identified common variants associated with HbF at three loci (BCL11A, HBS1L-MYB and HBB), which altogether explain half of the heritability of the trait. Rare and low-frequency markers are expected to account for a substantial fraction of the unexplained heritability and are not tested using GWAS methodologies. METHODS & RESULTS: To address this issue, 250,000 DNA sequence variants, mostly coding and rare or of low-frequency, were genotyped in 1,495 African Americans with SCD from the Cooperative Study of Sickle Cell Disease (CSSCD) using the Illumina ExomeChip array. After quality control, 132,419 polymorphic markers and 1,301 individuals remained, with genotyping rate >99.9%. Variants with a minor allele frequency (MAF) ≥0.05 were analysed by linear regression against HbF using the software PLINK, accounting for age, sex and the first 10 principal components. The analysis replicated the association of HbF at the BCL11A, HBS1L-MYB and HBB loci. Excluding these loci, the lowest P-values were found at two SNP near the KLF1 gene ($P=9.08\times10^{-6}$ and 1.75×10^{-5}), an erythroid transcription factor implicated in the globin switch and hereditry persistence of fetal hemoglobin. We analyzed missense, nonsense and splice site markers with MAF <0.05 using the Optimal Sequence Kernel Association Test (SKAT-O), but no genes reached exome-wide significance (all $P>2.5\times10^{-6}$). To test if any biological pathways were enriched in the association results, we performed gene-set enrichment analyses using gene-based association results. Preliminary analysis revealed multiple pathways nominally enriched, implicating among others Notch, FGF and SAPK/JNK signalling. CONCLUSION: We genotyped SCD patients with the Exome Chip to identify functional DNA sequence variants associated with HbF levels. The data generated here will be combined with other genetic results along with comprehensive functional data in order to identify new regulators of HbF production and biomarkers of SCD morbidity.

1114T

Targeted sequencing identifies loci associated with sarcoidosis in African Americans. *I. Adrianto[1], G.B. Wiley[1], A.M. Levin[2], S.B. Glenn[1], M.C. Iannuzzi[3], B.A. Rybicki[2], P.M. Gaffney[1], C.G. Montgomery[1].* 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Public Health Sciences, Henry Ford Health System, Detroit, MI; 3) Department of Medicine, SUNY Upstate Medical University, Syracuse, NY.

Sarcoidosis is a systemic granulomatous disease that primarily affects the lungs, but can manifest in multiple organ systems. While previous genome-wide association studies (GWAS) of both European and African American populations have identified common variants associated with sarcoidosis, rare variants (minor allele frequency<1%-5%) are hypothesized to have large effect sizes and may contribute to sarcoidosis risk. Therefore, we analyzed both common and rare variants to identify loci associated with sarcoidosis. We performed targeted sequencing using the Illumina HiSeq 2000 platform of 5Mb of genomic DNA within the loci identified through linkage, admixture, and GWAS in 187 sarcoidosis cases and 293 healthy controls of African Americans. All samples were sequenced with an average of 75x coverage. Rare variant association tests, which collapse rare variants within a gene of interest, were utilized using PLINK/SEQ. The sequence kernel association test (SKAT) was also applied to analyze both common and rare variants. An advantage of SKAT is its robustness to identify variants with different direction and magnitude of effects. Based on rare variant association tests, we identified 15 significant genes including PPT2 ($p=4.22\times10^{-4}$), FRMD3 ($p=4.46\times10^{-4}$), and ZNF451 ($p=5.66\times10^{-4}$). PPT2 has been associated with pulmonary function. FRMD3 is under-expressed in the lung tumor tissue of lung carcinoma patients suggesting its important role in the origin and progression of lung cancer. ZNF451 may play a role in transcriptional regulation. This gene is near RAB23, a sarcoidosis locus first identified in a German sample. From the SKAT test by taking into account both common and rare variants, we found 5 significant genes including HLA-DQA1 ($p=6.17\times10^{-4}$), AGER ($p=6.30\times10^{-3}$), and NOD2 ($p=4.33\times10^{-3}$). HLA-DQA1 plays a central role in the immune system and has been associated with sarcoidosis and other autoimmune disorders including rheumatoid arthritis, type 1 diabetes, and celiac disease. AGER has been associated with pulmonary function, whereas NOD2 that plays a role in the immune response, has been associated with early-onset sarcoidosis, inflammatory bowel disease, and psoriatic arthritis. This is the first targeted sequencing study to be done in sarcoidosis. We present the association of rare variants within genes in sarcoidosis. Replication and functional studies were required to define those associated genes in the specific pathogenesis of sarcoidosis.

1115F

Exome sequencing identifies variants in families with idiopathic scoliosis. *E. Baschal[1], K. Swindle[1], K. Calbart[1], C. Wethey[1], K. Gowan[2], K. Jones[2], N. Miller[1].* 1) Orthopaedics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO; 2) Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, Aurora, CO.

Idiopathic scoliosis (IS) is the most common disorder of spinal imbalance, which, when progressive, leads to significant morbidity and financial burden to individuals and families. Most cases are strongly suspected to be genetic, arguing for the development of novel approaches to finding causative genes for this disorder. As efforts to find the genetic causations of IS, including GWAS, have not led to major advances in disease understanding, we believe that causal variants may differ among families. To this end, we used exome sequencing to focus on multigenerational families affected with IS.

We completed exome sequencing for 7 multigenerational families with IS, sequencing three affected individuals in each family (average 6 affected per family). Illumina HiSeq reads were mapped to the reference human genome sequence (hg19) with large-scale alignment software (GSNAP). Sequence calls for variants (single-nucleotide polymorphisms, insertions and deletions) were performed using the Broad's Genome Analysis Toolkit (GATK). The program ANNOVAR was used to filter the variants by cross-referencing multiple genetic variation databases (e.g., dbSNP, 1000 genomes database, AVSIFT, etc.). Information was extracted about variant frequency (if previously known) and location within genes. Non-synonymous changes, those that cause an alternate splice site, and/or aberrant stop codon were considered for further analyses. For non-synonymous changes, variants were cross-referenced to the dbNSFP database to determine whether changes to the protein were considered tolerable or damaging. Finally, a segregation filter was applied. Filtering criteria required variants to be present in all three affected individuals in the family. This bioinformatic filtering resulted in the identification of 3 to 26 potential causative variants in each family (minor allele frequency less than 5%). No variants were present in all families. We performed confirmatory Sanger sequencing for select variants in individuals who originally underwent exome sequencing and in additional affected and unaffected family members. These findings suggest that genes with identified variants may contribute to the scoliosis phenotype, a significant step towards determining the genetic etiology of this disorder. Further investigation is needed to validate novel genes for IS by sequencing the genes in an additional familial IS cohort.

1116W

Asthma: An Omics View through Discordant Monozygotic Twins. *R. Chen[1], G.I. Mias[1], S. Runyon[2], J. Li-Pook-Than[1], G. Euskirchen[1], P. Lacroute[1], K.C. Nadeau[2], M. Snyder[1].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Pediatrics, Stanford University, Stanford, CA.

Asthma is a complex disease that has a clear hereditary component. As a common disease, asthma affects 11% of the population in the United States, or 13% in children under age 18. To determine the contribution of genetic and epigenetic factors for this disease, we performed whole genome, whole transcriptome, and targeted methylome profiling in 20 pairs of monozygotic twins, who are discordant in asthma, using high throughput sequencing. We determined the whole genomes of 12 pairs of monozygotic twins discordant in asthma (as well as 1 pair discordant in allergy and 2 concordant healthy pairs), and sequenced the whole transcriptome of peripheral blood mononuclear cells (PBMCs) from 20 (including the 12 pairs for whole genome sequencing) discordant monozygotic twins pairs. In addition, we also examined DNA methylation differences within each discordant twin pair by targeted methylome enrichment and high throughput sequencing. We observed a spectrum of genomic differences within each pair of monozygotic twin, including single nucleotide variants, insertions/deletions, copy number variants, structural variants, as well as different mobile element insertions, although none seemed to be a significant contributor to the asthma phenotype. Whole transcriptome sequencing revealed surprisingly similar expression patterns in PBMCs within each twin pair, indicating strong genetic determination for gene expression in monozygotic twins. Differentially expressed genes between the discordant twins are enriched in pathways involved in immune response, which is a key feature for symptoms of asthma. In addition, we also observed interesting differential DNA methylation patterns in each monozygotic twin pair using bisulfite sequencing. Our study provides a unique perspective to evaluate genetic and epigenetic factors associated with this complex, common disease of asthma at the omics level.

1117T

Whole genome sequencing identifies genes and non-coding regions as modifiers of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *J.X. Chong[1], M.J. Emond[2], T. Louie[2], R.L. Gibson[1], M.J. Bamshad[1,3], NHLBI GO Exome Sequencing Project.* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

We recently used exome sequencing of phenotype extremes to identify variants in *DCTN4*, *CAV2*, and *TMC6* that are associated with risk for chronic Pseudomonas aeruginosa (Pa) lung infection in individuals with cystic fibrosis (CF). To investigate the contributions of coding variants not captured by existing exome targets and non-coding variation on age of onset of chronic Pa infection, we performed whole genome sequencing on individuals with CF and early-onset (n=44) or late-onset (n=45) chronic Pa. We used SKAT-O with small sample adjustment to test for differences between extremes in the distributions of variants in genes and non-coding elements, down-weighting common variants. In the per-gene analysis, we analyzed non-synonymous, splice, and utr variants. The most significant (naïve p= 6.2×10^{-6}) association was observed with *CD22*. In the non-coding analysis, we used genome segmentation annotations from the ENCODE project to predict regulatory elements such as enhancers and insulators. We identified 44,749 predicted enhancers that were variable in at least five subjects. While current computational methods cannot reliably predict enhancer targets, the top two associations were a region upstream of *ADAM19* and an intronic region in *STAP1*. These genes play roles in host immunity and are thus biologically plausible candidates. Additionally, we used GREAT (Genomic Regions Enrichment of Annotations Tool) to assign enhancers to potential cis-regulatory target genes and identified a 10.7-fold enrichment (FDR q= 6.1×10^{-4}) for inositol phosphate kinase activity among the top 5% of enhancer associations. This signal is attributable to eight predicted enhancers in *ITPK1* and three in/near *ITPKB* (a kinase upstream of *ITPK1* in the inositol phosphate signaling pathway), yet neither gene was significant nor highly ranked by p-value in the per-gene analysis: *ITPK1* (naïve p=0.35) was ranked 6,158th and *ITPKB* (naïve p=0.15) 3,041st out of 15,800 genes. Interestingly, variability in *ITPK1* expression has been previously predicted to influence CF severity and to affect efficacy of CF drugs through regulation of alternative chloride channel activity, suggesting that variants in *ITPK1* enhancers could lead to differences in age of onset of chronic Pa infection. Nevertheless, validation in a replication cohort is a necessary next step. This analysis represents an early attempt to associate non-coding variation discovered by WGS using an extreme phenotype study design.

1118F

Exome chip analysis identifies rare variants associated with primary open angle glaucoma. *M.D. Courtenay[1], R.K. Lee[2], D.L. Budenz[3], J.L. Haines[4], M.A. Pericak-Vance[1], W.K. Scott[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL, USA; 3) Department of Ophthalmology, University of North Carolina, Chapel Hill, NC, USA; 4) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA.

Primary open angle glaucoma (POAG) is a complex ocular disorder with heterogeneous etiology comprised of genetic and environmental factors. Family-based linkage studies have identified few genes with rare mutations underlying POAG, and recently the GLAUGEN and NEIGHBOR consortium case-control meta-analysis implicated common variants in the *CDKN2BAS* and *SIX1/SIX6* genes. This study examined the potential effect of rare variants on the risk of developing POAG. Illumina HumanExome-12v1 arrays were used to genotype 112,385 SNPs in 273 Caucasian POAG cases (mean age 73, 75% normotensive POAG) and 795 controls (mean age 74.5). This array includes coding region variants across the genome, with a majority of variants (88%) being rare (minor allele frequency <5%). POAG cases were clinically evaluated and classified using the NEIGHBOR consortium criteria. After quality control, SKAT was used to assess the association between 16,422 genes and case-control status. SKAT's logistic regression analysis was adjusted for population structure using 7 principal components obtained from analysis with Eigenstrat, and SNPs in each gene were inversely weighted by allele frequency. No genes met Bonferroni criteria (P=3 × 10-6) for significant association, but thirteen genes had P values<0.001. None of the top genes have been previously associated with POAG or ocular development. The most significant gene overall (P=3 × 10-4) and in the normotensive POAG only subset (P=1.5 × 10-4) was *RNF216*, a ring finger protein that colocalizes and interacts with *RIP* to inhibit TNF-induced NF-KB activation. Interestingly, glaucoma has been associated with increased NF-KB activation and optineurin, another negative regulator of TNF-induced NF-KB, has also been previously associated with glaucoma. The next most significant overall gene was *JAG2* (P=5 × 10-4). Two rare missense mutations within this gene were more common in POAG cases than controls and contributed to the gene based result. *JAG2* is widely expressed and involved in notch signaling, which is important for embryonic development. *JAG2* was also recently implicated in the growth and dissemination of uveal melanoma. These initial results suggest that additional genes with multiple coding variants (both common and rare) contribute to the complex etiology of POAG.

1119W

Identification of Variants that Confer Susceptibility to Alzheimer Disease in the Amish through Exome Sequencing. *L. D'Aoust[1], A.C. Cummings[1], R. Laux[1], D. Fuzzell[1], L. Caywood[2], L. Reinhart-Mercer[2], W.K. Scott[2], M.A. Pericak-Vance[2], J.L. Haines[1].* 1) Center for Human Genetic Research, Vanderbilt University, Nashville, TN; 2) Hussman Institute of Human Genomics, Miller School of Medicine, University of Miami, Miami, FL.

Recent studies using genome-wide association study data sets have successfully identified multiple loci harboring common variation conferring susceptibility to late-onset Alzheimer Disease (LOAD). However, these loci individually have modest effect sizes and the majority of the genetic etiology of LOAD remains unknown. Rare variants with larger effects may explain additional unknown genetic risk. The genetically isolated populations of the Amish in Ohio and Indiana are advantageous for detecting rare variants because they will be enriched for any rare variants carried by the founders. We hypothesized that rare variants conferring susceptibility to LOAD in the Amish can be identified by exome sequencing. We performed whole exome sequencing on 162 individuals (53 affected) selected from these Amish populations. After careful quality control, we identified 62,897 variants in or very near genes for analysis. Association analysis was carried out using MQLS, which adjusts for relatedness using kinship coefficients, for the variants identified through sequencing. We focused our analyses on two classes of genes: the first including 15 genes previously implicated in LOAD through GWAS or carrying early-onset mutations; the second including genes in our four previously identified significant genetic linkage regions. The exomes sequenced harbor 131 variants in the fifteen known AD genes. Of these, 120 have rs numbers assigned, 37 and 28 are not present in the current release of 1000 Genomes and the Exome Sequencing Project, respectively. The most significant p-value among these genes is 0.0098 for position 10,054,789 on chromosome 1 in ABCA7. Within the linkage regions, the most significant p-value was 0.00017 on chromosome 3. We also tested the remaining variants outside these regions, with the most significant p-value being 3.94×10^{-6} for position 128,525,253 on chromosome 3. Four additional variants have p-values less than 1×10^{-4}. The most interesting variants from this screen are currently being genotyped in the entire Amish dataset of more than 1,200 samples. Our previous genetic linkage studies suggested the Amish harbor previously unidentified LOAD loci, and these data continue to support that hypothesis. Continued studies of the Amish are necessary to identify the exact variations underlying their LOAD.

1120T

Whole-genome sequencing of an Italian Multiple Sclerosis multiplex family identifies a novel functional variant in GRAMD1B. *F. Martinelli Boneschi[1,2], F. Esposito[1,2], A. Osiceanu[2], A. Zauli[2], D. Cittaro[3], M. Sorosina[2], A. Calabria[4], D. Lazarevic[2], V. Maselli[2], P. Brambilla[2], G. Comi[1,2], E. Stupka[3].* 1) DEPT NEUROLOGY & INSPE, SCIENTIFIC INST SAN RAFFAELE, MILAN, MI, Italy; 2) Laboratory of genetics of neurological complex disorders, INSPE, Scientific Institute San Raffaele, Milan, Italy; 3) Genome Function Unit & Center of Translational Genomics and bioinformatics, Scientific Institute San Raffaele, Milan, Italy; 4) San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), Scientific Institute San Raffaele, Milan, Italy.

While the role of common genetic variants is clearly established from recent studies on individuals with multiple sclerosis (MS), the contribution of rare variants to the disease susceptibility in multiplex families remains unclear. The aim of this study is to identify rare genetic variants contributing to MS susceptibility in an Italian multiplex family. SNP microarray genotyping and whole-genome sequencing in 4 MS patients and 4 unaffected individuals belonging to an Italian multiplex family descending from a first cousin marriage were performed. The Merlin software was used for the linkage analyses and SNPeff software was applied to prioritize rare variants. We identified 12,938 variants with high or moderate functional impact according to SNPeff software annotation; 5,453 of them are present with at least one copy in all the patients, of which 250 are novel (not present in dbSNP 137) or rare (GMAF <0.02). Six of these 250 variants are located under one of the two LOD peaks (LOD>1.5) identified on chromosome 8p21.2 and 11q23.3. Four of them are homozygous in cases and controls, while the remaining two suggested an autosomal recessive mode of inheritance with incomplete penetrance. The first one (rs201865228) maps to the OR8G5 gene, which is an olfactory receptor gene under positive selection (GERP++ score=-3.33), while the second one falls within the GRAMD1B gene, causing an amino acid substitution (S601P) (GERP++ score=4.72). The second variant is a novel one, and segregates within the family, being homozygote in 3 affected individuals and heterozygote in the remaining MS patient. Sanger sequencing confirmed the above segregation with the disease. Interestingly enough, the heterozygote MS patient carries the highest genetic load of MS risk variants. GRAMD1B is a very conserved gene from yeast to human. It encodes a membrane protein, which is part of the GRAM containing domain family protein. In the mouse it is highly expressed in the central nervous system and in specific immune cell subtypes, like dendritic cells and neutrophils. Transfection experiments and functional studies are ongoing to investigate the role of the gene. on the disease.

1121F

Comprehensive genomic profiling of 66 cardiometabolic phenotypes by whole genome sequencing in 3,621 samples from the UK10K project. *J. Min, The UK10K Consortium Cohorts Group.* School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom.

As part of UK10K Consortium Cohorts Group, 3,621 individuals from two deeply phenotyped population-based collections - TwinsUK and the Avon Longitudinal Study of Parents and Children (ALSPAC) - have been whole genome sequenced (WGS) to average 6.5x coverage using next-generation sequencing technology. Variants discovered through WGS of the TwinsUK and ALSPAC cohorts along with those known from 1000 Genomes were imputed into the full genomewide association study genotyped cohorts increasing the sample size for single point association analysis to 12,724 subjects. In a meta-analysis of directly sequenced and imputed variants, we tested 8,8 million variants for association with 66 cardiometabolic phenotypes. In single point analysis, we found 30 novel loci with P<5e-8 of which 13 with an effect allele frequency below 5%. In addition, novel low frequency trait associations underlying known loci such as between *LDLR* and LDL and *LIPG* and total cholesterol were found. In WGS data only, we used SKAT to analyse 26,367 gene regions using 41,692 non-overlapping windows with a maximum of 50 coding variants with MAF<5%. We identified 18 regions below a gene-based significance threshold of P<1.9e-6 including known associations such as *ADIPOQ* and Adiponectin (P<3.7E-11) and *PCSK9* and LDL (P =1.8E-7). Novel loci are being validated and replicated. Among the single point associations with common variants (MAF>5%), we showed a consistent pattern of enrichment of exonic and 5'UTR variants using 100,000 permutations matched on TSS and MAF (p<1e-5, r2<0.1). The enrichment of exonic variants was also seen for single point associations with MAF 1-5%. As a substantial proportion of causal alleles may exist beyond the exome, we combined genome-wide maps of functional annotations including DNAseI hypersensitive sites (DHS) from six ENCODE cell types to estimate the pathogenicity of variants in regulatory or conserved regions. We found DHS enrichments of single point associations relative to sites not in DHS peaks, DHS hotspots or exons (p<1e-5, r2<0.1). We will investigate phenotype cell type specificity by comparing enrichments of cell-type specific sites to sites shared among multiple cell-types. Our results provide insights into how large-scale whole-genome sequencing efforts are likely to contribute to the understanding of the genetic architecture of cardiometabolic phenotypes and may therefore significantly increase previous estimates of explained heritability.

1122W

Functional annotation combined with evolutionary principles facilitate whole genome sequence analyses of complex traits: the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium. *A.C. Morrison[1], F. Yu[2], J. Lu[3], A. Voorman[4], A.D. Johnson[5,6], J. Reid[2], X. Liu[1], D. Muzny[2], A.R. Folsom[7], C.J. O'Donnell[5, 6], B.M. Psaty[8,9], L.A. Cupples[5,10], A. Clark[3], R. Gibbs[2], E. Boerwinkle[1,2], The Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium.* 1) Human Genetics Center, University of Texas at Houston, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY; 4) Department of Biostatistics, University of Washington, Seattle, WA; 5) NHLBI Framingham Heart Study, Framingham, MA; 6) Division of Intramural Research, NHLBI, National Institutes of Health; 7) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 8) Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA; 9) Group Health Research Institute, Group Health Cooperative, Seattle, WA; 10) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Joint analysis of functional annotation of genome variation, particularly outside of non-protein coding regions, combined with empirical evidence of natural selection indicating phenotypic consequences should prove to be an efficient approach to analyzing whole genome sequence data. Here we intersect the results of sequencing the whole genomes of 2,700 deeply phenotyped individuals from the CHARGE consortium, annotation using multiple predictive functional genomic resources (e.g. dbNSFP and ENCODE), and application of population genetic measures of variation (e.g. Watterson's ϑ, nucleotide diversity (π), and integrated haplotype score (iHS)) to infer regions of natural selection. Predicted functional domains (e.g. transcription factor binding sites) with significant evidence of natural selection are used as the unit of inference for genotype-phenotype associations. Whole genome analyses of a priori annotated functional domains (e.g. transcription factor binding sites and microRNAs) reveal an abundance of evidence for selection acting on noncoding regions of the genome. Among the regions with the strongest evidence of natural selection, ~60% lie in intergenic regions, 33% in intronic regions, and slightly over 1% in genic regions. Focused analyses stratified by functional classes of noncoding variants detects pervasive purifying selection acting on enhancers, transcription factor binding sites, microRNAs and target sites but not on lincRNA or piRNAs, suggesting different evolutionary constraints for these domains. Plasma lipids, uric acid levels, and hemostatic factors are being analyzed using this approach. In preliminary analyses of the first 1,000 individuals with whole genome sequence, multiple non-protein encoding loci have been identified that show both evidence of natural selection and phenotypic associations. Detailed results will be shown in select regions as examples. In the first example, variants in a piRNA cluster (chr4: 10114931-10157431) upstream of the *SLC2A9* genes have high derived allele frequencies and are significantly associated with serum uric acid levels. In a second example, a CETPβ transcription factor binding site shows both strong evidence of natural selection and a significant effect on plasma lipid levels. These results document the power of combining whole genomic sequence annotation with evolutionary principles to identify functional regions of the genome influencing health and disease.

1123T

Exome sequence variants associated with blood levels of hemostatic factors: The Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium and the NHLBI Exome Sequencing Project (ESP). *N. Pankratz[1], J. Brody[2], M.H. Chen[3], B. Davis[4], P. Wei[4], J. Reid[5], C.J. O'Donnell[6], N.L. Smith[2], A.C. Morrison[4] on behalf of the CHARGE Hemostasis Working Group and the ESP Hemostasis Working Group.* 1) University of Minnesota, Minneapolis, MN; 2) University of Washington, Seattle, WA; 3) Boston University, Boston, MA; 4) University of Texas Health Science Center at Houston, Houston, TX; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) NHLBI's Framingham Heart Study, Framingham, MA.

Perturbed hemostasis, reflected by altered levels of blood coagulation proteins, can lead to morbidity and mortality from arterial and venous thrombosis. Hemostatic factors such as fibrinogen, coagulation factors VII and VIII, and von Willebrand factor (vWF), are intermediate phenotypes that are associated with clinical thrombotic events; thus, it is important to study these traits and identify genetic variants that increase susceptibility to clotting. To identify associations with rare and low frequency variants, we performed exome sequencing in independent samples from CHARGE and ESP. Sequencing in CHARGE was performed in a subset of participants from the Atherosclerosis Risk in Communities Study (ARIC; n=3,173 whites, 2,408 blacks), the Cardiovascular Health Study (CHS; n=741 whites), and the Framingham Heart Study (FHS; n=850 whites). ESP cohorts with available hemostasis measures included ARIC, CHS, FHS, Coronary Artery Risk Development in Young Adults (CARDIA), Multi-Ethnic Study of Atherosclerosis (MESA), and Women's Health Initiative (WHI), for a total of 2,649 EAs and 1,925 AAs. Available measures of fibrinogen, factor VII, factor VIII, and vWF varied across studies and projects. Initial results (n=3,851 whites total) are described for the meta-analysis of ESP and an initial subset of the CHARGE data for these four hemostatic factors using three approaches: 1) single variant test using linear regression for all variants above 5% minor allele frequency (MAF); 2) a 'T5' test where all non-synonymous or splice site variants with a MAF<5% are summed together to generate a gene score; and 3) the SKAT test which analyzes the same variants as the T5 test, but allows for effects to be in either direction and upweights the contribution of rarer variants. We found genome-wide significant associations for fibrinogen (*FGB*, p=4E-7), factor VII (*F7*, p=2E-74), factor VIII (*ABO*, p=1E-11), and vWF (*ABO*, p=6E=15) using the single variant test. These initial exome sequence analyses confirm prior genome-wide association analyses of common variants. The gene-based tests will be fully evaluated in the total combined sample of 6,225 whites and 3,908 blacks across ESP and CHARGE. These efforts represent the largest integration of exome sequence data from two large national projects to identify novel genes and variants associated with inter-individual variation in hemostatic factors.

1124F

Exome sequencing study identifies several candidate variants associated with Kawasaki disease. *Y. Park, J. Kim, J. Lee, Korean Kawasaki Disease Genetics Consortium.* Asan Institute for Life Sciences, University of Ulsan College of Medicine, Seoul, Korea.

Kawasaki disease (KD) is an acute systemic vasculitis of infants and children, manifested by fever and signs of mucocutaneous inflammation. The highly effective and standard treatment for KD is high-dose intravenous immunoglobulin (IVIG) therapy. However, about 20 % of KD patients have persistent or recurrent fever after the initial IVIG treatment, which increases the risk for coronary artery lesions. To identify susceptibility variants for KD, we sequenced whole exomes of 12 patients with KD as well as 17 control exomes. As unique variants to KD compared to control exomes, a total of 43 novel nonsynonymous single-nucleotide polymorphisms (nsSNPs) were detected by exome sequencing and validated by capillary resequencing. To replicate the findings, additional 18 KD patients and 18 healthy controls were analyzed by direct sequencing. We examined the associations between 30 KD patients and 35 controls in the combined data (exome sequencing data and small scale of case-control data) analysis, and identified strong signal of association with KD in 8 loci. Among 8 loci, 4 genes (*PDE3A, TNFSF8, IL16* and *CLEC11*) had variants unique in KD samples only and the other 4 genes (*HLA-DQB1, SNAPC4, MLL* and *HAVCR2*) showed high odds ratio (OR > 4.9). Currently we are investigating whether those 8 nsSNPs are significantly associated with Kawasaki disease in large independent samples of KD case and control. Our initial data suggest that whole exome sequencing can be used for the identification of causal variants of KD.

1125W

On the association of common polygenic variation with body mass index in 7062 Han Chinese women using low pass sequencing: genome-wide association, genetic risk scores, and GCTA. *R.E. Peterson[1], T.B. Bigdeli[1], Y. Li[2], W. Kretzschmar[2], F. Yang[3], H.H. Maes[1], A.H. Fanous[1], S. Bacanu[1], B.P. Riley[1], J. Wang[4], S. Shi[5], Y. Chen[6], J. Marchini[2], R. Mott[2], K.S. Kendler[1], B.T. Webb[1], J. Flint[2], CONVERGE consortium.* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, PO Box 980126, Richmond, VA, 23298, USA; 2) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, Oxfordshire, United Kingdom; 3) Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, No. 600 South Wanping Road, Shanghai, P.R. China; 4) Beijing Genomics Institute, Floor 9 Complex Building, Beishan Industrial Zone, Yantian District, Shenzhen 518083, P.R. China; 5) Huashan Hospital of Fudan University, No.12 Middle Wulumuqi Road, Shanghai, P.R. China; 6) CTSU, Richard Doll Building, Old Road Campus, University of Oxford, Headington, Oxford OX3 7LF, United Kingdom.

The dramatic increase in obesity prevalence among both children and adults in developed countries and the numerous negative consequences associated with elevated body weight warrant additional research into the genetic and environmental contributions to body mass index (BMI). A mega-analysis of genome-wide association studies (GWAS) of BMI in populations of European descent has confirmed robust association with 32 single nucleotide polymorphisms (SNPs). Of these, seven loci demonstrated association with BMI in a meta-analysis of East Asian populations, including *FTO*, *MC4R*, and *BDNF*. The purpose of this study was to potentially discover variants associated with BMI, and to replicate previously identified associations using the CONVERGE (**C**hina, **O**xford a**n**d **V**CU **E**xperimental **R**esearch on **G**enetic **E**pidemiology) project, a study of 6,000 Han Chinese women with recurrent major depression and 6,000 matched controls. Currently, 7062 individuals have been sequenced genome-wide at 1.2x coverage, imputed genotypes estimated, and have BMI information. We investigated the extent to which common polygenic variation accounts for phenotypic variation in BMI in two ways. First, we applied Genome-wide Complex Trait Analysis (GCTA) to estimate, as a random effect, the cumulative genetic effect due to all SNPs genotyped. In aggregate, common SNPs accounted for 10.0% of the variance in BMI (95%CI=[4.1,15.9], p=0.0004) in CONVERGE. Compared to estimates of 17% for European populations using the same method, this suggests population differences in the heritability of BMI. Second, we determined the predictive value of an aggregate genotypic score by randomly dividing the sample into training and testing sets. Using the best linear unbiased prediction (BLUP) method to estimate SNPs effects in the training set, we subsequently tested their predictive value via an aggregate risk score in the testing set. The score was modestly associated with BMI (p= 0.004) accounting for 0.22% of the variance in the testing set. Preliminary GWAS results of ~7.5M SNPs indicated 3 variants reached genome-wide significance. However, we did not find robust association for previously identified BMI-associated variants (p_{FTO}=0.005, p_{MC4R}=0.0004, p_{BDNF}= 0.646). These preliminary results suggest genetic variation associated with BMI may only partially overlap between European and Han Chinese populations, highlighting the importance of studying genetic contributions to complex traits in diverse populations.

1126T

Rare functional variants in complement genes mitigate C3 inactivation and confer high risk of advanced age-related macular degeneration.
S. Raychaudhuri[1,2,3,4,5], Y. Yu[6], E.C. Miller[7], R.M. Reynolds[6], P.L. Tan[8], S. Gowrisankar[3], J.I. Goldstein[3,9], M. Triebwasser[7], H.E. Anderson[10], J. Zerbib[11], D. Kavanagh[10], E. Souied[11], N. Katsanis[8], M.J. Daly[3,9], J. Atkinson[7], J.M Seddon[6,12]. 1) Division of Genetics, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 4) Partners HealthCare Center for Personalized Genetic Medicine, Boston, Massachusetts, USA; 5) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 6) Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, Massachusetts, USA; 7) Division of Rheumatology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 8) Center for Human Disease Modeling and Departments of Cell Biology and Pediatrics, Duke University, Durham, North Carolina, USA; 9) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 10) Institute of Genetic Medicine , Newcastle University, International Centre for Life, Newcastle upon Tyne, UK; 11) Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France; 12) Department of Ophthalmology, Tufts University School of Medicine, Boston, Massachusetts, USA.

Motivation: We sought to identify rare coding variants which influence risk of advanced age-related macular degeneration (AMD), in order to better define disease mechanism. We were motivated by the discovery of the R1210C mutation in CFH by our group. This mutation confers high risk of AMD (OR~20) and causes loss of C-terminal
Methods: To find additional rare coding functional variants we applied next-generation sequencing to target the exons of 681 genes within AMD-associated loci and pathways in 2,493 cases and controls. We only identified one gene, CFI, with a significant burden of rare variants in cases. We observed that 7.8% of AMD cases compared to 2.3% of controls are carriers of rare missense CFI variants in our data (OR=3.6, $p=2\times10^{-8}$). The enrichment for predicted loss of function variants was particularly noteworthy, being seen in 7 cases compared to 0 controls. Many of these mutations are seen in familial forms of atypical hemolytic uremic syndrome. We also observed significant association with rare missense alleles outside CFI. Genotyping in 5,115 independent samples confirmed associations to AMD with a K155Q allele in C3 (replication $p=3.5\times10^{-5}$, OR=2.8; joint $p=5.2\times10^{-9}$, OR=3.8) and a P167S allele in C9 (replication $p=2.4\times10^{-5}$, OR=2.2; joint p=6.5x10-7, OR=2.2). Then, in functional binding studies we demonstrated that the 155Q allele in C3 results in reduced C3 binding to CFH, mirroring the function of the R1210C mutation in CFH. We then demonstrate that this reduced binding results in reduced cleavage and inactivation of 155Q compared to 155K C3 in cofactor assays by CFH and CFI mediated cleavage.
Conclusions: The combination of functional studies, and rare variants of large effect in C9, CFI, CFH, and C3, implicate excessive alternative pathway complement activation in AMD pathogenesis, in particular by dysregulation of C3. This study informs the mechanistic underpinnings of this disorder.

1127F

Exome sequencing in Autism Spectrum Disorder.
S. Walker[1], R. Yuen[1], B. Thiruvahindrapuram[1], L. Lau[1], C.R. Marshall[1], B. Fernandez[2], E. Fombonne[3], W. Roberts[4], L. Zwaigenbaum[5], P. Szatmari[6], S.W. Scherer[1]. 1) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Memorial University of Newfoundland, Disciplines of Genetics and Medicine, St. John's, A1B 3V6, Canada; 3) Montreal Children's Hospital and McGill University, Department of Psychiatry, Montreal, H3Z 1P2, Canada; 4) The Hospital for Sick Children, Autism Research Unit, Toronto, M5G 1L7, Canada; 5) University of Alberta, Department of Pediatrics, Alberta, T5G 0B7, Canada; 6) McMaster University, Department of Psychiatry and Behavioural Neurosciences, Hamilton, L8S 4K1, Canada.

To discover rare genetic variants and new genes associated with Autism Spectrum Disorder (ASD), we are performing a detailed genomic analysis in a cohort of 1000 Canadian Families. By combining high resolution microarrays with exome and whole genome sequencing, we aim to investigate the role of *de novo* and rare inherited mutations, incorporating copy number (CNV), single nucleotide (SNV) and insertion/deletion variants.

From the first 700 individuals analysed by exome sequencing with SOLiD 5500xl, we find in the order of 24,000 single nucleotide variants per individual, of which approximately 350 are novel and in coding regions. We have uncovered numerous rare and novel inherited variants that appear to segregate with the phenotype in genes previously associated with ASD such as *NRXN1*, *NLGN4X*, *ARID1B* and *CHD8*, including mutations resulting in likely haploinsufficiency. *De novo* mutations have also been discovered implicating new genes such as *RIMS2*, *NTF3* and *LYPD6B* in the disorders. We have found multiple individuals carrying potentially pathogenic both CNVs and SNVs and in some cases more than one arising *de novo*. Additional incidental findings from our study have also instigated clinical follow-up such as early cancer screening in families with cancer syndromes.

Now, we are pursuing sequencing using semiconductor sequencing on the Ion Proton System. In the first trio family analysed we detected approximately 19,800 coding single nucleotide variants per person. In comparison with whole genome sequencing data from the same individuals, we identified 91% of the known exonic variants with an estimated false positive rate of 9%. Moreover, analyses of these data correctly recognised four known *de novo* mutations including three substitution variants in genes *KIAA1217*, *FAT3*, *STXBP5L* and a single base insertion in *USP54*.

Our data support a multigenic, multifactoral model for Autism susceptibility and highlight the necessity for extensive information of both genotypes and phenotypes to further our understanding of complex disorders.

1128W

Exome sequencing in schizophrenia quartets families identifies the patterns and rates of causal mutations and CNVs.
D. Zhou[1], Z. Zhang[1], Y. Liu[2], L. He[1,2]. 1) Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Bio-X Institutes, Shanghai Jiao Tong University, Shanghai 200030, People's Republic of China; 2) Institutes of Biomedical Sciences, Fudan University, 303 Mingdao Building, 138 Yixueyuan Road, Shanghai 200032, People's Republic of China.

Schizophrenia is known to be a multifactorial disorder with an underlying heritability that is deemed to be around 80%, without totally displaying Mendelian inheritance. Previous studies have commonly thought that many genes (or other functional genomic elements) are involved. However, the common variants and mutations in known genes confer only a limited effect on the phenotype, and these could not explain a substantial fraction of schizophrenia heritability. Here we examined the effects of the novel rare mutations, CNVs on schizophrenia pathogenesis by sequencing the whole exomes of about 300 subjects from 50 schizophrenia quartets families (only one cases in parents, more than two cases in children). We identified the rare mutations or large copy number alternations in each family. We will then use protein-protein interaction (PPI) network to evaluate the connection of the causal genes.

1129T

Identification of low frequency variants associated with albuminuria and kidney function in Danes with and without diabetes through exome-array analysis. *T.S. Ahluwalia[1], J. Bork-Jensen[1], N. Grarup[1], R. Ribel-Madsen[1], J.M. Justesen[1], M.N. Harder[1], T.H. Sparsø[1], T.O. Kilpeläinen[1], T. Skaaby[2], C. Christensen[3], I. Brandslund[4,5], M. Aadahl[2], M.E. Jørgensen[6], A. Linneberg[2], L. Husemoen[2], T. Lauritzen[7], T. Jørgensen[2,8,9], T. Hansen[1,8,10], O. Pedersen[1,8,11]*. 1) Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 2) Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark; 3) Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, Denmark; 4) Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark; 5) Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark; 6) Steno Diabetes Center, Gentofte, Denmark; 7) School of Public Health, Department of General Practice, University of Aarhus, Aarhus, Denmark; 8) Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 9) Faculty of Medicine, University of Aalborg, Aalborg, Denmark; 10) Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark; 11) Faculty of Health Sciences, University of Aarhus, Aarhus, Denmark.

Type 2 diabetes (T2D) is an increasingly common disease with 347 million people affected worldwide. About 35% of individuals with T2D attain microvascular complications of the kidney which may lead to end stage renal failure and death. Diabetic kidney disease has been shown to be partially caused by a genetic susceptibility component. We therefore sought to identify coding variants associated with kidney function through an exome-array based association study. Five cohorts were investigated: ADDITION-DK ($n_{non\text{-}T2D}$=610, n_{T2D}=1,574), Vejle (n_{T2D}=1,951), Inter99 ($n_{non\text{-}T2D}$=5,555, n_{T2D}=156), Health2006 ($n_{non\text{-}T2D}$=2,637), and Health2008 ($n_{non\text{-}T2D}$=652). Genotyping was performed by the Illumina Infinium HumanExome BeadChip, including 263,894 variants and 16,550 custom-designed single nucleotide polymorphisms (SNPs) selected from an exome sequencing study of 2,000 Danes. After filtering the variants for minor allele frequency (MAF) ≥ 0.1% and genotyping call rate of > 0.90, the final associations were examined in 72,775 remaining variants. Kidney function was estimated by the Cockroft Gault formula, denoted by estimated glomerular filtration rate (eGFR, ml/min/1.73m^2), while albuminuria, by urinary albumin-to-creatinine ratio (UACR), was estimated from baseline measures. UACR measures were available for 13,153 individuals and eGFR measures for 7,402 individuals. Both traits were rank-normalized and analyzed separately by linear models in each cohort. Adjustments for age, sex, and four principal components were made and an additive genetic model was used. Meta-analysis of these results was then performed in three groups: T2D, non-T2D, and one including all individuals, using fixed effects meta-analysis. In a meta-analysis comprising all individuals, the strongest association with UACR (Pmeta=3.3×10^{-6}) occurred at a rare (MAF=0.9%) missense variant in the cubilin gene (*CUBN*), which was not in linkage disequilibrium with previously identified common *CUBN* SNP (r^2=0.001). Although the variant did not reach our study-wise significant threshold of $P<3.4 \times 10^{-7}$ (Bonferroni correction for 72,775 SNPs and 2 traits), it may represent a secondary signal. The *CUBN* gene is involved in the receptor-mediated endocytotic reabsorption of albumin in the kidney. We also replicate 13 GWAS-identified common loci associated with eGFR or UACR (P<0.05). A deeper understanding of the genetic basis of diabetic kidney disease may help to develop novel therapeutic strategies.

1130F

Exome-wide association study for the identification of genes for Primary Open Angle Glaucoma (POAG). *Z. Li[1], CC. Khor[1,2,3,4], Y. Ikeda[5], L. Jia[6], LJ. Chen[7], M. Nakano[8], R. George[9], T. Do[10], K. Abu-Amero[11], LS. Tajudin[12], C. Kee[13], O. Mineo[14], M. Takanori[15], M. Zhang[16], JB. Jonas[17], TN. Chau[18], N. Fuse[19,20], TY. Wong[1,4,25], JJ. Liu[2,4], DF. Garway-Heath[21], CP. Simmons[22], L. Vijaya[9], S. Kinoshita[5], CP. Pang[7], NL. Wang[6], RR. Allingham[23], MA. Hauser[24], K. Tashiro[8], T. Aung[1,25], EN. Vithana[1,25]*. 1) Singapore Eye Research Institute, Singapore National Eye Center, Singapore; 2) Infectious Diseases; Human Genetics; Genome Institute of Singapore, Singapore; 3) Department of Paediatrics, National University Health System & National University of Singapore, Singapore; 4) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 5) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan,; 6) Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology & Visual Sciences Key Lab, Beijing, China; 7) Department of Ophthalmology & Visual Sciences, the Chinese University of Hong Kong, Hong Kong, China; 8) Department of Genomic Medical Sciences, Kyoto Prefectural University of Medicine, Kyoto, Japan; 9) Vision Research Foundation, Sankara Nethralaya, Chennai, India; 10) Vietnam National Institute of Ophthalmology, Hanoi, Vietnam; 11) Department of Ophthalmology, College of Medicine, King Saud University, Riyadh, Saudi Arabia; 12) Department of Ophthalmology, School of Medical Sciences, Universiti Sains Malaysia, Kota Bharu, Malaysia; 13) Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Korea; 14) Ozaki Eye Hospital, Hyuga, Japan; 15) Mizoguchi Eye Clinic, Sasebo, Japan; 16) Shantou University / Chinese University of Hong Kong Joint Shantou International Eye Center, Shantou, China; 17) Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karls-University Heidelberg, Heidelberg, Germany; 18) Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; 19) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Sendai, Japan; 20) Department of Ophthalmology, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan,; 21) NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 22) Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, UK; 23) Department of Ophthalmology, Duke University Medical Center, Durham, NC, USA; 24) Center for Human Genetics, Department of Medicine,Duke University Medical Center, Durham, NC, USA; 25) Department of Ophthalmology, National University Health System & National University of Singapore, Singapore.

Purpose: The common, late onset form of POAG has a substantial hereditary predisposition. The purpose of this study was to identify protein-changing mutations within the human genome that show strong association with POAG. **Methods:** We used the exome-wide association approach by deploying the customized Human Exome Genotyping Bead Chip manufactured by Illumina; this particular array contains ~250,000 *loci* of base content as well as an additional 30,000 protein-changing mutations specific only to individuals of East Asian descent. Through collaborations with investigators from China, Hong Kong, Vietnam, Japan, India, and the USA, we have collected and genotyped total of 3,822 cases and 10,426 controls. All cases were >40 years old and comprised of both Normal Tension Glaucoma (NTG) and High Tension Glaucoma (HTG). After uniform and stringent quality control filters were applied for both individual samples and SNP markers, 3,540 cases and 9,744 controls across 7 independent sample collections (Singapore, China, Hong Kong, Vietnam, Japan, India, USA) remained. We measured the association between SNP genotypes and POAG risk using unconditional logistic regression. Further compensations were made for the top axes of genetic stratification for each independent collection to ensure minimal confounding due to differing ancestries between POAG cases and controls. Meta-analysis summarizing the results across all cases and controls was performed using fixed effects modeling weighted in an inverse-variance manner. **Results:** We have confirmed strong evidence of association with variants of several genes, including *CDKN2BAS* (OR =0.74, $P = 1.27 \times 10^{-9}$) and several novel *loci*. Novel variants underlying POAG are being validated in an additional 4,000 cases and 11,000 controls that have been independently collected. **Conclusions:** We have performed a meta-analysis of a large multi-ethnic and racial dataset that has confirmed the associations between *CDKN2BAS* and POAG. In addition we have identified several novel genetic variants that are in the process of validation at this time. **Grant support:** This work is supported by Biomedical Research Council (BMRC) in Singapore, Ref: BMRC 10/1/35/19/675.

1131W

Rare Variant Association of von Willebrand Factor Levels in a Healthy Young Cohort of European Individuals Using HumanExome BeadChip Data. *A.B. Ozel[1], K. Desch[2], D. Siemieniak[3], D. Ginsburg[1,2], J. Li[1].* 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Pediatrics, University of Michigan, Ann Arbor, MI; 3) HHMI, University of Michigan, Ann Arbor, MI.

von Willebrand factor (VWF) is an abundant plasma glycoprotein that regulates hemostasis. Plasma levels of VWF are highly variable among healthy people, with 65% of its variance attributable to inherited factors. We previously performed genome-wide association and linkage analyses of WVF in a healthy sibling cohort of 1,152 subjects and a second healthy cohort of 2,310 individuals (Desch et al., 2012). VWF plasma levels showed significant association with variants at the *ABO* locus (p<7.9E-139) and *VWF* locus (p<5.5E-16), consistent with previous reports. Linkage analysis using sibling data identified significant signals at a ~35cM interval on 2q12-2p13 (LOD=5.3), and a ~10cM interval on 9q34 at the *ABO* locus (LOD= 2.9), which explained 19.2% and 24.5% of the variance in VWF levels, respectively. The presence of a strong linkage signal and its absence in association studies suggest a causal locus harboring many genetic variants that are individually rare, but more common in aggregate. To investigate the role of rare variants (MAF<=5%) we collected genotype data on 940 European individuals in the first cohort using the Illumina HumanExome Beadchips. Most of the >240K variants were rare coding variants, with an average call rate of 99.7% after data clean-up. Common variants on the exome chip replicate previous findings of significant association at *ABO* (p<1E50). Preliminary analysis of 211,149 low-frequency variants (MAF<= 5%, HWE<1E-6) in 15,960 genes using the Sequence Kernel Association Test optimal (SKAT-O) gene-burden test showed a gene-wise significant association for 20 rare variants at *SLFNL1* (Schlafen-Like 1) on chromosome 1 (p<2E-6). Schlafen-family genes may have regulatory roles in haematopoiesis due to their increased expression levels during myeloid differentiation. The top signal in the chromosome 2 linkage interval was for 21 variants at the *RETSAT* gene (Retinol Saturase) (p<8E-3), which plays a role in metabolic adaptation and body weight maintenance. Further tests involving other gene-burden tests (e.g., Variable-Threshold (VT), the method by Madsen & Browning) are under way. These analyses of low-frequency variants could provide new insight into the regulation of hemostasis through VWF and reveal novel loci of extensive heterogeneity that may explain a significant portion of the missing heritability for VWF.

1132T

Exome Chip-based Association Analysis Identifies Novel Coding Variants Associated with Adiposity Traits in Hispanic Americans: The IRAS Family Stud. *N. Palmer[1], C. Gao[2], J. Norris[5], J. Rotter[6], L. Wagenknecht[3], D. Bowden[1], E. Speliotes[7], C. Langefeld[4].* 1) Biochemistry Dept, Wake Forest School of Medicine, Winston-Salem, NC; 2) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 3) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Epidemiology, Colorado School of Public Health, Aurora, CO; 6) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 7) Division of Gastroenterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

Despite recent successes of genome-wide association studies (GWAS) in identifying genetic variants that affect adiposity traits, a substantial proportion of the heritability remains unexplained. A potential role for low-frequency and rare variants has been proposed. Exome chip genotyping, a cost-effective approach to evaluate coding variation, was performed in 1,414 Hispanic Americans from the Insulin Resistance Atherosclerosis Family Study (IRASFS). Evaluation of 81,560 autosomal variants for the full complement of adiposity-related traits, including body mass index (BMI), waist circumference, waist-to-hip ratio (WHR), computed tomography (CT) measures of visceral (VAT) and subcutaneous (SAT) adipose tissue, and DEXA measured percent body fat, we identified two loci at exome-wide significance (P<0.05/81,560=6.13E-07). A novel coding variant in *IDH1* (rs34218846, V178I, 7% MAF; P=2.51E-07-2.14E-08) obtained genome-wide significance (P<5.0E-08) for BMI, waist and WHR. This gene was further supported by complementary GWAS and imputation data (P=0.019-4.97E-06). *IDH1* encodes cytosolic isocitrate dehydrogenase 1 and is postulated to play an important role in fat and cholesterol biosynthesis as demonstrated in cell and knock-out models. In addition, a rare missense variant in *ALDH1A3* (rs1130737, R15G, 0.01% MAF; P=2.39E-07) was associated with WHR. Although aldehyde dehydrogenase 1 has been examined for its role in sex- and depot-specific fat formation, gender stratification did not improve the observed association (P_{women}=2.72E-03). Analysis of CT measured adiposity identified variants approaching exome-wide significance in *NCF2* (P= 2.30E-05-8.51E-06) with SAT. Burden testing of non-synonymous variation at this locus did not improve evidence of association (P=1.9E-04). No significant signals were observed in the evaluation of percent body fat (P>2.60E-05). These results highlight that examining exonic variation can identify novel variants and genes that influence adiposity. Since the exome chip captures coding variation present predominantly in Europeans and African Americans, a comprehensive assessment of exonic variation through exome sequencing in this Hispanic American population will complement and extend the current findings.

1133F

A tool for co-segregation analysis using Whole Exome Sequencing data. *T. Gambin[1], B. Yuan[1], E. Boerwinkle[2,3], J. Lupski[1,4,5].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 3) The Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Texas Children's Hospital, Houston, TX.

Whole Exome Sequencing (WES) has proved to be a powerful method in studying Mendelian diseases. However, the identification of causative variants remains challenging due to the very large number of variants observed in each individual's exome. To narrow down the original set of variants to a short list of candidate causal mutations, a large number of tools have been developed, most of which are designed for annotating, filtering and prioritizing variants using allele frequencies in control/disease populations and computational predictions of deleteriousness and conservation of sequence variants and indels. Multiple methods are capable of utilizing trio information to detect de novo variants and/or exclude mutations violating expected Mendelian segregation. However, few published and freely available tools address the problem of selection/prioritization of variants that co-segregate in a family with a Mendelian condition under pre-specified inheritance model. All of them have significant limitations, such as low specificity or unavailable source code. Here, we present an open source R package for co-segregation analysis that incorporates genotype data from WES, pedigree structure and phenotypic information about individuals. In the first step, the algorithm filters out all common variants and generates a 'project level VCF' based on the genotype data from all family members, i.e. the file with genotype calls across all samples in pedigree at every variant position called in at least one sample. Next, for each variant, the pattern of co-segregation of genotype and phenotype information is evaluated under different inheritance models. At the same time, variants detected to be likely de novo mutations are distinguished from others. The list of variants observed to be co-segregating with disease can then be filtered/prioritized based on other criteria, such as predicted functional effects, relevant tissue expression and a priori candidate gene. Our algorithm pays particular attention to a low quality genotype calls by being stringent in looking for variants in affected individuals and lenient in checking the presence of the same variants in unaffected ones. The proposed algorithm is context dependent and can 'learn' from early samples and then applied to newly recruited samples. We are successfully applying this algorithm to data from NHGRI's Centers for Mendelian Genetics, and, once published, the code will be freely available.

1134W

Summarizing polygenic relative risks due to risk alleles for common complex diseases for the MedSeq project. *SW. Kong[1], CA. MacRae[2], IH. Lee[1], HL. Rehm[3], P. Kraft[4], J. Krier[2], JL. Vassy[2], RC. Green[2], IS. Kohane[1,5].* 1) Informatics Program, Boston Children's Hospital, Boston, MA; 2) Department of Medicine, Brigham & Women's Hospital, Boston, MA; 3) Department of Pathology, Brigham & Women's Hospital, Boston, MA; 4) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 5) Center for Biomedical Informatics, Harvard Medical School, Boston, MA.

Known disease-causing mutations and loss-of-function variants are of primary interest for personal whole-genome sequencing (WGS) in a clinical context. However, estimating the genetic liability for common complex diseases using known risk alleles would also be informative for clinical sequencing. The recent success of polygenic risk scoring methods for common diseases such as schizophrenia and type 2 diabetes has demonstrated that a larger proportion of the total liability can be explained by hundreds to thousands of risk alleles. Here we present an outline of a clinical WGS report for common risk alleles within the MedSeq Project, a randomized clinical trial that aims to develop standards and procedures for the evaluation and reporting of genome sequencing data. We have developed a novel method for summarizing known risk alleles focused at present on cardiac diseases- the Cardiac Risk Report (CRR). For 8 clinical phenotypes that are known to be prevalent in both primary care and cardiomyopathy populations, we provide multiplicative polygenic risk scores (MPRS) derived from 161 published risk alleles with small or moderate effect size (odds ratio < 1.2, on average). For each clinical phenotype, we used the product of odds ratio of risk allele each raised by its count as the MPRS. We then normalized the MPRS by the population median from the 379 individuals of European ancestry in the 1000 Genomes Project. Interestingly, the distribution of MPRS for type 2 diabetes showed a significant difference between ethnic groups, suggesting that an individual's disease risk from genetic variation should be carefully interpreted in the context of his or her genetic ancestry. In addition to normalized MPRS, the CRR includes contextual data such as the disease prevalence and heritability, all of which are important elements in the interpretation of modern multi-locus disease risk models. Our approach for the MedSeq Project demonstrates how complex trait risk variants from an individual genome can be interpreted and reported in clinical context.

1135T

Whole exome sequencing of families with multiple cases of adolescent idiopathic scoliosis implicate novel mutations in *FBN2*, *MESP2* and *SNTG1*. *A. Sasson[1], J.P. Dormans[2], F.J. Salley[2,3], C.E. Kim[3], S. Deliard[4], J. Talarico[2], J.C. Perin[1], RM. Chiavacci[3], H. Hakonarson[3,4,5], S.F.A. Grant[3,4,5].* 1) Center for Biomedical Informatics, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; 2) Department of Orthopedic Surgery, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Untreated scoliosis, particularly in more severe cases, has a detrimental influence on health throughout life plus a negative socioeconomic impact with respect to work and marital status. Approximately three quarters of structural scoliosis is clinically classified as idiopathic, which is the most common spine deformity arising during childhood. One of the main sub-forms of the disorder is adolescent idiopathic scoliosis (AIS), which presents in children aged 10 to 16 years old. There is strong evidence for a genetic component to idiopathic scoliosis, widely considered to be polygenic. It often appears in several members of the same family; indeed, one in three cases is familial. Also studies on twins showing that concordance of monozygotic twins is greater than that of dizygotic twins. Classical candidate gene studies and GWAS have only achieved limited success in identifying genetic determinants of idiopathic scoliosis. As such we elected to perform whole exome sequencing of members of five families with multiple cases of AIS on the Illumina HiSeq platform, based on recruitment from the Department of Orthopedic Surgery at the Children's Hospital of Philadelphia. Variants considered in our analyses had to be unique i.e. not reported in public databases, and were common to all cases, but not present in unaffecteds, in each given family. Subsequent filtering with PolyPhen revealed a shortlist of candidate variants in each family. Interestingly, within the respective shortlists for three of the families, we observed a novel variant in a different previously implicated scoliosis-related gene in each setting. In Family 1, where we sequenced three affected and four unaffected members, we observed a L1309F non-synonymous variant in *FBN2* (PolyPhen score=1). In Family 2, where we sequenced three affected and two unaffected members, we observed a K26Q non-synonymous variant in *SNTG1* (PolyPhen score= 0.759). In Family 3, where we sequenced two affected and three unaffected members, we observed a S224F non-synonymous variant in *MESP2* (PolyPhen score=0.906). Efforts are now underway to validate these novel observations further, test in additional sub-forms of idiopathic scoliosis and to investigate their functional role in the pathogenesis of the trait.

1136F

Identification of *de novo* variants contributing to nonsyndromic cleft lip and palate. *E.J. Leslie[1,6], K.M. Steinberg[2], D.C. Koboldt[2], C. Harris[2], D.E. Larson[2], R.S. Fulton[2], G.L. Wehby[3], J.T. Hecht[4], T.H. Beaty[5], A. Scott[5], M.L. Marazita[6], G.W. Weinstock[2], J.C. Murray[1].* 1) Dept. Pediatrics, University of Iowa, Iowa City, IA; 2) The Genome Institute, Washington University, St. Louis, MO; 3) Depart of Health Management and Policy, University of Iowa , Iowa City , Iowa; 4) Dept. Pediatrics, University of Texas Medical School at Houston, Houston, TX; 5) Dept. Epidemiology, Johns Hopkins University, Baltimore, MD; 6) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a common birth defect, affecting 1/1000 individuals worldwide. The complex inheritance of NSCL/P reflects the combined action of multiple genetic and environmental risk factors. Efforts to identify genetic risk factors have included numerous linkage, candidate gene, and genome-wide association (GWA) studies. To date, there have been four independent GWA studies and a meta-analysis which have collectively identified genome-wide significant associations for at least a dozen loci. Despite this progress, identifying specific, etiologic variants remains a challenge. We selected thirteen regions from previous GWA and candidate gene association studies for targeted capture and deep sequencing to identify common and rare etiologic variants for NSCL/P. We sequenced these regions, totalling 6.7Mb, in 1200 case-parent Asian trios from the Philippines and China plus 400 trios of European descent. Sequencing trios affords us the opportunity to identify *de novo* mutations which may be more deleterious than other rare variants and could contribute to the missing heritability in complex traits. *De novo* variants were called using the software tool Polymutt, which considered relationships between subjects. We identified 135 *de novo* variants absent from dbSNP or the NHLBI Exome Variant Server. These 135 variants included three missense *de novo* variants in *PAX7*, *ABCA4*, and *PIK3R5* and a splice site variant in *IRF6*. Of the remaining noncoding variants, 51 occurred in putative regulatory elements defined by histone modifications, transcription factor ChIP-Seq, and DNaseI hypersensitivity clusters. Notably, two of the *de novo* variants occurred within the same craniofacial enhancer, located 450kb upstream of *IRF6*. These data show that *de novo* variants are likely contributors to the etiology of NSCL/P. Future analysis of these variants may facilitate identification of inherited rare and common causative variants at these targeted regions.

1137W

Exome sequencing to identify de novo and rare recessive mutations in sporadic ALS. *K. Meltz Steinberg[1], D.C. Koboldt[1], D.E. Larson[1], G.E. Sanderson[1], R. Pamphlett[2], E.R. Mardis[1].* 1) The Genome Institute, Washington University, St. Louis, MO; 2) The Stacey Motor Neuron Disease Laboratory, Department of Pathology, The University of Sydney, Sydney, New South Wales, Australia.

Amyotrophic lateral sclerosis (ALS) is characterized by progressive motor neuron degeneration. Five to ten percent of patients have a family history of ALS (FALS), but most cases are sporadic (SALS). While the almost two thirds of FALS cases can be explained by mutations in a few genes, such as *SOD1*, the causes of SALS are still largely unknown. The late onset of disease further complicates efforts to elucidate the genetic causes of SALS using traditional family based studies. To further investigate the genetic contribution to SALS we performed exome capture and sequencing on 45 trios with no family history of ALS. The average age of the patients was 47.5, and two patients were positive for hexanucleotide repeat expansion of *C9orf72*, which is known to contribute to the ALS phenotype. Using the trio-aware genotype caller, Polymutt, we identified 22 high confidence de novo coding mutations, and twelve of these were identified as deleterious or damaging using bioinformatic prediction software. We also identified a splice site mutation and one nonsense mutation. Of note, some of the de novo mutations characterized are in genes involved in axonal guidance and apoptosis. In addition, we identified a de novo start codon mutation in *CHRM1*, a gene that harbors another de novo missense variant from a recently published, independent SALS cohort. Leveraging the trio information, we were also able to identify rare (MAF<0.01) compound heterozygous mutations in 18 different genes that are predicted to be highly damaging and cause a loss of function. We also identified five rare homozygous stop gain mutations and three rare homozygous splice donor/acceptor mutations. These de novo and rare recessive mutations may provide insight into the genetic contribution to SALS.

1138T

Novel intellectual disability genes identified by exome sequencing. *R. Rabionet[1], L. Domenech[1], O. Drechsel[2], M. Viñas[4], A. Puig[1], M. Gehre[1], S. Ossowski[2], I. Madrigal[3], M. Guitart[4], M. Mila[3], X. Estivill[1].* 1) Center Genomic Regulation (CRG), UPF and CIBERESP, Barcelona, Spain; 2) Center Genomic Regulation (CRG) and UPF, Barcelona, Spain; 3) Servei de Genètica, Hospital Clínic de Barcelona, Barcelona, Spain; 4) Laboratori Genètica, UDIAT-Centre Diagnòstic, Corporació Sanitària Parc Taulí, Sabadell, Spain.

Intellectual disability (ID) is a genetically heterogeneous disorder affecting 1-3% of the population. About 30% of cases of ID can be explained by structural variants, and more than 100 genes have been implicated in its pathophysiology. Nevertheless, a large proportion of cases remains unexplained. A recent study has shown that *de novo* point mutations are a frequent cause of ID, making next generation sequencing technologies very powerful tools to identify rare *de novo* genetic causes of ID. We recruited 33 trio cases and 6 sibling pairs affected with idiopathic ID, all of them negative for the fragile X expansion and without cytogenetically visible abnormalities. Exome sequencing was performed, and non-synonymous variants present in the cases were selected and filtered based on their frequency in known databases. Then, *de novo* variants were identified, as well as potential X-linked or recessive genes (carrying homozygous or compound heterozygous variants in the case); the variants' functionality score was predicted (Condel), and they were annotated against a list of known ID-related genes. Novel variants were also filtered for their presence in an in-house database of exome data. We detected an average of 8520 non-synonymous variants in the cases, of which an average of 27 were considered *de novo*. Frequency and functionality based filtering reduced the number of potential candidate ID genes harboring *de novo* variants to 0-10 per case. The presence and *de novo* state of the identified candidate variants was confirmed by Sanger sequencing. In 4 cases we confirmed the presence of a *de novo* variant in previously described ID genes, and in 7 additional families, 9 genes with damaging *de novo* variants were considered as potential novel ID genes. Three cases carried variants in genes for known ID syndromes. Three male cases carried X-linked variants in known ID genes and 2 cases were homozygous or compound heterozygous for rare potentially damaging variants in ID genes. In 6 cases we identified X-linked (2) or recessive (3) candidate novel ID genes; one X-linked gene was mutated in two families. Of the 14 potential novel ID genes, 11 have functions related with neurite outgrowth, gene silencing or epigenetic control of gene expression, making them good candidates for novel ID genes. Further evidence for their involvement in ID will be collected by analysis of interaction networks, brain expression, and by targeted resequencing in additional isolated ID cases.

1139F

Rare coding variant association study of Inflammatory Bowel Diseases. *M.A. Rivas[1,2], T. Green[2], C. Stevens[2], J.H. Cho[3], J.D. Rioux[4], R.J. Xavier[5], M.J. Daly[2,6], NIDDK IBD Genetics Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Broad Institute, Cambridge, MA, USA; 3) Yale School of Medicine, New Haven, Connecticut, USA; 4) Université de Montréal and Research Centre, Montreal Heart Institute, Montreal, Quebec, Canada; 5) Center for Computational & Integrative Biology, Cambridge, MA, USA; 6) Analytical and Translational Genetics Unit Massachusetts General Hospital, Boston, MA, USA.

In a prior pilot study we reported evidence of independent rare coding variants in common variant association study (CVAS) regions associated to inflammatory bowel disease (IBD). In this study we explored the extent to which rare and low frequency variants contribute to susceptibility of IBD through a combined analysis of targeted exon resequencing and exome array data. We performed targeted exon resequencing analysis of 770 genes proximal to CVAS regions in 2495 IBD patients and 1,121 control subjects. We also used the Illumina HumanExome Beadchip array in 3,352 subjects (895 IBD, 2447 controls) to test association of coding variants with IBD. In the exome-chip data we did not find any genes passing exome-wide significance threshold (P<2.5×10-6) when testing putative loss of function (LoF) or protein altering mutations. For genes in CVAS regions we find suggestive evidence of association of putative LoF mutations at *IFIH1* (P=.0075) and *DPP4* (P=.0047). Interestingly, the putative LoF mutations at *IFIH1* in the exome chip data are the same mutations reported in Nejentsev et al. 2009 to confer protection to Type 1 Diabetes, whereas in IBD we find evidence of conferred risk. Furthermore, we find suggestive evidence of protein altering association at *NOD2* (P=4.3×10-6) and *IL23R* (P=4×10-4).

The goal of the targeted exon resequencing experiment was to identify additional rare putative loss of function mutations that in aggregate may highlight signal for association to disease and provide avenues for downstream functional follow-up. We observed a common protein truncating variant mutations driving association of CVAS signal at *FUT2* (p.W154X) and *GSDMB* (c.IVS6-2A>G); furthermore, additional putative loss of function mutations at NOD2 exclusive to cases (3 additional copies, 0 observations in controls and 0 additional observations in the Exome Variant Server from 6500 subjects) supporting that loss of function of NOD2 confers risk to IBD. We identified a premature stop-gain variant p.R381X in *IL23R* at the same site of the common missense protective associated variant p.R381Q exclusive to controls. Similarly, we find a putative loss of function mutation, p.R179X, at *RNF186*, a CVAS region associated to ulcerative colitis, exclusive to control subjects (6 copies). The presence of loss of function alleles exclusive to controls highlights key targets for inhibition.

1140W

Exome Array Identifies Novel Loci and Rare Variants Associated With Age-related Macular Degeneration. *Y. Yu[1], S. Raychaudhuri[2,3,4,5,6], R. Reynolds[1], J.I. Goldstein[4,7], E. Souied[8], M.J. Daly[4,7], J.M. Seddon[1,9].* 1) Ophthalmic Epidemiology and Genetics Service, Department of Ophthalmology, Tufts Medical Center, Boston, MA; 2) Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 5) Partners HealthCare Center for Personalized Genetic Medicine, Boston, Massachusetts, USA; 6) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 7) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 8) Hôpital Intercommunal de Créteil, Hôpital Henri Mondor, Université Paris Est Créteil, Créteil, France; 9) Department of Ophthalmology, Tufts University School of Medicine, Boston, Massachusetts, USA.

Meta-analysis of Genome-wide association studies (GWAS) has identified 19 common variants for age-related macular degeneration (AMD), however, a large proportion of the heritability of AMD is still unexplained and the functional consequences of these variants are still unclear. To fine-map candidate loci and examine the association of rare functional variants to AMD, we genotyped 3693 cases and 3942 controls of European ancestry using Illumina Human ExomeBeadchip with custom content. We confirmed rare AMD causal mutations in several alternative complement pathway genes recently reported by our group, including R1210C in *CFH* ($P = 3.8 \times 10^{-7}$, OR =19.3), K155Q in *C3* ($P = 1.6 \times 10^{-10}$, OR =3.4) and P167S in *C9* ($P = 5.0 \times 10^{-7}$, OR =2.0). In addition, we identified a novel common SNP on chromosome 16 significantly associated with the risk of AMD, and a new rare non-synonymous variant in a gene related to the innate immune pathway for which no common variants have been associated with AMD previously. This study supports the involvement of both common and rare variants in AMD pathogenesis. Additional variants in the low-frequency spectrum are likely to explain the missing heritability of AMD.

1141T

Exome sequencing in children with severe viral respiratory infections.
S. Asgari, P. McLaren, J. Fellay. School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

Background: Human genetic variation plays a key role in determining individual responses after exposure to infectious agents. Respiratory viruses are the most common pathogens leading to hospitalization in children under the age of 5. Infections are usually mild and self-limiting, however, 0.1-0.2% of all children require admission to Intensive Care Unit (ICU). The majority of these children are previously healthy without known risk factor. Here, we use exome sequencing to search for genetic variants conferring unusual susceptibility to viral respiratory infection in the pediatric population. Methods: Study participant are recruited since 2011 in Swiss and Australian pediatric ICU. Inclusion criteria are age 0-3, ICU admission due to viral respiratory infection, mechanical ventilation, and absence of known risk factor or co-morbidity. So far, we captured and sequenced the exome of 10 patients using the Agilent SureSelect 50Mb enrichment kit and the Illumina HiSeq2000. Paired-end, 120 bp reads were aligned to the hg19 reference genome using BWA. Quality control and variant calling were performed using GATK. We used VAAST to run a case-control analysis of genetic variation between our exomes and data from the 1000 Genomes Project. Finally, we used DAVID for pathway analysis. Results and discussion: Samples were sequenced at high coverage (72x on average). After filtering for high quality variants based on the QC metrics given by GATK, we observed an average of 8434 synonymous-coding, 7889 non synonymous-coding, 71 splice-site, 81 stop-gain and 7 stop-loss variants per sample. Using a VAAST score threshold of 30, the case-control analysis resulted in the identification of 288 genes enriched for potentially deleterious variants in cases. From this list, we filtered out 38 highly variable genes (HLA, mucins, olfactory receptors, T-cell receptors and zinc fingers) as well as an additional set of 73 genes that systematically obtained high VAAST scores in permutation analysis, resulting in a final number of 177 candidate genes. Functional classification of these candidates using DAVID showed enrichment for innate immunity genes. Our ongoing recruitment and sequencing effort will allow us to explore candidate variants, genes and pathways in a larger study population.

1142F

Testing rare coding variation for an impact on HIV-1 viral load through exome sequencing.
P.J. McLaren[1,2], P.R. Shea[3], D.B. Goldstein[3], J. Fellay[1,2]. 1) School of Life Science, EPFL, Lausanne, Vaud, Switzerland; 2) Institute of Microbiology, University Hospital Lausanne, Lausanne, Vaud, Switzerland; 3) Center for Human Genome Variation, Duke University School of Medicine, Durham, NC, USA.

Background: Common variants (maf>5%) in the MHC and CCR5 regions are known to influence set point HIV-1 viral load (spVL) yet explain only a portion of the total trait variance. The impact of rare coding variation on HIV-1 disease progression has not been as thoroughly investigated. Here we utilize exome sequencing in 391 HIV-1 infected individuals with stable spVL to look for rare and functional variants that mediate control of HIV-1 infection.

Methods: Set point HIV-1 viral load was calculated as the average of at least 3 measurements obtained during the chronic phase of infection. We captured and sequenced all coding exons to high coverage (>70x) in 391 HIV-1 infected individuals of the Swiss HIV Cohort Study using the Illumina Truseq 65Mb enrichment kit and the Illumina HiSeq2000. Paired-end, 120 bp reads were aligned using the Burrows-Wheeler Aligner (BWA). Quality control and variant calling were performed using the GATK. Variant functional annotation was performed using snpEff version 2.1. Individual variants were tested for association using linear regression. Gene and gene network burden testing was performed using SCORE-Seq and SKAT.

Results: Individual variant testing showed strong evidence for association in the MHC region. Conditional regression demonstrated that this signal is explained by two independent SNPs, rs1057151 (p=6.7e-11) in HLA-B and rs3207555 (p=4.0e-10) in HLA-C. Accounting for these two SNPs, no other variants show evidence for association. Testing for a relationship between burden of rare variants (maf<5%) and spVL in 16,839 genes individually did not uncover significant associations. Similarly, there was no evidence for enrichment of rare variants affecting spVL across a network of 2,971 genes identified as interacting with HIV-1 in biochemical screens.

Conclusions: Outside of the MHC, no significant impact of rare variation on spVL was detected by exome sequencing in 391 individuals. Larger samples are likely required to fully explore the role of rare coding variation on this phenotype. Additional classes of variation not detected by GWAS or current sequencing technologies may also contribute to host HIV-1 control.

1143W

Exome sequencing reveals a novel putative risk mutation for meningococcal disease.
A. Ndungu[1], K.S. Elliott[1], T.C. Mills[1], A. Rautanen[1], P. Hutton[4], C. Garrard[3], A. Gordon[5], C.J. Hinds[2], A.V.S Hill[1], S.J. Chapman[1,6], GAinS Investigators. 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Intensive Care Unit, Barts and The London Queen Mary School of Medicine and Dentistry, London; 3) Intensive Care Unit, John Radcliffe Hospital, Oxford; 4) John Radcliffe Hospital, Oxford; 5) Imperial College, London; 6) Respiratory Medicine, Churchill Hospital Site, Oxford Radcliffe Hospital, Oxford, UK.

Meningococcal disease is a major cause of meningitis and sepsis worldwide and follows invasive infection with the Gram-negative bacteria *Neisseria meningitidis*. Asymptomatic nasopharyngeal carriage of *N. meningitidis* is relatively common in the general population (approximately 5-10% in adults), yet in a small minority of individuals the pathogen is able to penetrate mucosal surfaces and invade the bloodstream leading to life-threatening disease. Host genetic variation is increasingly recognised as playing an important role in determining susceptibility to infectious agents, and common polymorphisms in the gene encoding complement factor H have previously been described at the genome-wide level to associate with a small increase in the risk of meningococcal disease in children. Meningococcal disease is an extreme phenotype which may be enriched for large-effect genetic variants of low frequency. To investigate this possibility further, we performed whole exome sequencing on 13 adults from the UK with well-defined invasive meningococcal disease that were recruited as part of the GAinS (Genomic Advances in Sepsis) project. One further case was excluded from sequencing because of a known primary immunodeficiency resulting from a complement pathway mutation. All cases were critically ill and admitted to a UK intensive care unit with sepsis. Over 84,000 mutations were identified as missense, nonsense, splice site variants or variants occurring in UTRs. These accounted for 17% of total variants while synonymous, intergenic and intronic variants accounted for the rest. From these, we identified a homozygous novel missense mutation (predicted to be deleterious by PolyPhen) in a complement pathway protein in one individual. Validation of this result via functional assays will further understanding of the genetic basis of meningococcal disease. Overall,taking into account the sequenced individuals and excluded patient, two of 14 critically ill adult patients with meningococcal sepsis in this study were found to have predicted deleterious mutations in complement pathway components. This finding raises the possibility that a significant proportion of severe bacterial disease may occur in association with unique large-effect mutations and provides proof-of-principle to support the use of next generation sequencing to study these phenotypes.

1144T

Cumulative effect of coding sequence variation in *TLR6* and *ENG* influences risk of infectious complications in patients with *Staphylococcus aureus* bacteremia.
W.K. Scott[1,2], D.M. Dykxhoorn[1,2], S. Guo[1], C.L. Nelson[3], T. Rude[4], F. Ruffin[4], A.S. Allen[3,5], Q. Yan[4], V.G. Fowler[3,4]. 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami FL; 3) Duke Clinical Research Institute, Duke University Medical Center, Durham NC; 4) Department of Medicine, Duke University Medical Center, Durham NC; 5) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham NC.

Staphylococcus aureus bacteremia (SAB) is a serious, common infection. Both host and bacterial genetic variation likely influence development of complications in SAB patients. Studies of common single nucleotide variants (SNV) have failed to detect significant, reproducible associations with SAB complications. Because such studies may miss the impact of rare coding SNV due to weak linkage disequilibrium (LD), we used whole-exome sequencing to determine if the cumulative effect of common and rare coding SNV influences risk of complicated SAB. To reduce the effect of bacterial strain, we analyzed patients infected with *S. aureus* clonal complexes 5 or 30. In the first phase of the 200-sample study, 35 white cases of complicated SAB were compared to 35 age- (within deciles) and sex-matched white SAB controls without complications. After DNA capture with the Agilent SureSelect 72Mb exome/UTRs kit, samples were sequenced 3 per lane on the Illumina HiSeq 2000. DNA sequences were aligned with BWA and genotypes called with GATK and VQSR recalibration. SNV with VQSLOD>-2 and quality score of 99 were analyzed. Genotypes with read depth<8 and probability<99 were removed before analysis. Association of genes with complicated SAB was tested by Cochran-Armitage (CA) -Sum and -Max tests using RVASSOC software. CA-Sum sums individual SNV chi-square test statistics across the gene and evaluates significance by permutation. CA-Max considers only the SNV with the largest chi-square statistic. While no gene was significantly associated with complicated SAB after Bonferroni correction ($p=0.05/17{,}830$ genes$=2.8\times10^{-6}$), the most significant genes ($p<10^{-4}$ on both tests) were *TLR6* (CA-Sum $p=1.7\times10^{-5}$, CA-Max p=7.9×10-5) and *ENG* (CA-Sum $p=2.9\times10^{-5}$, CA-Max $p=4.9\times10^{-5}$). The similarity in CA-Sum and CA-Max results suggests that the associations are driven by one SNV in each gene; the most significant are common synonymous SNV in *TLR6* (T361T and K421K, in strong LD) and *ENG* (L69L). Although the biological significance of these SNV is unclear, these genes are plausible risk factors for complicated SAB. *TLR6* (dimerized with *TLR2*) recognizes *S. aureus* cell surface proteins, activating innate immune responses. Mutations in *ENG*, a cell surface receptor involved in TGF-beta signaling, cause hereditary hemorrhagic telangiectasia, in which SAB is a common comorbidity. These results suggest that coding SNV in genes underlying the immune response may influence development of complicated SAB.

1145F

Implication of *CDH1* in two familial cases of nonsyndromic cleft lip with or without cleft palate.
L.A. Brito, G. Yamamoto, M. Aguena, M.R. Passos-Bueno. Department of Genetics, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a complex trait and the most common craniofacial birth defect. Familial cases account for 20-30% of the cases, mostly presenting non-mendelian pattern of inheritance. Association of common variants has been reported, but they are far from explaining the whole heritability of the disease. In this sense, rare variants might help to bridge the missing heritability gap. We carried out the exome sequencing of 4 affected individuals (HiScan SQ - Illumina Inc; mean coverage of 100x) of a large Brazilian family segregating NSCL/P in an autosomal dominant and incompletely penetrant fashion (family 1), and identified a nonsense mutation in the exon 6 (c.G760A:p.D254N) of E-cadherin gene (*CDH1*), which had never been described in any population. This mutation is located in a conserved region (LJB-PhyloP conservation score = 0.99) and is predicted to damage the protein function (LJB-SIFT score = 1; LJB-Polyphen score = 0.99 [scores ranging from 0 to 1, where 1 is the most deleterious prediction]). We screened, by Sanger method, 222 unrelated controls and 263 unrelated patients, and found the same mutation in only one patient. This patient belongs to a family with multiple affected members, from a region geographically close to that of family 1, raising the possibility of founder effect. Considering both families, we estimated the penetrance at 53%. CDH1 is largely known to be involved with familial diffuse gastric cancer, and it has also been implicated in breast, ovarian, endometrial and prostate cancer. A few authors have reported the presence of cleft lip and/or palate in individuals belonging to families segregating familial diffuse gastric cancer, suggesting a partially common etiology between these traits. Our families do not report any case of gastric cancer, which, in turn, suggests that *CDH1* might also be implicated with NSCL/P. Recently, one group described potentially deleterious mutations in CDH1 in NSCL/P individuals from European descent without known history of gastric cancer in their families. In conclusion, we describe the impact of a rare mutation underlying two familial cases of NSCL/P, in a gene which had never been implicated by common-variant association studies. In addition, our results support the involvement of CDH1 in NSCL/P etiology and encourage the mutational screening of this gene in familial cases of NSCL/P. FAPESP, CNPq.

1146W

Rare variants and risk for asthma in 7,224 individuals from ethnically diverse populations.
C. Igartua[1], RA. Myers[1], C. Ober[1], DL. Nicolae[1, 2, 3], The Eve Consortium. 1) Human Genetics, University of Washingon, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

Asthma is a chronic respiratory disease characterized by inflammation and obstruction of the airways, with heritability estimates between 20-80%. Although many common risk alleles have been identified for asthma, together they explain very little of the heritability. Because the genetic architecture of asthma likely involves a combination of common and rare alleles, we evaluated the role of rare variants on asthma risk in subjects of European, African, and Latino ancestry that are part of the EVE Consortium, a multi-center sample that includes 5,523 case-control and 567 parent-child trios. Subjects were genotyped using the Illlumina Infinium Human-Exome Chip array, which contains ~250k variants enriched for rare coding variants that are predicted to affect protein function. A gene-based association test for each gene on the exome chip that had at least 2 functional SNPs (nonsynonymous, stop or splicing) was performed using the optimized SNP-set Kernel Association Test (SKAT-O), and applying the first 2 principal components to adjust for population stratification. We considered a total of 142,744 variants in 13,727 genes in each study sample. Only the *PGLYRP4* gene (10 variants) in the African Caribbean sample was significant after multiple testing correction (p=2.04×10-6). When functional variants were weighted based on their allele frequencies (rarer variants weighted more), 3 genes approached genome-wide significance in African Americans: *RAPGEF6* (6 variants; p=1.1×10-5), *CUL7* (19 variants; p=1.2×10-5) and *SLC22A5* (12 variants; p=4.9×10-5). We also investigated genes near reported asthma GWAS SNPs and found only one instance (*SLC22A5*) where the association was driven by exonic variation. No genes were significant in the other ethnic groups or in the combined sample. These results indicate that rare variants may contribute more significantly to asthma in individuals of African ancestry, possibly due to the larger number of rare alleles in those populations. Our study, which was limited to variants present on the Human Exome array and conducted in relatively small samples, does not provide evidence for large-effect functional low-frequency exonic variants in the etiology of asthma.

1147T

Targeted next generation sequencing and functional genomics in alopecia areata identifies ULBP6 as a critical node in its genetic architecture. *L. Petukhova[1,2], E. Drill[1,3], Z. Dai[1], L. Bian[1], M. Duvic[6], M. Hordinsky[7], D. Norris[8], V. Price[9], R. Clynes[1,4], A.M. Christiano[1,5].* 1) Dept Dermatology, Columbia Univ, New York, NY; 2) Dept Epidemiology, Columbia Univ, New York, NY; 3) Dept Biostatistics, Columbia Univ, New York, NY; 4) Dept Medicine, Columbia Univ, New York, NY; 5) Dept Genetics & Development, Columbia Univ, New York, NY; 6) Dept Dermatology, MD Anderson Cancer Center,Houston,TX; 7) Dept Dermatology, Univ of Minnesota,Minneapolis,MN; 8) Dept Dermatology, Univ of Colorado,Denver,CO; 9) Dept Dermatology, UCSF,San Francisco,CA.

Alopecia areata (AA) is a highly prevalent and poorly understood autoimmune disease which targets the hair follicle causing disfiguring hair loss. There is an enormous unmet medical need for the 5.3 million patients in the US who suffer from AA, arising primarily from a lack of understanding of disease pathogenesis. Our initial GWAS in AA revealed the first disease association to ULBP3/6 genes in any human disease. These genes are ligands for the NKG2D activating receptors for a repertoire of leukocytes. We biologically validated our statistical evidence by showing a marked upregulation of ULBP3/6 in lesional hair follicles and the presence of CD8+NKG2D+ T cells within the immune infiltrate. These findings, together with the previous demonstration of MICA overexpression in AA hair follicles, placed the NKG2D axis squarely at the center of AA pathogenesis, and invited a functional genomics approach to uncover causal variants predisposing to disease. In order to better understand the genetic variation driving the tagSNP associations identified in our GWAS at the ULBP locus, we selected a subset of 124 cases from our GWAS cohort for targeted deep resequencing with RainDance technology, amplifying 72Kb of sequence encompassing the entire region of association. As preliminary analysis of this dataset, we looked at the distribution of rare variants (p<0.01 in EVS and 1000G) across this region. We identified two rare missense variants, one of which is highly overrepresented in our cohort (p=0.005) and is located within ULBP6. Of the 127 rare or novel variants located within intergenic regions we identified 34 that fall within transcription factor binding sites, 7 of which are overrepresented in our cases, which cluster into two regions. One of these regions is a CTCF binding site, which is known to influence chromosome structure providing a mechanism for the regulation of gene expression. We have begun to assess the biological consequences of the ULBP6 protein coding variant, by developing a battery of functional assays aimed specifically at examining mRNA expression levels, protein levels, effects on cell surface display, receptor binding affinity and killing efficiency. We are additionally developing cellular assays to determine the effects of the regulatory variants. This work will clarify how GWAS identified genetic variation influences NKG2D-mediated cytotoxicity in the pathogenesis of autoimmune disease.

1148F

Exploring the genome for the secrets of human longevity. *E.B. van den Akker[1,2], S.J. Pitts[3], M.H. Moed[1], S. Potluri[3], J. Deelen[1,4], J.J. Houwing-Duistermaat[5], D.R. Cox[3,6], M.J.T. Reinders[2], M. Beekman[1,4,7], P.E. Slagboom[1,4,7], Genome of The Netherlands Consortium.* 1) Molecular Epidemiology, Leiden University Medical Center, Leiden, Zuid Holland, The Netherlands; 2) The Delft Bioinformatics Lab, Delft University of Technology, Delft, Zuid Holland, The Netherlands; 3) Rinat-Pfizer Inc, South San Francisco, CA 94080, USA; 4) Netherlands Consortium for Healthy Ageing, Leiden, The Netherlands; 5) Leiden University Medical Center, Medical Statistics and Bioinformatics, Leiden, The Netherlands; 6) In Memoriam; 7) These authors contributed equally to this work.

The clustering of longevity in families is being studied worldwide and the genetic contribution to longevity in the population at large is estimated at 25%. This is illustrated by the fact that first-degree family members of nonagenarians and centenarians also have a life-long survival advantage that can be attributed to a lower risk of coronary artery disease, cancer and type-2 diabetes. The survival advantage in 420 Dutch long-lived families from the Leiden Longevity Study (LLS), which have been selected on the basis of living nonagenarian siblings (≥2), could not be explained by the absence of risk alleles at known disease susceptibility loci. To investigate whether rare private genetic variants contribute to their longevity, we explored the whole-genome sequence of 220 unrelated Dutch nonagenarians (mean age 94 years) in comparison with 95 younger controls from the general Dutch population. The nonagenarians were selected from the LLS study population on the basis of having the best family history of survival and the controls were randomly selected from Dutch biobanks included in the BBMRI-NL consortium consisting of healthy individuals. High quality whole-genome sequences of DNA from leukocytes were generated by Complete Genomics (USA) at >30x coverage. The mean numbers of SNPs, deletions, insertion and substitutions per individual are remarkably similar between controls and the nonagenarians, which may be regarded as supercontrols. The mean numbers of variants observed in the exome of controls were 3,344,320.57 SNPs, 224,143 deletions, 217,134 insertions and 81,475 substitutions and in nonagenarians 3,357,346.38 SNPs, 225,241 deletions, 213,847 insertions and 80,436 substitutions, respectively. Further, we investigated differences between the groups in the presence of longevity alleles at known loci (a.o. *TOMM40/APOE* and *FOXO3A*). Next, we explored all genes in a burden analysis and paid special attention to IIS and mTOR pathway related genes. For this analysis we selected all detected genetic variants with an impact on protein function according to ANNOVAR. Our preliminary data analysis indicated a tendency that for many exonic annotations the nonagenarians carry fewer variants than controls.

1149W

Simulation of Finnish population history, guided by empirical genetic data, to assess power of rare variant tests in Finland.
R. Wang[1,2,3], V. Agarwala[2,4,5], J. Flannick[2,3,6], D. Altshuler[2,3,6], J.N. Hirschhorn[1,2,3]. 1) Division of Endocrinology, Boston Children's Hospital; 2) Program in Medical and Population Genetics, Broad Institute; 3) Department of Genetics, Harvard Medical School; 4) Harvard-MIT Division of Health Sciences and Technology, MIT; 5) Program in Biophysics, Graduate School of Arts and Sciences, Harvard University; 6) Department of Molecular Biology, Massachusetts General Hospital.

The Finnish population has been extensively utilized in genetic studies. It is considered to be a relatively homogenous founder population, and hence well suited for gene mapping, especially for variants of lower frequency. However, without realistic population demographic and phenotypic models, it is difficult to assess the implications of the founder effect for association studies of rare variants in exome genotype or sequence data. In this study, we developed a population genetic framework to address these issues. We first confirmed the Finnish founder effect in deep resequencing data by showing that allele frequency spectra are shifted towards higher end in the Finns compared to non-Finnish Europeans (NFEs). Next, building on recent simulations that generate representative sequence data for European populations, we used empirical sequence data to develop a simultaneous forward simulation of sequence variation in the NFE population and the Finnish population. We then simulated phenotypes in these populations under different disease models characterized by different correlation between phenotypic effects of variants and their effects on fitness. With simulated genotype and phenotype data, we conducted rare variant statistical tests (single variant, burden tests), and assessed their performance by calculating power. Our results showed that founder populations such as the Finns could provide additional power especially when phenotypic effects of variants correlate well with effects on fitness. The single variant test shows good performance in a founder population under different disease models, and it is particularly powerful if rare variants play a big role in explaining phenotypic variance. Finally, we compared the efficiency of genotype data (exome chip) and sequence data (exome sequencing) under different scenarios. Our results suggest that exome chip is currently much more cost-efficient than exome sequencing, especially in a founder population. Our study has highlighted the usefulness of understanding the population-genetic properties of a study population and exploring a range of genetic models for recognizing the features and limitations of association studies in that population. As the field of human genetics moves forward to explore new and expanded sources of variation, such models offer a context with which to interpret the data and to plan future studies for gene discovery.

1150T

De novo and inherited retrotransposon insertions associated with autism revealed by whole genome sequencing.
X. Jin[1,2], A. Ewing[3], J. Ju[1], R. Yuen[4], J. Wu[5], Y. Jiang[6], M. Wang[1], A. Shih[7], Y. Li[1], J. Wang[1], S. Scherer[4], Z. Sun[5], H. Yang[1], H. Kazazian[8], J. Wang[1]. 1) BGI, Shenzhen, Guangdong, China; 2) School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China; 3) Center for Biomolecular Science and Engineering, University of California at Santa Cruz, Santa Cruz, California, USA; 4) The Hospital for Sick Children, Toronto, ON Canada; 5) The Institute of Genomic Medicine, Wenzhou Medical College, Wenzhou, China; 6) Department of Paediatrics and Neurobiology, Duke University School of Medicine, Durham, NC, USA; 7) Autism Speaks, New York, NY, USA; 8) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

Retrotransposons are genetic elements that are capable of moving by a 'copy-and-paste' mechanism to spread throughout genome. Around 45% of the human genome is composed of retrotransposons. Recent studies have revealed that somatic mosaicism driven by retrotransposition may reshape the genetic circuitry that underpins normal and abnormal neurobiological processes. Autism spectrum disorder (ASD) is a lifelong developmental condition that affects about 1 in 88 individuals. Multiple researches have confirmed the contribution of de novo and rare inherited mutation to the risk for ASD. Here we sequenced the whole genome of 232 individuals from 74 families with ASD collected in China and the US. We developed a novel method to identify both de novo and inherited retrotransposon insertions from paired-end whole-genome sequencing data. We indentified 2 high-confidence de novo retrotransposon insertions in genes MARK2 and RIMS2, both of which have been reported to contain de novo functional mutations in recent large-scale exome studies of autism. We also detected 1281 to 1725 (average 1530) inherited retrotransposon insertions in each family. Several inherited retrotransposition insertions were located in known autism related genes such as CTNNA3, AUTS2 and TCF4. Our findings suggest that de novo and inherited retrotransopon insertions may play an important role in autism genetics.

1151F

Exome sequencing of 55 multiply affected coeliac families and large scale resequencing follow up.
V. Mistry[1], N.A. Bockett[1], M. Muddassar[2], K.A. Hunt[1], S.L. Neuhausen[3], P.J. Ciclitira[4], V. Plagnol[5], D.A. van Heel[1]. 1) Centre for Digestive Diseases, Barts and The London School Of Medicine and Dentistry, Blizard Institute, 4 Newark Street, Whitechapel, London E1 2AT, United Kingdom; 2) Division of Genetics and Molecular Medicine, Kings College London School of Medicine, 8th Floor Tower Wing, Guy's Hospital, London SE1 9RY, United Kingdom; 3) Department of Population Sciences, City of Hope, Duarte, California 91010, USA; 4) King's College London, Division of Diabetes and Nutritional Sciences, Gastroenterology, The Rayne Institute, St Thomas' Hospital, Westminster Bridge Road, London SE1 7EH, United Kingdom; 5) University College London Genetics Institute, Gower Street, London WC1E 6BT, United Kingdom.

Coeliac disease is a highly heritable common autoimmune disease involving chronic small intestinal inflammation in response to dietary wheat. The HLA region, and 40 newer regions identified by GWAS and dense fine mapping (many immune genes), account for ≈ 40% of heritability. We hypothesized that rare mutations of larger effect size (OR ≈ 2 - 5) might exist, especially in multiply affected pedigrees. We exome sequenced 75 subjects from 55 multiply affected families. We selected interesting variants/genes for further follow up using a combination of: linkage, shared variants between multiple related subjects and gene burden tests for multiple potentially causal variants. We next performed highly multiplexed amplicon sequencing (Fluidigm) of all RefSeq exons from 24 candidate genes in 2,304 coeliac cases and 2,304 controls. High coverage data enabled direct genotyping (99.98% of all sample genotype calls had a read depth >40) and extensive quality control. 1,335 unique variants with a 99.98% genotyping call rate were observed in 4,478 (post quality control) samples, of which 939 were present in coding regions of 24 genes (Ti/Tv 2.99). 91.7% of coding variants were rare (MAF in 2,230 controls, <0.5%) and 60% were novel. No common or low frequency variants were seen at novel sites. Gene burden tests (C-Alpha, SKAT) performed with rare functional variants identified no rare significant associations ($P<1\times10^{-3}$) at the resequenced candidate genes. Our strategy of sequencing multiply affected families, and deep follow up of candidate genes, has not identified new disease risk mutations. Common variants (and other factors, e.g. environmental) may instead account for familial clustering in this common autoimmune disease.

1152W

Pathway analysis using whole exome sequencing in Parkinson disease.
K. Nuytemans[1], V. Inchausti[1], L. Maldonado[1], W. Perry[1], E.R. Martin[1,2], G.W. Beecham[1,2], L. Wang[1,2], W.K. Scott[1,2], J.M. Vance[1,2]. 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA.

Previously, we reported several biological pathways (KEGG) to be associated with Parkinson Disease (PD) using gene expression and GWAS data (Edwards 2011). The top 3 pathways included 'axon guidance', 'focal adhesion' and 'calcium signaling pathway'. These analyses are based on common, often intergenic or intronic genomic variants with low risk effects and downstream differences in transcript level. In contrast, variants identified in whole exome sequencing (WES) are enriched for 'rare risk variants'; variants with a low frequency but possible higher risk effect on disease. We set out to determine whether the combination of exome variants in the same pathways also provides evidence for association with increased PD risk. We performed WES in 315 PD patients and 344 controls (all unrelated). Association of sets of variants in pathways/gene groups with increased risk for PD was assessed using the Cochran-Armitage sum and max test implemented in the RVASSOC program and the Sequence Kernel Association test (SKAT). Analyzed pathways included 13 top KEGG pathways identified in the previously reported pathway study as well as candidate gene groups 'axonal transport', 'mitochondrial' (MitoCarta) and 'lysosomal' genes (Human Lysosome Gene Database). On average, ~85% and ~73% of exome variants in these pathways have a frequency <5% and <1% respectively. Only minimal significant results (0.05>p>0.01) were obtained for 5 out of 13 tested KEGG pathways when analyzing all exome variants. However, we observed evidence for association with PD risk for the 'mitochondrial' and 'lysosomal' gene groups (p<0.001). Permutation tests (N=500) on random sets of the same number of genes or variants as the original groups showed that the original groups are not significantly more associated than the random gene sets. Interestingly, when filtering on variant frequency (<5% or <1%) we obtained similar results for all but the 'lysosomal' gene group. A significant p-value for this group was observed in both rare variant analyses (p<0.001; $p_{perm}<0.02$). No strong evidence for association of previously reported pathways with PD risk could be identified when analyzing all exome variants. Further filtering on frequency indicated that rare variants in the 'lysosomal' gene group are significantly contributing to disease. Additional analyses on smaller candidate gene groups and pathways with variants filtered on function will be performed to elucidate the observed signal in this analysis.

1153T

Whole-genome Sequencing in a Multiplex Nuclear Family of Schizophrenia to Identify Its Rare Susceptibility Variants. *S.C. Yu[1], H.Y. Chen[2], S.L. Yu[3], C.M. Liu[4,5], H.G. Hwu[1,4,5], W.J. Chen[1,4,5,6].* 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, Nat, Taipei, Taiwan; 2) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan 115, R.O.C; 3) Department of Clinical Laboratory Sciences and Medical Biotechnology, National Taiwan University College of Medicine, 1 Chang-Te Street, Taipei, 100, Taiwan; 4) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University; 5) Graduate Institute of Brain and Mind, College of Medicine, National Taiwan University Hospital, National Taiwan University; 6) Genetic Epidemiology Core, Center of Genomic Medicine, National Taiwan University.

Background: Genome-wide association study (GWAS) has limitations for exploring rare genetic variants. With the advent of next-generation sequencing (NGS), it becomes feasible to search the whole genome for rare genetic variants specific to schizophrenia patients. A previous study among a large number of families of sib-pair co-affected with schizophrenia in Taiwan has revealed several linkage signals by incorporating several endophenotypes. This implies existence of genetic heterogeneity in schizophrenia. We hence postulated that high density nuclear families of schizophrenia may help identify certain inherited rare variants that can be used for replication in other multiplex families of schizophrenia. **Method:** We selected one high density schizophrenia family from Taiwan Schizophrenia Linkage Study (TSLS), which recruited schizophrenia patients and their first-degree relatives throughout Taiwan from 1998 to 2002. NGS was performed to sequence the whole genomes of a 5-member nuclear family, in which the mother and 2 children affected with schizophrenia, and the father and another child unaffected. Both dominant and recessive inheritance models were used to explore schizophrenia related genomic variations. **Results:** The results showed that total 384 variants (2 exonic single nucleotide variants; SNVs) selected under recessive inheritance model, i.e., with schizophrenia patients having homozygous variants and health persons having heterozygous variants. In dominant inheritance model, there were 13715 variants (70 exonic SNVs) selected, i.e., with schizophrenia patients having the variant and health persons having wild-type allele only. In addition, 10 non-synonymous exonic SNVs that exhibited inheritance in the family but not seen in 1000 genomes project were identified for those in accordance with dominant inheritance model. **Discussion:** Our findings demonstrated the utility of NGS in identifying inherited rare genetic variants that are potentially associated with multiplex schizophrenia. Our use of a multiplex family can help exclude those variants due to typing error, or de novo mutations, and hence select those with high-penetrating susceptibility genetic variants for schizophrenia. In particular, the 10 non-synonymous exonic SNVs warrant future replication in other multiplex families from TSLS. This may lead to discovery if rare but inheritable susceptibility variants for schizophrenia and their underlying pathophysiology.

1154F

Contribution of rare variants in the development of gout in Japanese males. *A. Taniguchi, C. Sekita, H. Kaneko, W. Urano, N. Ichikawa, H. Yamanaka.* Institute of Rheumatology, Tokyo Women's Medical Univ, Shinjuku-ku, Tokyo, Japan.

Background: Genome-wide association studies (GWAS) have identified multiple loci associated with serum levels of uric acid (SUA) and gout. Especially, *SLC22A12, SLC2A9, ABCG2, SLC17A1*, and *SLC17A3* have been reported to be associated with SUA or gout in multiple ethnicities including Japanese. However, these polymorphisms explain about 10% of variation. Contribution of rare variants in the development of gout has not extensively been studied especially in Japanese. Objective: The aim of the present study was to identify low frequency variants associated with gout in Japanese males. Methods: Genomic DNA was extracted from 325 Japanese male patients with gout and 597 Japanese male controls. Rare variants were defined as minor allele frequencies less than 0.01 in controls. Resequencing of *SLC22A12, SLC2A9, ABCG2, SLC17A1*, and *SLC17A3* was preformed using genomic DNA samples from a hundred of gout and a hundred of controls. Every exon and exon/intron boundaries were amplified by PCR and the PCR products were sequenced using the Big Dye Terminator cycle sequencing kit and the ABI PRISM 3100xl Genetic Analyzer (Life Technologies). Based on the results of resequencing, genotyping of each variant was performed by TaqMan methods (Life Technologies) using all samples. Results: We identified seven rare variants in *SLC17A1, SLC17A3*, and *ABCG2* that are responsible for uric acid secretion in the renal proximal tubules. Of those, two were nonsense and five were missense mutations. A significant accumulation of rare variants was identified in gout patients compared to controls ($p=2.461\times10^{-6}$). Six rare variants were identified in *SLC2A9* and *SLC22A12* that are responsible for uric acid reabsorption in the renal proximal tubules. One mutation was nonsense and the rest of the mutations were missense. We found a significant accumulation of those variants in controls (p=0.037). Conclusion: It is suggested that rare variants in *SLC17A1, SLC17A3*, and ABCG2 promote gout and those in *SLC2A9* and *SLC22A12* are protective for gout. The present study indicated that two types of rare variants that were protective for gout or promote it were associated in the development of gout in Japanese males.

1155W

Targeted sequencing of the pericentromeric region of chromosome 2 in Finnish constitutional delay of growth and puberty families. *D. Cousminer[1], L. Dunkel[2], A. Palotie[1,3], E. Widén[1].* 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 2) Department of Pediatrics, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 3) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

Constitutional delay of growth and puberty (CDGP) is the most common cause of pubertal delay, representing about 65% of boys and 30% of girls with late puberty. CDGP is defined as the absence of pubertal development at an age 2 standard deviations above the population mean, an extreme variant of normal pubertal timing. While both environmental and genetic factors contribute to variability of pubertal timing and tempo, high heritability estimates of up to 80% predict a strong influence from genetic factors. At the tail ends of the population distribution, one or several genes with large effect sizes are predicted to have a strong influence on the phenotype. Recent studies showed that the majority of Finnish families segregated CDGP in an autosomal dominant manner, and a subsequent linkage study identified a susceptibility locus in the pericentromeric region of chromosome 2. We followed up on this linkage signal by performing targeted sequencing of over 60 Mb under the linkage peak in the proband and affected parent of the 13 best-linked families. Poorly sequenced regions were imputed against Finnish 1000 Genomes reference samples. Following sequencing and imputation, transmitted coding variants and non-coding regulatory variants will be compared among the probands to create a list of potentially causative candidate variants. These candidate variants will be assessed for association with pubertal delay by genotyping independent cases and controls. Preliminary results show that the 13 probands are not more closely related to one another than expected, nor is there a high probability of consanguinity among their recent ancestors. Furthermore, principal component and multidimensional scaling analyses indicate that the probands fit against a background of population controls and do not cluster more closely together than the population in general. Prior to imputation, there were no shared transmitted variants or indels among all 13 probands at a known 1000 Genomes allele frequency of <10%. There were also no shared 3'UTR, noncoding transcript, or nonsynonymous transmitted variants with unknown 1000 Genomes frequency that did not appear in reference genomes in the ensemble genome browser. Predicted regulatory regions remain to be analyzed, and it is possible that only a subset of probands share a causative mutation. We hope the results of this study may elucidate part of the as-yet poorly known genetic landscape underlying delayed puberty.

1156T

Population genetics of rare variants and complex diseases. *R. Hernandez[1], M. Maher[1], L. Uricchio[1], D. Torgerson[2].* 1) Bioeng. & Therapeutic Sci, UCSF, San Francisco, CA; 2) Department of Medicine, UCSF, San Francisco, CA.

Identifying drivers of complex traits from the noisy signals of genetic variation obtained from high throughput genome sequencing technologies is a central challenge faced by human geneticists today. We hypothesize that the variants involved in complex diseases are likely to exhibit non-neutral evolutionary signatures. Uncovering the evolutionary history of all variants is therefore of intrinsic interest for complex disease research. However, doing so necessitates the simultaneous elucidation of the targets of natural selection and population-specific demographic history. Here we characterize the action of natural selection operating across complex disease categories, and use population genetic simulations to evaluate the expected patterns of genetic variation in large samples (n=10,000). We focus on populations that have experienced historical bottlenecks followed by explosive growth (consistent with many human populations), and describe the differences between evolutionarily deleterious mutations and those that are neutral. We find that the genes associated with several complex disease categories exhibit stronger signatures of purifying selection than non-disease genes. In addition, loci identified through genome-wide association studies of complex traits also exhibit signatures consistent with being in regions recurrently targeted by purifying selection. Through simulations, we show that population bottlenecks and rapid growth enables deleterious rare variants to persist at low frequencies just as long as neutral variants, but low frequency and common variants tend to be much younger than neutral variants. This has resulted in a large proportion of modern-day rare alleles that have a deleterious effect on function, and that potentially contribute to disease susceptibility. The key question for sequencing-based association studies of complex traits is how to distinguish between deleterious and benign genetic variation. We used population genetic simulations to uncover patterns of genetic variation that distinguish these two categories, especially derived allele age, thereby providing inroads into novel methods for characterizing rare genetic variation driving complex diseases.

1157F

Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders.
H.N. Cukier[1], N.D. Dueker[1], S.H. Slifer[1], J.M. Lee[1], P.L. Whitehead[1], E. Lalanne[1], N. Leyva[1], I. Konidari[1], R.C. Gentry[1], W.F. Hulme[1], D. Van Booven[1], V. Mayo[1], N.K. Hofmann[1], M.A. Schmidt[1,2], E.R. Martin[1,2], J.L. Haines[3], M.L. Cuccaro[1,2], J.R. Gilbert[1,2], M.A. Pericak-Vance[1,2]. 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA.

Autism spectrum disorders (ASDs) encompass a constellation of neurodevelopmental conditions and studies to date demonstrate that the underlying etiology is extremely heterogeneous. To help unravel this genetic complexity, we performed whole exome sequencing in 40 ASD families with multiple, distantly related, ASD affected individuals. In contrast to previous autism exome studies which have primarily focused either on simplex families to discover de novo alterations or consanguineous families that carry recessive mutations, our study required that each family contain at minimum one pair of affected cousins. A total of 164 individuals were captured with the Agilent SureSelect Human All Exon kit, sequenced on the Illumina HiSeq 2000, and the resulting data processed and annotated with BWA, GATK, and SeattleSeq. Each family had approximately 90,000 changes. Variants were filtered to those in identity by descent (IBD) regions delineated by SNP genotyping data. Initial analyses focused on novel and rare (MAF < 0.05) variants predicted to be detrimental, either by altering amino acids or splicing patterns. In accordance with a dominant model of inheritance, exome sequencing identified 742 heterozygous changes, 499 of which were validated either by Sanger sequencing or genotyping on the Infinium HumanExome BeadChip. We identified numerous potentially damaging, ASD associated risk variants in genes previously unrelated to autism. A subset of these genes has been implicated in other neurological disorders including depression (*SLIT3*), epilepsy (*CLCN2*, *PRICKLE1*), mental retardation (*AP4M1*, *CEP290*), schizophrenia (*WDR60*), and Tourette syndrome (*OFCC1*). This reinforces the theory that there are shared genetic components across distinct neurological disorders. Additional alterations were found in previously reported autism candidate genes, including three genes with alterations in multiple families (*CSMD1*, *FAT1*, and *STXBP5*). Compiling a list of ASD candidate genes from the literature, we determined that variants occurred in ASD candidate genes 1.65 times more frequently than in random genes captured by exome sequencing ($p=8.55\times10^{-5}$). By studying these unique pedigrees, we have identified novel DNA variations related to ASDs, demonstrated that exome sequencing in extended families is a powerful tool for ASD candidate gene discovery, and provided further evidence of an underlying genetic component to a wide range of neurodevelopmental and neuropsychiatric diseases.

1158W

Nextgen RNA sequencing of monocytes coupled with association data identifies several genes in Systemic Lupus Erytrhematosus susceptibility.
AK. Maiti[1], C. Sun[1], P. Motghare[1], M. Arango[2], J-M. Anaya[2], SK. Nath[1]. 1) Gen Epidemiology Unit, A & CI, OMRF, Oklahoma City, OK; 2) Universidad del Rosario-Corporación para Investigaciones Biológicas, Bogotá, Colombia.

Systemic Lupus Erythematosus (SLE) is a devastating autoimmune disease. While there is no cure, disease progression may be delayed by proper treatment early in the disease process. Thus, identification of distinct target for drug development could have immense impact in developing disease specific therapy. Along with lymphocytes and other blood cells, monocytes play a critical role in SLE pathogenesis. Although several large scale genomic studies such as microarray gene expression have reported results for peripheral blood mononuclear cells (PBMC) or lymphocytes, no studies have been performed using only monocytes from SLE patients. Since the ratio of the monocytes to lymphocytes is very low in PBMC, major changes in gene expression observed in microarray studies could come from only lymphocytes, thus monocyte biology in SLE has not been studied in detail. To understand the monocyte biology in SLE, we performed Nextgen RNA sequencing from monocytes of SLE patients and healthy controls and identified several differentially expressed genes (e.g. IL9R, LTBP3, DTX3, MAVS, FGF4 and ARHGEF40 that are not studied extensively for their role in the pathogenesis of SLE. We also identified several novel structural variations, such as deletions, insertions and splice variations in several genes and nongenic regions that could be potentially associated with SLE. When these novel variations are correlated (LD>80%) with known SLE associated variants from our SLE GWAS data we observed that nearest SNP of some of these novel variants carrying genes/regions are strongly or suggestively associated with SLE, such as, HIP1 ($p=6.7 \times 10$ (-16), DTX3 ($P= 3 \times 10$(-3)), LTBP3 (6.78×10 (-6), PTPN3 (9×10 (-4). Among genes, some of them (e.g DTX3, LTBP3) are both identified as differentially expressed genes in monocytes and suggestively associated with SLE. Also, several similar variations are identified in multiple cases although no information could be obtained from previous studies since SNPs from these regions are not present in GWAS CHIPs but could be potentially associated with lupus. Association studies with these newly identified variations with a large number of samples are in progress to identify novel SLE associated variants. Taken together, RNA sequencing of monocytes not only identified differentially expressed genes, but also identified novel variants that could be pathogenic to SLE and implicate their roles in genotype-phenotype correlation.

1159T

Multiplexed targeted capture of *FBN1* and *FBN2* reveals association with adolescent idiopathic scoliosis.
J.G. Buchan[1], D.M. Alvarado[3], P. Yang[3], C. Cruchaga[4], M.B. Harms[2], K. Ha[3], T. Zhang[1], M.C. Willing[5], D.K. Grange[5], T.E. Druley[1,5], A.C. Braverman[6], M. Lovett[1], M.B. Dobbs[3,7], C.A. Gurnett[2,3,5]. 1) Genetics, Washington University, Saint Louis, MO; 2) Neurology, Washington University, Saint Louis, MO; 3) Orthopedic Surgery, Washington University, Saint Louis, MO; 4) Psychiary, Washington University, Saint Louis, MO; 5) Pediatrics, Washington University, Saint Louis, MO; 6) Medicine, Washington University, Saint Louis, MO; 7) St. Louis Shriners Hospital for Children.

Adolescent idiopathic scoliosis (AIS) is a common pediatric spine deformity that affects up to 3% of the population. Despite a strong genetic component, few genes have been associated with AIS and the pathogenesis remains poorly understood. Marfan syndrome and congenital contractual arachnodactyly are Mendelian disorders that are highly associated with scoliosis and are caused by mutations in the large fibrillin genes, *FBN1* and *FBN2*. To determine whether rare genetic variants in fibrillin contribute to the pathogenesis of AIS, we sequenced *FBN1* and *FBN2* using a novel and cost-effective method of targeted capture called multiplex direct genomic selection (MDiGS) in combination with exome sequencing. In individuals of European ancestry, an increased frequency of private *FBN1* and *FBN2* variants was identified in AIS cases (8.2%; n=233) compared to controls (2.0%; n= 393) (OR=4.27; $p=2.7\times10^{-4}$). Private *FBN1* and *FBN2* variants were associated with more severe spinal deformity ($p=3.2\times10^{-4}$), but were not associated with systemic features of Marfan syndrome or congenital contractural arachnodactyly. Activation of the TGF-β signaling pathway was demonstrated by elevated pSMAD2 in muscle of AIS patients with private *FBN1* variants. Our findings underscore the functional importance of rare fibrillin variants in AIS pathogenesis and provide the possibility of novel therapeutic strategies for treating AIS.

1160F

ImmunoSeq: Discovery of novel rare variants implicated in autoimmune and inflammatory diseases by targeting regulatory regions in immune cells. *A. Morin[1,2], T. Kwan[1,2], K. Tandre[3], M.L Eloranta[3], V. Arsenault[2], M. Caron[1,2], L. Létourneau[2], C. Wang[5], G. Bourque[1,2], C. Laprise[4], A. Montpetit[2], A.C. Syvänen[5], L. Rönnblom[3], M. Lathrop[1,2], T. Pastinen[1,2].* 1) Department of Human Genetics, McGill University, Montréal, Quebec, Canada; 2) McGill University and Genome Québec Innovation Centre, Montréal, Quebec, Canada; 3) Department of Medical Sciences, Section of Rheumatology, Uppsala University, Uppsala, Sweden; 4) Département des sciences fondamentales, Université du Québec à Chicoutimi Saguenay, Quebec, Canada; 5) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

Genome-wide association studies (GWAS) have identified many common SNPs associated to complex traits, yet they only partially explain their genetic component. Whole-exome sequencing has been used in the discovery of rare variants associated to complex traits, but its potential success is limited, since GWAS have shown that the associated SNPs are predominantly located in noncoding regions. A recent study showed that rare coding variants have limited impact on the development of autoimmune diseases. Comprehensive DNase I hypersensitive site (DHS) mapping by the ENCODE project identified all classes of cis-regulatory elements, and recent studies showed enrichment of GWAS noncoding variants in DHS. The discovery of rare variants located in regulatory regions specific to immune cells should explain part of the heritability of autoimmune complex traits. Our approach combines selective DNA capture of relevant regulatory regions coupled to next-generation sequencing. Genome-wide DHS mapping data from the ENCODE and NIH Roadmap Epigenomics Projects were used to select regulatory regions of immune cells. We designed a custom DNA capture panel (Roche SeqCap EZ developer) to target these selected regions, as well as exonic and HLA regions. Captured DNA (163Mb) is indexed (5-fold per lane) for sequencing on Illumina HiSeq2000 system, yielding average coverage of ~28x. We first applied the 'Immunoseq.' assay to 30 healthy individuals from Swedish (Uppsala BioResource), where we have existing transcriptome, methylome and ChIP-seq data from three primary immune-cell subsets (monocytes, B-cells and T-cells). The capture panel was also applied to 40 trios from the SaguenayLac-Saint-Jean asthma familial collection. Rare variants were defined as novel if absent in genotyping data, dbSNP137, or from 1000 Genomes imputation data. We observe a large proportion (15080/21626=70%) of rare variants falling into regulatory regions. A subset of these rare variants (8%, n=1747) are highly conserved (GERP ≥ 4), and of these, 69 are proximal to 52 genes identified by autoimmune disease GWAS. This highlights the 1) potential impact of non-coding variation in the development of autoimmune and chronic inflammatory diseases, and 2) power of our targeted capture approach to identify novel rare variants in regulatory regions. On-going work focuses on correlating these rare, non-coding variants to gene expression and regulatory element activity as well as risk of inflammatory disease.

1161W

Novel candidate genes putatively involved in stress fracture predisposition detected by whole Exome sequencing. *E. Friedman[1,2], D.S. Moran[3], D. Ben- Avraham[4], R. Yanovich[3], G. Atzmon[4].* 1) Oncogenetics Unit, Inst Gen, Chaim Sheba Med Ctr, Tel Hashomer, Israel; 2) The Sackler School of Medicine, Tel-Aviv University, Ramat Aviv; 3) The Military Physiology Unit, Heller Institute, Sheba Medical center, Israel; 4) Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10463.

While clearly genetic factors are involved in stress fracture (SF) pathogenesis, few studies that focused on candidate genes reported on the contribution of sequence variants in these genes to this common overuse injury. We employed an unbiased screening approach by using exome sequence capture arrays followed by next generation sequencing (HiSeq2000) of two pooled DNA samples from Israeli soldiers: cases with high grade SF (n= 34) and ethnically and age matched controls with no evidence of SF (n= 60). Of the 144,217 and 202,406 sequence variants (control and case pools, respectively) 67,408 variants passed the various QC filtering stages with 3-20,000 reads/variant. Of 1900 variants with more than 600 reads/variant in both DNA pools, 145 sequence variants (in 127 genes) displayed statistically significant ($p<0.05$) differences in rates between cases and controls. Subsequent validation of these 145 sequence variants individually for the 55 controls and 32 SF soldiers who formed the pooled samples using the Sequenom platform, validated 11/145 SNPs. A second, independent, individually genotyped validation cohort with 72 controls and 104 SF soldiers using the Sequenom platform to query the same 145 SNPs, validated only 1 SNP: a missense mutation in the BTN3A1 gene. Combined analysis the two datasets for all cases and controls with adjustment for inter-batch variability resulted in 8/145 SNPs that were statistically different between cases and controls, 5 of which were in the originally detected and validated 11 SNPs. Of these, three SNPs were Synonymous SNPs located within the SEC24D, PAWR, LRRC55 genes. SEC24D and PAWR encode for proteins involved in carbohydrate and lipid post translational modifications and Ubiquitin C network. The LRRC55 protein acts in the calcium-activated potassium channel. In conclusion, exome sequencing of DNA pools provided novel candidate genes seemingly involved in SF pathogenesis and predisposition.

1162T

A possible role of Transposable Elements in dysregulating the genomic architecture of Schizophrenia. *F. Macciardi[1], G. Guffanti[2], S.G. Potkin[1], M. Pato[3], I. Guella[1], M. Vawter[1], J.A. Knowles[3], T. VanErp[1], C. Pato[3], S. Gaudi[4].* 1) Dept. of Psychiatry & Human Behavior, University of California Irvine (UCI), Irvine, CA; 2) Department of Psychiatry Division of Epidemiology & Division of Child and Adolescent Psychiatry Columbia University/ NYSPI, New York, NY; 3) Department of Psychiatry and the Behavioral Sciences - Keck School of Medicine at USC, Los Angeles, CA; 4) Department of Infectious, Parasitic and Immune-Mediated Diseases, Italian National Institute of Health, Rome, Italy.

Background. Between half and two-third of our genome is composed of repetitive, low-complexity Transposable Elements (TEs, e.g., LINEs, SINEs, SVAs and HERVs). Since their discovery TEs were thought to act as 'controlling elements' of nearby genes. While the notion that DNA could be mobile was accepted, the idea of control was not. The role of TEs as regulatory elements is now more recognized. TEs create variability by retro-transpositions and by SNVs in fixed TEs. By retrotransposing to new insertional sites, TEs create structural variants and provide novel promoters, splice sites, exons or polyadenylation signals. **Methods.** We identified and annotated SNVs and Retrotrasposition Insertion Polymorphisms (RIPs) for various classes of TEs (LINEs, SINE, HERVs and SVAs) from whole-genome sequence data in post-mortem brains of 10 cases and 10 controls. We examined TE sequence differences (including RIPs) between SZ patients and controls, using re-alignment and de novo assembly and characterized the genomic context of RIPs (exon, introns, 5' and 3' UTRs, non-coding regions). **Results**. We found that RIPs are largely underrepresented in SZ patients compared to controls for all TE classes. The number of RIPs in controls ranges from 3,000 to at least 4,500, within the boundaries calculated from the 1,000 Genome Project for dbRIP-reference and RIP non-reference TEs, a larger estimate than from previous studies. In cases RIPs are 1/10 to 1/20 than in controls. **Discussion.** The low number of RIPs in SZ subjects suggests a neuro-developmental defect affecting embryonic brain development. Recent stem cell models show that a high number of RIPs for L1s and Alus is a key factor for somatic neuronal development, although the precise mechanisms are not yet clear. It is thus possible that a reduction of RIPs in SZ at the early embryonic level may be less relevant than the somatic retrotransposition that occurs in neurons at later developmental stages and results in an increased transcriptome heterogeneity among neurons. Our preliminary results reveal a new, unexpected dimension of the regulatory genome that may play a pivotal role in the etiology of schizophrenia. Since their identification, TEs represent a controversial concept due to their biology and mechanisms of action that challenge genetic dogma. Our findings, however, suggest that TEs can represent an important new genomic risk factor in schizophrenia.

1163F

AMD risk and association with variation in 202 drug target genes. *P.L. St Jean[1], W.H. Cade[2], F. Grassmann[3], M. Schu[4], S. Slifer[2], Z. Ye[5], M.H. Brilliant[5], M.M. DeAngelis[6], L.A. Farrer[4], J.L. Haines[7], T.E. Kitchner[5], M.A. Pericak-Vance[2], B.H.F. Weber[3], L. McCarthy[1], C.-F. Xu[1], M.G. Ehm[1].* 1) Quantitative Sciences, GlaxoSmithKline, RTP NC, USA; Stevenage UK; 2) University of Miami Miller School of Medicine, Miami, FL, USA; 3) Institute of Human Genetics, University of Regensburg, Regensburg, Germany; 4) Boston University Schools of Medicine and Public Health, Boston, MA, USA; 5) Marshfield Clinic Research Foundation, Marshfield, WI, USA; 6) Department of Ophthalmology and Visual Sciences , Moran Eye Center, University of Utah, Salt Lake City, UT USA; 7) Center for Human Genetics Research, Vanderbilt University Medical School, Nashville, TN, USA.

Age-related macular degeneration (AMD) is a common cause of blindness and visual impairment in older adults occurring in ~12% of people over age 80. Advanced forms of AMD are geographic atrophy (GA) involving retinal pigment epithelial atrophy, and neovascular (NV) complications characterized by abnormal blood vessel growth and leakage of blood and protein. Approved treatments of AMD, focused on the neovascular forms, include ranibizumab and aflibercept which are both administered by injection. Developing drugs effectively treating GA, as well as non-injectable treatments for neovascular forms, would have a substantial, positive impact for people with advanced AMD. Our objective is to identify genetic variants in drug targets gene that are associated with AMD risk with the hope that some of these genes may prove to be effective targets for developing therapeutics for AMD. Exons and flanking regions of 202 genes encoding current or prospective GSK drug targets were sequenced in 14,000 European subjects. Illumina 500K or Affymetrix 500K data were available on 6900 of these subjects and were used to create reference haplotype sets allowing us to impute into several AMD collections of European origin. Association between AMD risk and dosage was conducted on a total of 2,696 cases and 3,335 controls for variants with an R2 quality metric threshold of $\geq$0.5. A meta-analysis of effect estimates was performed using the inverse variance method in METAL while testing for heterogeneity across studies. A predefined statistical threshold of $p \leq 3.65E-5$, accounting for the effective number of variants, was used to identify significant results while a less stringent threshold of $p<0.005$ was used to identify suggestive results. Across the collections, 46,957 variants with an R2 quality$\geq$0.5 were analyzed. Thirteen variants were associated at $p \leq 3.65E-5$ and these reside in regions flanking, but not within the 202 target genes. Ten of these variants are on chromosome 6 and map to reported AMD risk genes C2, CFB, SKIV2L; however, they are 400-800 kb from the closest GSK target gene. Several variants within GSK target genes were associated at $p<0.005$ and many of these genes are implicated in AMD disease mechanisms such as inflammation and lipo-protein processing. Further support of the involvement of these target genes in AMD risk will be assessed in independent replication studies.

1164W

Matrix metalloproteinase 2, 3, 9, 10, 13 gene polymorphisms and risk for polycythemia and essential thrombocytosis patients. *E. Uctepe[1], S. Maral[2], O. Bender[1], T. Yasar[1], E. Gunduz[1], M. Gunduz[1,3].* 1) Medical Genetics Department, Turgut Ozal University, Ankara, Turkey; 2) Internal Medicine Depertmant,Fatih University, Ankara, Turkey; 3) Departments of Otolaryngology Head and Neck Surgery,Turgut Ozal University, Ankara, Turkey.

Chronic myeloid disorders such as polycythemia vera (PV), essential thrombocytosis (ET) and idiopathic myelofibrosis arises from clonal proliferation of neoplastic stem cells in the bone marrow. Matrix metalloproteinases (MMPs) are a family of zinc-dependent endopeptidases that have potential to degrade all types of extracellular matrix (ECM) and also play a role in remodeling of the ECM. It is known that MMPs play a role in bone marrow fibrosis. The primary goal of our study is to determine the relationship between chronic myeloproliferative diseases and MMP gene polymorphisms. The demonstration of a relationship will determine whether polymorphisms lead to susceptibility to disease and with future work may aid in the development of new therapeutic modalities. Patients were selected from outpatient clinics of the Turgut Ozal Medical University Hospital between December 2010 and May 2011. Twenty-eight patients which previously diagnosed and follow-up with a diagnosis of polycythemia vera, seventeen with secondary polycythemia, and twelve with essential thrombocytosis were enrolled in the study, along with a control group of 22 healthy patients. DNA isolation from peripheral blood, DNA amplification with polymerase chain reaction (PCR) and agarose gel electrophoresis was performed. Using PCR-RFLP, MMP-2, -3, -9, -10, and MMP-13 gene polymorphisms were analyzed. There was a statistically significant difference between the groups in terms of Gln279Arg polymorphisms rates of MMP 9 . ($\psi2 = 22.975$, $p = 0.001$). The highest MMP 9 Gln279Arg polymorphism rate was observed in the ET group. It were determined that MMP 13 gene 77A> G polymorphism had significantly higher prevalence rates in the ET group. ($\psi2 = 12.605$, $p = 0.05$). There was nostatistically significant difference between the groups in terms of MMP-2 -735 C> T, MMP -3-1612 5A/6A, and MMP-10 180G> A polymorphism rates. (respectively: $\psi2 = 6.447$, $p = 0.375$, $\psi2 = 10.870$, $p = 0.092$). In conclusion, Gln279Arg MMP-9 and 77A> G MMP-13 gene polymorphisms were associated with ET and PSV diseases. Hence, it is thought that these gene polymorphisms may play a role in the mechanism of bone marrow fibrosis and may be a factor that increases the risk of thrombosis. Illumination of the molecular basis of the relationship between MMP-thrombosis and MMP-fibrosis provides a better understanding of the pathophysiology of PSV and ET diseases and will allow new approaches to diagnosis and treatment.

1165T

Race and sex effects on the relationship between sarcopenia and BMD.
H. He[1], Y.J. Liu[1], J. Li[1], H. Shen[1], Q. Tian[1], H.W. Deng[1,2]. 1) Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, New Orleans, LA; 2) Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China.

The relationship between reduced muscle mass (sarcopenia) and bone mineral density remained unclear. The main purpose of this study was to determine this relationship and examine effects of fat mass (FM), lean mass (LM), and muscle strength on regional and whole body BMD in different race cohorts and to determine if these relationships are altered by sex and/ or race. The study population was collected from different clinical centers and consisted of 17,891 individuals from three ethnic populations, including 8604 Caucasians, 5013 Chinese and 4274 African-Americans. Partial correlation analyses, controlling for race, study site, gender, regular exercise, smoke and alcohol use were conducted to investigate the relationships between BMD and body composition variables. Multiple regression analysis was used to examine the independent effects of FM, LM and grip strength on regional and whole body BMD. Regression models were stratified by race and sex and adjusted for age, height, city, menopause status, current smoking, alcohol use and regular exercise. Men had a greater lean mass, lower fat mass and lower fat percentage than women. Whole body and regional BMDs were significantly greater in men than in women. African-Americans had the highest mean bone density, follows by Caucasians and then Chinese. As expect, grip strength was higher in men than in women and highest in Caucasians. Age was negatively associated with height, grip strength, lean mass, whole body and regional BMDs, and positively associated with fat mass and fat percentage. LM and FM were positively associated with BMD at all sites ($r = 0.14$-0.37, $p < 0.001$). The partial correlation was stronger for LM and BMDs than for FM and BMDs. LM was a significant independent (Std β= 0.11- 0.34, $p < 0.001$) contributor to BMDs across race and sex and had greater effects on BMDs than FM. FM was also a significant ($p < 0.001$) determinant of BMD, except in African-American and Caucasian men. Compared to LM, Grip strength contributed less effect to whole body BMD and wrist ultradistal radius BMD. Non-significant race×FM and race×LM interaction terms were found. Sarcopenia was associated with a 3-fold higher risk of osteoporosis(OR=3.02; 95%CI= 2.622, 3.473). LM is the strongest predictor of BMD for all race and sex group. Sarcopenia is associated with low BMD and osteoporosis. Further studies may assess whether maintaining lean mass contributes to prevent osteoporosis.

1166F

Unexpected pleiotropy: Do asthma and dental caries share a genetic basis? *J.R. Shaffer[1], R.J. Weyant[2], R. Crout[3], D.W. McNeil[4], M.L. Marazita[1,5,6]*. 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Dental Public Health and Information Management, University of Pittsburgh, Pittsburgh, PA; 3) Periodontics, West Virginia University, Morgantown, WV; 4) Dental Practice and Rural Health, West Virginia University, Morgantown, WV; 5) Oral Biology, University of Pittsburgh, Pittsburgh, PA; 6) Clinical and Translational Science, University of Pittsburgh, Pittsburgh, PA.

The surprising association between asthma and dental caries (i.e., tooth decay) has been repeatedly observed in epidemiological studies. However, at present, the cause of this relationship is largely speculative. As part of an initiative by the Center for Oral Health Research in Appalachia (COHRA), we explored the relationship between dental caries scores (assessed by intra-oral examination) and parent-reported asthma in a cohort of 940 rural children <10 years of age. These children comprise part of a larger sample of 2- and 3-generation families, all of which were genotyped on an Illumina whole-genome SNPchip. Approximately 14% of children reported asthma, which was significantly associated with both whether or not a child experienced dental caries ($p=0.04$; ψ^2 test) and the number of carious (i.e., decayed or restored) tooth surfaces ($p=0.001$; Wilcoxon test). Using a variance components approach, we determined that both asthma and dental caries showed high heritability (90% for asthma, 40-60% for dental caries; both consistent with other studies). Moreover, estimates of genetic correlation (ρ_G) indicated that asthma and dental caries may partly share a common genetic basis ($\rho_G=0.36$ for asthma and number of carious surfaces, $p = 0.006$; $\rho_G=0.35$ for asthma and yes/no caries, $p=0.03$). In other words, the suites of genes affecting susceptibility to these two complex diseases may partly overlap. Next, we explored whether the ORMDL3-GSDMB locus on chromosome 17, which has been repeatedly implicated in childhood asthma in large-scale association studies, was associated with dental caries. The SNP rs8082130 in this region was associated with number of carious tooth surfaces ($p=0.0004$). Other loci implicated in asthma (e.g., HLA-DQA1/A2, RORA, SMAD3, and IL2RB) showed nominal associations with dental caries. These results support the hypothesis that asthma and dental caries are linked, and suggest that a shared genetic basis may partly explain this relationship. However, the exact nature of this relationship remains speculative: do asthma and dental caries share a root etiology, or are asthma-liability genes influencing dental caries through a mechanistic artifact, such as exposure to antiasthma medications? While additional work is needed to tease out the causality of this association, this study illustrates that pleiotropic actions of the genetic determinants of complex diseases may yet to be fully appreciated. R01-DE014899; U01-DE018903.

1167W

A pilot T1D risk prediction study using custom panel and advanced multivariate predictive models. *C. Kim[1], Z. Wei[1], J. Glessner[1], K. Thomas[1], H.K. Åkerblom[2], M. Knip[2], H. Hakonarson[1,3,4], I. Ilonen[4,5].* 1) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) The Children's Hospital University of Helsinki, Finland; 3) The Department of Pediatrics, The Perelman School of Medicine, Philadelphia, PA; 4) Turku Immunogenetics Laboratory, University of Turku, Turku, Finland; 5) equal contribution.

Recent progress in genetic research has expanded our knowledge of the genes affecting risk of Type 1 Diabetes (T1D). Up to 50 gene loci outside the HLA region have been identified at the moment. However, the effect of each of these genes is relatively small and their combined effect has been estimated to be only around 10-15 percent. Given the relatively modest individual contributions of T1D loci, towards genetic risk, the simple univariate methods for prediction of T1D are not optimal. We have designed and validated a better multivariate non-linear genetic test that can be used to determine which infants are most at risk of T1D in the future. By scoring approximately 400 to 500 of the most strongly associated SNPs ($P \leq 10\text{-}5$) from GWAS with a support vector machine (SVM), we can achieve a negative predictive value (NPV) of virtually 100 percent in siblings of affected patients. These 'SVM SNPs' comprise approximately 300 informative markers that map to the major histocompatibility complex (MHC) with the remaining being distributed throughout the genome. Given these promising performance, we hypothesize that T1D can essentially be ruled out in children with a strong family history. As a proof-of-concept, here we conduct a risk prediction study for children enrolled in the full TRIGR study and early pilots. We designed a custom chip that covers 384 most relevant T1D SNPs. We genotyped 198 Finnish samples with unknown disease status using this chip for prediction. To evaluate and optimize model performance, we also genotyped 185 samples (132 controls and 53 cases) collected at the Children's Hospital of Philadelphia (CHOP). The WTCCC T1D dataset (2938 controls and 1963 T1D cases) was used to train predictive models. We employ SVM and L1 penalized logistic regression (LR) for risk prediction. The SVM model achieved AUC of 0.796 (95 percent CI=[0.715, 0.868]), and the LR model, AUC of 0.797 (95 percent CI=[0.717, 0.867]), when testing on the CHOP samples. The predicted risk status for the Finnish samples demonstrates significant correlation with autoantibodies, suggesting that in addition to HLA typing and family history, the T1D disease status could be predicted with greater accuracy when coupled with the predictive score from the 384 markers than by the HLA status and family history measures alone. Whole exome sequencing is underway to further improve the predictive accuracy of the algorithm and to better understand the pathogenesis of the disease.

1168T

Prediction of complex phenotypes, such as skin, eyes and hair color based on the analysis of polymorphisms of pigmentation genes for forensic purposes. *F.T. Goncalves, F.A. Lima, R.S. Gonzales, C. Fridman.* Legal Medicine, Ethics and Occupational Health, Medical School, University of São Paulo, Sao Paulo, Sao Paulo, Brazil.

Human pigmentation traits, including color variation in skin, eye, and hair, belong to the most visible and differentiating human traits. The genetic basis underlying variation in human pigmentation traits has been the subject of intensive research by investigators in a variety of life science communities, including forensic purposes, in which the use of these trials can help identifying missing person or guide some police investigations. The variations in skin pigmentation are dependent on geographic location and population ethnicity. The two types of melanin synthesized in well defined chemical reactions are the protective dark colored eumelanin and the sulphur containing light red-yellow pheomelanin. The events leading to the synthesis of melanin are controlled by signaling cascades that involve a host of genes encoding ligands, receptors, transcription factors, channel transporters and many other crucial molecules. The aim of this study was evaluate polymorphisms in MC1R, TYR, ASIP, SLC24A5 and SLC45A2 genes in a sample of 132 individuals of admixed population from Brazil, intending to use the data in forensic genetics casework in several situations. No significant differences were detect in all polymorphisms and eye, hair and skin color association, except the SNP rs1426654 of the SLC24A5 gene. The polymorphic homozygous genotype (GG) for gene SLC24A5 was associated with dark skin (OR, 0.07; CI, 0.009-0.49) compared with light skin, while the heterozygous genotype (AG) showed association with black hair (OR 0.22, CI 0.07-0.66 p = 0.008) when compared with brown hair. The SLC24A5 gene encodes the protein NCKX5, which involves a cellular transport complex of ions sodium/calcium. It is postulated that this gene participates in the control of calcium in melanosomes. As the production of melanin is dependent on the amount of calcium in these cells, we infer that the polymorphism in the SLC24A5 gene would result in a transmembrane protein with increased permeability to calcium in the cell, which would lead to a higher concentration of this ion in melanosome, that increases the production of melanin resulting in dark skin and hair color, as described in homogenous population of Africa and Europe. We are increasing our sample to confirm these results.
Supported: FAPESP, LIM40-HCFMUSP.

1169F

Common dysfunctional variants of urate exporter gene ABCG2/BCRP are a major cause of early-onset gout. *H. Matsuo[1], K. Ichida[2], T. Takada[3], A. Nakayama[1], H. Nakashima[4], T. Nakamura[5], Y. Kawamura[1], Y. Takada[6], K. Yamamoto[7], S. Shimizu[1], M. Sakiyama[1], T. Chiba[1], N. Hamajima[8], Y. Sakurai[4], T. Shimizu[9], N. Shinomiya[1].* 1) Dept. Integrative Physiol. Bio-Nano Med., Natl Defense Med Col, Tokorozawa, Japan; 2) Dept. Pathophysiol., Tokyo Univ. Pharm Life Sci, Tokyo, Japan; 3) Dept. Pharm., Univ. Tokyo Hospital, Tokyo, Japan; 4) Dept. Prev. Med. Public Health, Natl Defense Med Col, Tokorozawa, Japan; 5) Lab. Math., Natl Defense Med. Col., Tokorozawa, Japan; 6) Central Res. Inst., Natl Defense Med Col, Tokorozawa, Japan; 7) Med. Inst. Bioregulation, Kyushu Univ., Fukuoka, Japan; 8) Dept. Healthcare Admin., Nagoya Univ., Nagoya, Japan; 9) Midorigaoka Hosp., Takatsuki, Japan.

Background and objectives Gout is a common disease which mostly occurs after middle age, but more people nowadays develop it before the age of thirty. The major genetic causes of early-onset gout are still unclear, but some genetic predisposition has been supposed. ATP-binding cassette (ABC) transporter, subfamily G, member 2 (ABCG2/BCRP) is a high-capacity urate exporter which regulates serum uric acid levels. We conducted a study to determine whether common variants of ABCG2 cause early-onset gout. **Design, setting, participants & measurements** 705 Japanese male gout cases with onset age data and 1,887 male controls were genotyped, and the ABCG2 functions which are estimated by its genotype combination were determined. We then conducted a logistic regression analysis among them to investigate the involvement of ABCG2 dysfunction in early-onset gout. **Results** The onset age was 6.5 years earlier with severe ABCG2 dysfunction than with normal ABCG2 function ($P = 6.14\times10^{-3}$). Patients with mild to severe ABCG2 dysfunction accounted for 88.2% of early-onset cases (twenties or younger). Severe ABCG2 dysfunction particularly increased the risk of early-onset gout, conferring an adjusted odds ratio [OR] of 22.2 (95% confidence interval [CI] 5.89-83.7, $P = 4.66\times10^{-6}$). Moderate (1/2 function) and mild (3/4 function) dysfunction of ABCG2 also increased the risk of early-onset gout, conferring ORs of 15.3 (95% CI 7.53-30.9, $P = 4.08\times10^{-14}$) and 6.47 (95% CI 3.31-12.7, $P = 4.89\times10^{-8}$), respectively. **Conclusions** Common dysfunction of ABCG2 proved to be a major cause of early-onset gout. Early screening of ABCG2 dysfunction and appropriate interventions will greatly benefit high-risk individuals.

1170W

Quantitative genetic analysis of preterm birth in a sample of 2 million births yields precise heritability estimates. *W. Wu[1], A. Fraser[2], E. Clark[3], A. Rogers[4], G. Stoddard[5], D. Witherspoon[1], T. Manuck[3], K. Chen[6], S. Esplin[3], K. Smith[2], M. Varner[3], L. Jorde[1].* 1) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 2) Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT; 3) Department of Obstetrics and Gynecology, University of Utah School of Medicine, Salt Lake City, UT; 4) Department of Anthropology, University of Utah, Salt Lake City, UT; 5) Study Design and Biostatistics Center, University of Utah School of Medicine, Salt Lake City, UT; 6) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.

Preterm birth (PTB), defined as birth prior to 37 completed weeks' gestation, occurs in more than 10% of all births in the United States. PTB is the leading cause of neonatal mortality, and it causes a broad spectrum of lifelong morbidity in survivors. The etiology of spontaneous PTB (SPTB), PTB not due to obstetric or medical indication, is complex and has an important genetic component. Previous studies have shown that the heritability of SPTB is approximately 30%. However, these studies defined SPTB as a maternal phenotype, were assessed within selected families, were based on twin pairs, or were based on a dichotomous binary case/control design. Using the Utah Population Database (UPDB), we assessed SPTB in more than 2 million post-1945 births in Utah. We analyzed SPTB using the newborn's gestational age as a quantitative trait. Our estimate of SPTB heritability is 13.33% (95% confidence interval 10.96% - 15.69%), and the broad-sense heritability is 24.4%. These figures are lower than those of previous studies. In addition to estimating heritability, we partitioned other genetic and environmental components contributing to gestational age variance: dominance variance = 11.12%, common maternal environmental variance = 15.23%, and individual environmental variance =60.33%. Our results suggest that the environmental component of SPTB may be larger than previously realized, providing guidance for future studies. In particular, high-risk pedigrees with multiple affected cases, rather than samples of unrelated cases, have a greater likelihood of revealing genes that underlie this multifactorial condition.

1171T

Visceral adiposity linked to chromosome 9p24.2 in adults from the Fels Longitudinal Study. *B. Towne[1], J. Blangero[2], A.C. Choh[1], J.E. Curran[2], C. Bellis[2], T.D. Dyer[2], E.W. Demerath[3], M. Lee[1], S.A. Czerwinski[1].* 1) Wright State University School of Medicine, Dayton, OH; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) University of Minnesota, Minneapolis, MN.

Elucidating the genetic underpinnings of overweight and obesity continues to be a challenge. Part of the problem has been an over-reliance on crude measures such as weight or BMI. Using more direct and delineated phenotypic measures, however, especially of metabolically active fat depots, can provide new insights into the genetics of body composition and associated disorders. We present here results from genome-wide linkage analysis of visceral adipose tissue (VAT) volume in the entire trunk in 739 healthy adults (329 males; 410 females) aged 18 to 96 years (mean 46.1 years) in the Fels Longitudinal Study. VAT data were obtained using a multi-slice MRI protocol. Mean VAT was 3611 cc in males and 1735 cc in females. Subjects were all SNP genotyped using the Illumina Human 610-Quad BeadChip, and a subset of 17,583 SNPs identified as being in linkage equilibrium in our study population were used in whole-genome multipoint linkage analysis conducted using SOLAR (Almasy and Blangero, 1998). Significant linkage (LOD = 3.53) of VAT to markers on chromosome 9p24.2 was found, with the linkage peak being between SNPs rs7870248 and rs10758400. The 1-LOD support interval around this linkage peak contains several plausible positional candidate genes potentially relevant to visceral adiposity and cardiometabolic dysregulation, including *VLDLR* (very low density lipoprotein receptor), *GLIS3* (GLIS family zinc finger 3), and *RCL1* (RNA terminal phosphate cyclase-like 1). Of particular interest is *GLIS3* which is involved in the regulation of pancreatic beta cell development and insulin gene transcription. Recent GWA studies have found SNPs in *GLIS3* to be associated with increased risk of T1DM and T2DM. Those findings, and the results presented here, suggest that there may be a genetically mediated relationship between visceral adiposity and diabetes mellitus risk that warrants further investigation. Supported by NIH grants R01HD012252, R01DK064391, R01DK064870, R01AR052147, and R37MH59490.

1172F

Leveraging Genetic Information to Assess the Effect of Diabetes on Memory Scores in the Health and Retirement Study. *S. Walter[1], J. Daniel[1], L. Kubzansky[1], S.C. Chang[1], M.C. Cornelis[1], D.H. Rehkopf[2], M.M. Glymour[1,3].* 1) Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA, USA; 2) Department of Medicine, Stanford University, Stanford, CA, USA; 3) Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, USA.

Background: Extensive observational analyses suggest that type II diabetes increases risk of dementia and memory impairment, but this association may be spurious and reflect unmeasured common prior causes of diabetes and dementia. The research question is not easily amenable to experimental approaches, but natural experiments based on genetic instrumental variables (IV) can lend insight into whether lifelong elevations in diabetes risk are associated with differences in late life memory functioning. Methods: Data are from 6555 non-diabetic and 1947 diabetic white participants in the nationally representative Health and Retirement Study with complete information on genetics, self-reported diabetes status, and memory functioning. Based on 40 Single Nucleotide Polymorphisms (SNPs) shown in previous research to predict type II diabetes, we calculated a genetic risk score (GRS) as the sum of the number of risk alleles multiplied by the per allele effect size on type II diabetes (based on prior publications). Self-reported type II diabetes was used to confirm the GRS predicts diabetes within HRS. Memory score was based on a composite of immediate and delayed word recall plus a proxy report for extremely impaired respondents. Memory score was regressed on the GRS plus age, sex, and population stratification eigenvectors to test whether genetic variations in diabetes risk predict memory functioning. Results: Mean GRS was 4.07 (standard deviation [SD]= 0.44) in non-diabetics and 4.22 (SD=0.45) among diabetics. Each one unit change in the GRS was associated with a 13.0 percentage point (95% CI: 11.1, 15.0) increase in the probability of diabetes ($p=3.7\times10^{-38}$). The GRS was not associated with memory functioning (beta=0.003; 95% CI: -0.006, 0.013; p=0.5136). These associations provide no evidence of an effect of diabetes on memory but are consistent with an upper bound estimate of the effect of approximately a tenth of a standard deviation in memory score. Conclusions: A genetic risk score associated with a large increase in lifetime risk of developing diabetes was unassociated with memory functioning in this sample of older adults.

1173W

Employing a phenome-wide association study approach to investigate the pleiotropic nature of mitochondrial DNA variation. *S.L. Mitchell[1], S.A. Pendergrass[2], R. Goodloe[1], K. Brown-Gentry[1], R. McClellan[1], J. Boston[1], M. Allen[1], P. Mayo[1], N. Schnetz-Boutaud[1], D.G. Murdock[1], D.C. Crawford[1].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.

Hematologic abnormalities, including sideroblastic anemia, have been observed comorbid with mitochondrial disorders. Rare mitochondrial (mt) DNA point mutations in COX1 and MT-TL2 have been reported in patients with anemia. Data also support a role for common mitochondrial variation in human disease, such as type 2 diabetes & Alzheimer's, highlighting the potential pleiotropic nature of the mitochondrial genome. Thus far, phenome-wide association studies (PheWAS) to identify pleiotropy have focused predominantly on the nuclear genome. To characterize the pleiotropic effects of mtDNA variation, we, as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed the National Health and Nutrition Examination Surveys (NHANES), epidemiologic surveys conducted by the Centers for Disease Control and Prevention, which collect demographic, lifestyle, and health data from participants using questionnaires, medical exams, and laboratory measures; DNA is collected for a subset of participants. We selected and genotyped 63 mtSNPs to tag ancestral mitochondrial haplogroups for the major NHANES populations: non-Hispanic whites (NHW; n=8856), non-Hispanic blacks (NHB; n=4325), and Mexican Americans (MA; n=4766). A simplified mitochondrial phylogenetic tree rooted at the most recent common ancestor was constructed using these SNPs and a custom algorithm was developed to automate haplogroup classification. Preliminary tests of association were performed in NHW using linear or logistic regression to identify associations between European haplogroups and a range of phenotypes (n=132). Preliminary results revealed an association between mean cell hemoglobin and haplogroups U ($p=9.0\times10^{-4}$, $\beta=0.19$) and H ($p=3.6\times10^{-3}$, $\beta=-0.12$). The correlated phenotype mean cell volume was also associated with both haplogroup U ($p=3.2\times10^{-3}$, $\beta=0.42$) and haplogroup H ($p=1.5\times10^{-4}$, $\beta=-0.41$). The association of mitochondrial haplogroups with hemoglobin and mean cell volume is consistent with evidence that mitochondria are essential for heme synthesis. Notably, the haplogroup H-defining SNP mt7028 and the haplogroup U-defining SNP mt12308 are located in COX1 and MT-TL2, respectively. These preliminary results suggest that PheWAS is a valid approach for identifying novel mitochondrial genotype-phenotype correlations relevant to human health.

1174T

RNA-sequencing identifies novel differentially expressed coding and non-coding transcripts in Sjögren's syndrome. *C.J. Lessard[1,2], I. Adrianto[1], M.G. Dozmorov[1], G.B. Wiley[1], J.A. Ice[1], H. Li[1,2], J.A. Kelly[1], A. Rasmussen[1], S.B. Glenn[1], K.S. Hefner[3], D.U. Stone[2], G.D. Houston[2], D.M. Lewis[2], J.A. Lessard[4], J.M. Anaya[5], B.M. Segal[6], N.L. Rhodus[7], L. Radfar[2], J.B. Harley[8], J.A. James[1,2], C.G. Montgomery[1], R.H. Scofield[1,2], P.M. Gaffney[1], J.D. Wren[1,2], K.L. Sivils[1,2].* 1) Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Hefner Eye Care and Optical Center, Oklahoma City, OK; 4) Valley Bone and Joint Clinic, Grand Forks, ND; 5) Universidad del Rosario, Bogota, Colombia; 6) Hennepin County Medical Center, Minneapolis, MN; 7) University of Minnesota, Minneapolis, MN; 8) Cincinnati Children's Hospital Medical Center and the Department of Veterans Affairs Medical Center, Cincinnati, OH.

Sjögren's syndrome (SS) is a common, clinically heterogeneous autoimmune disease characterized by exocrine gland dysfunction that involves both innate and adaptive immune responses. SS etiology is complex, with environmental, genetic, and genomic factors contributing. Of the many genetic associations reported in complex diseases, >80% map to non-protein coding DNA sequences; however, many reside in regions shown to be transcriptionally active. We used RNA-seq to identify differentially expressed (DE) protein-coding (~3% of the genome) and non-coding transcripts in 57 SS cases and 37 healthy controls. RNA samples were isolated from whole blood and prepared for sequencing using the NuGEN Encore kit and sequenced on the Illumina HiSeq 2000. Raw FASTQ files were aligned to the human genome (hg19) using TOPHAT. DE transcripts were determined using DESeq with a false discovery rate (FDR) q-value of 0.05 and a fold change of >2. After alignment, the reads were summarized for 55,076 transcripts across the human genome annotated by Ensembl. A total of 2614 DE transcripts were identified. Of the protein-coding regions, SRP14 was the most statistically DE locus in the case-control analysis (q=2.03×10E-20, fold change (FC)=2.32). Two other DE protein-coding transcripts of interest were identified: UQCRB (q=1.94×10E-19, FC=2.86) and ATP5I (q=1.88×10E-18, FC=2.34). Among the 408 DE non-protein coding transcripts, we observed DE of a long non-coding RNA (lncRNA) at 2p25.1 (q=3.69×10E-5, FC=2.55). lncRNAs are important regulators of the human genome with diverse functions; however, most have yet to be characterized. Bioinformatics evaluation in the 2p25.1 region showed transcription factor binding sites and transcription of lncRNA sequences using immunologically relevant cell lines. To formulate functional hypotheses for the lncRNA at 2p25.1, we evaluated co-expression patterns with protein coding sequences and identified T cell activation and development as the most likely pathways influenced. In this SS RNA-seq study, we identified multiple candidate loci and, for the first time, DE lncRNA regions in SS. Although the function of the lncRNAs identified in this study are unknown, many others have been described to function as scaffolds, decoys, signals, and guides for various proteins by conferring nucleotide sequence specificity not possible by motifs alone. Future studies in SS are warranted to elucidate the functional consequences of these lncRNA.

1175F

Genetically distinct subtypes of Rheumatoid Arthritis. *B. Brynedal[1], L. Klareskog[2], L. Padyukov[2], L. Alfredsson[1], H. Källberg[1].* 1) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Complex diseases are heterogeneous in both their presentation and etiology. The identification of subgroups might therefore reveal more homogeneous populations with a smaller and distinct set of pathways. Earlier studies have demonstrated major heterogeneity regarding risk alleles in the HLA-DRB1 gene and immunologically defined (presence (ACPA+) or absence (ACPA-) of antibodies toward citrullinated peptides) subsets of Rheumatoid Arthritis (RA). We set out to characterize how the genotypic profile differs between subsets of RA patients and whether their genetic distinction is restricted to the HLA region. We focused our analysis on the combination of established risk factors of RA using two measures; genetic distance (GD) and genetic risk score (GRS). We used data from the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) case-control study encompassing 2747 cases (1856 ACPA+ and 891 ACPA-) and 1590 controls. The GRS was compared between groups (ACPA+, ACPA- and controls) including and excluding variants in the extended HLA region. We similarly calculated the GD between pairs of groups with and without HLA variants, and derived the empirical significance using permutation of group labels. ACPA+ and ACPA- RA were expectedly found to genetically distinct with a significant GD, and significantly different GRS. The exclusion of HLA variants removed much of this distinction, illustrating that RA subsets share a substantial part of their etiology. The GD between both groups of RA patients and controls is significant both including and excluding HLA variants. ACPA-RA patients and controls do not have a significantly different GRS, but excluding the variants close to the HLA region increases the GRS of ACPA-RA patients more than controls rendering the difference significant. This also illustrates that ACPA- RA patients carries proportionally more 'RA protective' HLA variants than controls. We conclude that previous published genetic risk factors show that there are both differences and similarities between subsets of RA defined by the absence or presence of ACPA. The major difference related to presence of ACPA is associated with alleles in the HLA region. Omitting the alleles in chromosome six makes ACPA+ RA and ACPA-RA more similar but different from controls. This exemplifies the importance of defining clear subsets of disease.

1176W

Assessing human craniofacial morphology as complex phenotype in genetic association studies. *F.I. Martinez[1, 2].* 1) Anthropology Program and ICIIS, Pontificia Universidad Catolica de Chile, Santiago, Chile; 2) Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, UK.

The availability of current genotyping and sequencing technologies combined with high-resolution medical imaging (such as MRIs and CT Scans) has led to a renewed interest in identifying genetic variants responsible for the genetic components of human craniofacial morphology. Recent GWAS have revealed, for the first time, genetic variants correlated with normal variation in human craniofacial phenotype. Some of these loci map to genes with previously known role in craniofacial development, thus implying strong support to the findings and suggesting an unambiguous link between genotype and phenotype. There is general agreement that many genes with small effect (hundreds or thousands) should influence complex phenotypes such as body height or face morphology. Therefore, it is relevant to note that linear measurements are the sole proxy for facial phenotype currently used in GWAS. If many genes with small effect influence complex phenotypes, then the reliability of linear measurements as proxies seems an unrealistic assumption. Is it feasible to assume that linear distances account for the effect of few loci between many genes implicated in craniofacial developmental pathways? Or, is it more likely that observed variability captured by linear measurements reflects mainly the effect of non-genetic patterns of morphological variation in the population such as age or sexual dimorphism? In order to explore these questions, a small sample of 100 virtually reconstructed medical CT scans was analyzed by means of geometric morphometrics and linear measurements. In addition, single nucleotide polymorphism rs6180 (I526L) in the exon 10 of the growth hormone receptor gene (GHR) was chosen as independent variable to run correlation analysis with craniofacial morphology. This polymorphism was selected due its previous link to cranial morphology, human height and the role of GHR in the GH/IGF-1 axis, that promotes bone growth. A significant statistical association was found between rs6180 and craniofacial morphology. Permutation of the data was performed in order to explore the influence of sample structure in the results. Although the findings suggest a putative effect of the rs6180 polymorphism in craniofacial morphology, the correlation is actually driven by the pattern of craniofacial growth in the population. Results show that linear measurements should be regarded with more caution when used as proxy for craniofacial morphology in association studies.

1177T

SNPs and smoking: What can the aggregate of genome-wide SNPs tell us about genetic liability to smoking initiation and quantity smoked?
A.G. Wills, M.C. Keller. Institute for Behavioral Genetics, University of Colorado, Boulder, CO. Department of Psychology, University of Colorado, Boulder, CO.

Twin studies have established the genetic contribution to cigarette smoking initiation and progression and have suggested that the dimensions underlying liability to each of these stages of the smoking trajectory may be correlated. Using whole-genome SNP data from a combined sample of 10,162 European Americans that had participated in the Atherosclerosis Risk in Communities Study (ARIC) and the Multi-Ethnic Study of Atherosclerosis (MESA), we examined the degree to which the aggregate of genome-wide SNPs contributed to trait variation, or 'SNP heritability', for smoking initiation and quantity (cigarettes/day multiplied by years smoked). We then estimated the genetic correlation, defined as the additive genetic correlation tagged by the aggregate of SNPs, between initiation and quantity. Using the entire sample, common SNPs explained 11% (p=.0001) of the variance in smoking initiation and, among smokers (n = 5967), only 4% (p=.2) of the variance in quantity smoked. However, when the quantity measure was analyzed separately for each sex, common genome-wide SNPs could not explain variation in quantity smoked for female smokers but explained 15% of the variance in quantity smoked among male smokers. Given that the standard error (.12) in our female analysis of smoking quantity was large, we continued analysis on the sexes combined for the bivariate analysis of initiation and quantity. In the bivariate analysis, we re-coded initiation as an ordinal variable that included both early and late onset users in addition to those that did not initiate to allow for the estimation of the genetic correlation between initiation and quantity. The correlation between the set of SNPs that explained variance in initiation and those that impacted quantity smoked was high (rg = .83) and not significantly different from 1 (p = .30). Thus, common genetic variation, as indexed by genome-wide SNPs, contributed to cigarette smoking liability, and the genetic factors that influenced smoking initiation were largely shared with those that impacted quantity smoked. Grant support: NICHD, 2T32HD007289, Stallings; NIMH, 1R01MH100141, Keller.

1178F

Endometrial expression profiling in women with recurrent early pregnancy loss.
G. Kosova[1], C. Billstrand[1], V.J. Lynch[1], M.D. Stephenson[2,3], C. Ober[1,2]. 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Obstetrics and Gynecology, University of Chicago, Chicago, IL; 3) Current address: Department of Obstetrics and Gynecology, University of Illinois College of Medicine at Chicago, Chicago, IL.

Recurrent early pregnancy loss (REPL), defined as at least two documented intrauterine miscarriages of less than 10 weeks of gestation, affects up to 5% of couples trying to conceive. The endometrium's ability to acquire a receptive state during the mid-secretory phase is a crucial component of successful implantation and, therefore, has been the focus of many studies that aim to identify predictive markers of early pregnancy. In order to better understand the molecular mechanisms that are perturbed in REPL, we examined the gene expression profiles of mid-secretory phase endometrium in patients with different diagnoses of endometrial dysfunction. Using the Illumina HumanHT-12v4 BeadChip arrays, gene expression levels were compared between women with 1) abnormally elevated cyclin E levels (n= 9) vs normal cyclin E levels (n=23), and 2) out-of-phase histological dating (n=10) vs normal histology (n=22). Differentially expressed (DE) genes were identified using a likelihood ratio test within a fixed-effects linear model framework. Overall, 81 genes were DE in the first comparison and 56 in the second at $P<0.001$. The two genes with the largest differential expression were *S100P* (3.5-fold up-regulation in women with high cyclin E), and *LEFTY2* (4.6-fold up-regulation in out-of-phase endometria). *S100P* expression peaks in the endometrium during the window of implantation (Tong et al. Fertil Steril 94:1510). *LEFTY2* levels normally decrease during that phase, and aberrantly high expression levels were observed in patients with implantation failure and infertility (Tabibzadeh et al. JCEM 85:2526). Gene enrichment analyses of the DE genes (at $P<0.01$) for each comparison revealed over-representation of biological processes including immune response, signal transduction, leukocyte activation and cell adhesion among down-regulated genes in women with elevated cyclin E; and cell adhesion, cell motility and TGF-β signaling among up-regulated genes in women with out-of-phase histology. Lastly, when the analyses were limited to genes that evolved function in the endometrium of Eutherian mammals, significant enrichments of DE genes were observed in both groups ($P=0.005$ in cyclin E, $P=0.003$ in histology). Taken together, the results of this study suggest that this is a promising approach for identifying genes and pathways whose tight regulation during the endometrial secretory phase is necessary for optimal implantation and successful pregnancy outcome.

1179W

Association study of ERCC3 genetic variations with nasal polyposis in asthmathics.
J.H. Kim[1], B.L. Park[2], C.S. Park[3], H.D. Shin[1]. 1) Research Institute for Basic Science, Sogang University, Seoul, South Korea; 2) Department of Genetic Epidemiology, SNP Genetics, Inc., Seoul, Republic of Korea; 3) Division of Allergy and Respiratory Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea.

Nasal polyps, as an inflammatory condition and one of the clinical symptoms of aspirin exacerbated respiratory disease (AERD), are edematous protrusions arising mainly from the mucous membranes of the nasal sinuses. Despite the well-defined triggers, the exact mechanisms involved in the development of asthma with nasal polyps still need to be clarified. The ERCC3, one of nucleotide excision repair family genes, has been considered to play roles in the inflammation mechanisms and cancers. To investigate the associations of ERCC3 polymorphisms with nasal polyposis in asthmatics, single nucleotide polymorphisms (SNPs) were genotyped in a total of 475 asthmatics including 161 nasal polyp-positive and 314 polyp-negative subjects. Results of logistic analysis revealed that SNPs and haplotypes were significantly associated with nasal polyp development in the overall asthma group. Despite the need for further replications and functional evaluations, our preliminary findings suggest that ERCC3 may contribute to nasal polyposis susceptibility in asthmatics.

1180T

Examining ocular SNPs for regulatory enhancer potential: What does a non-coding GWAS hit really mean?
C.N. Simonti[1], J.A. Capra[1,2]. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN.

Genome-wide association studies (GWAS) have enabled the discovery of genes and functional non-coding regions underlying many complex ocular diseases, such as age-related macular degeneration (AMD) and glaucoma. However, the mechanisms behind the vast majority of these associations remain to be elucidated. There are strong indications that many SNPs associated with ocular diseases may affect enhancer or regulatory regions. For example, two of AMD-associated SNPs in high linkage disequilibrium (LD; calculated from 1000 Genomes; $r^2 \geq 0.8$) have been shown to alter a transcription factor (TF) binding site in a manner that changes gene expression. As anticipated, most ocular diseases have associated non-coding SNPs near eye developmental genes. To assess the regulatory potential of all known ocular trait-associated SNPs,we intersected them with a genome-wide set of predicted regulatory enhancers identified from analysis of histone marks, evolutionary conservation, and sequence motifs in variety of cellular contexts. Ten percent of SNPs associated with ocular traits from the NHGRI GWAS catalog fall into predicted enhancers, a slightly higher fraction than the 8.5% of all trait-associated SNPs that overlap predicted enhancers. The majority of ocular disease SNPs in predicted enhancers did not have a gene within its LD region suggesting a regulatory function. Most of the remaining predicted enhancer SNPs were in sizable introns or known TF binding sites. Comparing the genomic context of SNPs associated with different ocular traits revealed several trait-specific patterns. For example, AMD is distinct from other ocular diseases in its immunologic and cholesterol-related contributing factors. Reflecting its uniqueness, AMD was the only ocular phenotype associated with SNPs nearby complement or cholesterol genes, several of which were in predicted enhancers. GWAS hits for glaucoma are enriched in predicted enhancers when compared to other ocular diseases like AMD, pathological myopia, and corneal astigmatism. SNPs associated with glaucoma demonstrating a distinct profile suggests new functional information may in fact be gleaned from such comparative analyses. Our analysis of the regulatory potential of ocular SNPs highlights contexts and variants relevant to the associations and suggests follow-up studies. For example, profiling the expression driven by the identified enhancer variants in ocular tissues would be a useful focus for future inquiries.

1181W

AutDB: A modular database for accelerating ASD genetic research. *E. Larsen, U. Kuppuswamy, S. Banerjee-Basu*. MindSpec Inc., McLean, VA.

A major focus of research in the post-genomic era is to decipher the heterogeneous genetic landscape underlying the pathogenesis of complex human diseases such as Autism Spectrum Disorders (ASD). A number of genes in which rare and/or common genetic variants thought to potentially play a role in ASD onset and pathogenesis have been identified. The advent of new techniques such as next generation sequencing (NGS) has resulted in a significant increase in the number of ASD genes with rare genetic variants. With the accelerated growth of genetic data obtained from ASD individuals adding to the already complex genetic landscape of this disease, there is a critical need for databases specialized in the storage and assessment of this data. The autism genetic database AutDB (http://autism.-mindspec.org/autdb/Welcome.do) was developed to serve as a publically available web-based modular database for the on-going curation and visualization of ASD candidate genes. Since its release in 2007, AutDB has been become widely used by individual laboratories in the ASD research community, as well as by consortiums such as the Simons Foundation, which licenses it as SFARI Gene. AutDB has been designed using a systems biology approach, integrating genetic information within the original Human Gene module to corresponding data in subsequent Animal Model, Protein Interaction (PIN) and Copy Number Variant (CNV) modules. The number of ASD susceptibility genes in the Human Gene Module of AutDB has increased from 304 genes in December 2011 to 546 genes in June 2013, which demonstrates both the continued discovery of ASD candidate genes and the ongoing curation of these genes into AutDB. In addition, the usage of NGS techniques has contributed to a dramatic increase in the number of rare variants identified in ASD candidate genes (from 1202 in Dec 2011 to 3061 in June 2013) compared to common variants (from 534 to 791 over the same period). Functional profiling of ASD genes with rare variants provides insight into the enriched molecular functions of these genes, including ionotropic glutamate receptor binding, serotonin receptor activity, beta-tubulin binding, and voltage-gated sodium channel activity. AutDB serves as a valuable resource for understanding the ever-evolving genetic landscape of ASD and provides researchers with information useful in bioinformatics analyses such as those described above that will aid in unraveling the molecular mechanisms underlying the disease.

1182T

General financial risk factors for risk and ambiguity associated respectively with dopamine D4 receptor (*DRD4*) and serotonin transporter (5-HTTLPR). *R.P. Ebstein[1], M. Monakhov[2], S.H. Chew[2]*. 1) Psychology Dept, National Univ Singapore, Singapore; 2) Economics Dept. National Univ Singapore, Singapore.

Only a handful of investigations have focused on the neurogenetic underpinnings underlying financial decisions that are risky (uncertain outcomes with certain probabilities) or ambiguous (uncertain outcomes with uncertain probabilities) despite the fact that such attitudes are heritable and stable traits. Little is known regarding specific genes that correlate with the propensity to take financial risk especially under ambiguity. Two widely studied common polymorphisms in human behavioral genetics, the *DRD4* exon III VNTR and the serotonin transporter gene (*SLC6A4*) 5-HTTLPR promoter repeat, are biologically plausible candidates for contributing to risk attitude. A large number of Han Chinese university students (N=2,351) completed a battery of incentivized economic tasks and a risk preference phenotype (gR) was extracted from 12 moderate gain gambles using principal component analysis (PCA). Overall, risk proneness was significantly associated with the long 5HTTLPR *SLC6A4* repeat whereas carriers of the *DRD4* 4R/4R genotype showed less risk taking compared to all other exon III genotypes (mainly 2R). Notably, when moderate gain tasks are further informed by comparing gambles with known and unknown probabilities, 5HTTLPR is associated with risk only for ambiguous gambles whereas the *DRD4* is associated with risky gambles only with the known deck. Altogether these results suggest a general risk factor for moderate gain domain, fine-tuning the analysis shows that there are two general risk factors for risky and ambiguous gambles. The neurogenetic approach distinguishes between these two preferences and identifies distinct genes, indexing dopamine and serotonin neural pathways that respectively underpin risky and ambiguous decision-making. Overall, these neurogenetic findings strengthen the notion that risk attitude towards gambles with known probabilities and uncertain probabilities not only have a different anatomical architecture but also differ at the neurochemical and DNA level.

1183F

A molecular genetic study of amyotrophic lateral sclerosis/parkinsonism-dementia complex (ALS/PDC) in Kii peninsula of Japan. *H. Ishiura[1], Y. Fukuda[1], B. Ahsan[1], J. Mitsui[1], Y. Kokubo[2], S. Kuzuhara[3], S. Yoshida[4], T. Kihira[4], J. Yoshimura[5], K. Doi[5], K. Higasa[5], S. Morishita[5], A. Toyoda[6], A. Fujiyama[6], R. Kuwano[7], K. Hara[7], M. Nishizawa[7], J. Goto[1], S. Tsuji[1]*. 1) Dept Neurology, Univ Tokyo, Tokyo, Japan; 2) Department of Neurology, Mie University, Mie, Japan; 3) Department of Medical Welfare, Suzuka University of Medical Science, Mie, Japan; 4) Department of Neurology, Kansai University of Health Science, Osaka, Japan; 5) Department of Computational Biology, Univ Tokyo, Tokyo, Japan; 6) Comparative Genomic Laboratory, National Institute of Genetics, Shizuoka, Japan; 7) Brain Research Institute, Niigata University, Niigata, Japan.

[Background] ALS/PDC is a neurodegenerative disorder characterized clinically by a combination of upper and lower motor neuron involvement, parkinsonism, and dementia, and neuropathologically by wide-spread neurofibrillary tangles in the central nervous system. ALS/PDC has been described to be prevalent in the island of Guam, southern West New Guinea, and Kii peninsula of Japan. In Kii peninsula of Japan, there are two major disease foci (southernmost and north eastern parts of the peninsula). Presence of multiplex families has been observed, in particular, in the north eastern region, suggesting involvement of genetic component to the disease. [Method] Twenty six patients and 66 unaffected family members from 12 multiplex families, 15 apparently sporadic patients and 24 unaffected family members, and 27 regional control subjects from the north eastern region and 1 familial patient and 14 apparently sporadic patients in the southernmost region were included in the study. Extensive mutational analyses were conducted employing whole genome resequencing of 7 patients with pathological diagnosis of ALS/PDC. In addition, recently identified repeat expansions in *C9ORF72* was analyzed by repeat-primed PCR analysis. [Results] Whole genome resequencing analysis of 7 patients with the pathological diagnosis of ALS/PDC showed coverages of 35X - 73X. Of 3.5 - 3.7 million SNVs called in each patient, no shared novel nonsynonymous substitution was revealed. Repeat-primed PCR analysis revealed three patients with expansion of hexanucleotide repeats in *C9ORF72*. These patients are from the southernmost region of Kii peninsula. Haplotype analysis revealed an extended shared haplotype spanning 3.3 - 63 Mb, suggesting a fonder effect in this region. [Conclusion] Although *C9ORF72* mutations can explain in part a high prevalence of ALS in the region, the majority of the patients in these two regions cannot be accounted for by the *C9ORF72* mutations. Furthermore, whole genome sequence analysis did not reveal novel nonsynonymous mutations shared among the seven patients, raising various possibilities including genetic heterogeneity, structural variation, and repeat expansion mutation.

1184W

HLA-DQB1*03:02 is associated with narcolepsy in the Japanese population. *T. Miyagawa[1], H. Toyoda[1], A. Hirataka[1], SS. Khor[1], M. Yamasaki[1], Y. Honda[2], M. Honda[2,3], K. Tokunaga[1]*. 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Somnology, Tokyo Medical University, Tokyo, Japan; 3) Sleep Research Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan.

Narcolepsy, a sleep disorder characterized by excessive daytime sleepiness, cataplexy, and rapid eye movement (REM) sleep abnormalities, is tightly associated with human leucocyte antigen (HLA)-DQB1*06:02. Almost all the patients carry HLA-DQB1*06:02. Since HLA-DQB1*06:02 allele is common in the general populations (10-30%), it has been suggested that it is almost necessary but not sufficient for the development of narcolepsy. To investigate the influence that additional HLA-DQB1 alleles have on susceptibility or resistance to narcolepsy, We examined HLA-DQB1 in 625 Japanese narcoleptic subjects and 1,418 Japanese control subjects The strongest association was with HLA-DQB1*06:01 ($P = 2.2\times10^{-10}$, OR = 0.37) as reported in previous studies. We also found additional predisposing effects of HLA-DQB1*03:02 ($P = 2.7\times10^{-7}$, OR = 1.93) as the second hit in this study. A conditional analysis controlling for the effect of HLA-DQB1 revealed an independent association with HLA-DPB1*05:01 in narcolepsy ($P = 1.8\times10^{-3}$, OR = 1.44). We also performed case-control analyses of HLA-DQB1 for narcolepsy without cataplexy (n = 110) and idiopathic hypersomnia without long sleep time (n = 82). Predisposing effects of HLA-DQB1*06:02 ($P = 7.3\times10^{-9}$, OR = 2.81) and protective effects of HLA-DQB1*06:01 ($P = 2.2\times10^{-3}$, OR = 0.50) were observed in narcolepsy without cataplexy, as with narcolepsy. No significant association was seen between idiopathic hypersomnia without long sleep time and HLA-DQB1 alleles. These results indicate complex HLA associations contribute to the genetic predisposition to narcolepsy and narcolepsy without cataplexy.

1185T

Association between the dopamine D4 receptor (*DRD4*) and political ideology in a large Singapore Han Chinese sample. *M. Monakhov[1], S.H. Chew[1], R.P. Ebstein[2].* 1) Economics Dep't., National University of Singapore, Singapore; 2) Psychology Dep't., National University of Singapore.

The liberalism - conservatism ideological divide has become pervasive in American politics viz., red and blue states, and understanding the determinants of this continuing conflict is of keen interest not only to political scientists and psychologists, but also to the media and the public at large. Importantly, there are marked individual differences in political ideology that have been attributed to both dispositional and situational factors. In addition to situational factors contributing to political orientation, accumulating evidence shows that social attitudes are partially heritable suggesting that genes may contribute to individual differences in liberalism - conservatism. A specific gene polymorphism, the dopamine D4 receptor (*DRD4*) exon III VNTR was recently identified by Fowler and his group as provisionally contributing to individual differences in political ideology, but contingent on the subject's number of friends. By leveraging our ongoing investigation of economic decision-making, we had the opportunity to undertake an independent replication of the association between *DRD4* and political ideology in a large sample of ~1,800 Han Chinese university-students in Singapore. We employed the same self-report question, as did Fowler's group, to measure political ideology and demonstrate an association between the *DRD4* exon III repeat region and ideology for students who grew up in a singular Singaporean political culture and are characterized by a distinctive allele distribution compared to their American counterparts. In the combined male and female sample, carriers of *DRD4* exon III VNTR genotype 4R/4R reported more conservative political orientation, compared to all other genotypes (ordered logit coeff. = -0.18, p = 0.04, N = 1803). The association was evident also in an additive genetic model (regression of political attitude on number of 4R alleles in the genotype, coef = -0.15, p=0.027). The association was observed only in females (coeff.=-0.385, p=0.002, N=901) and not in male subjects. Cliff's delta (effect size measure for ordinal data) was 0.057 in combined sample and 0.11 in females subsample. Similar to the first study by Fowler's group, we also observed some evidence that the gene association depends on subject's number of friends. Finally, we have evidence that the main effect of DRD4 on political ideology in women is mediated in part by the NEO-PI-R personality trait of Neuroticism (Sobel test: p=0.035).

1186F

An approach to complex disease modifier genes and mouse-to-human translation: transcriptome analysis of mouse strains reveals Alzheimer disease modifier gene. *T. Morihara[1], N. Hayashi[1], H. Akatsu[2], M. Silverman[4], M. Yokokoji[1], N. Kimura[5], M. Sato[1], K. Kamino[3], Y. Yamagichi[6], T. Tsunoda[7], T. Tanaka[1], M. Takeda[1].* 1) Psychiatry, Osaka University, Suita, Osaka, Japan; 2) Choju Medical Institute, Fukushimura Hospital, Toyohashi-shi, Aichi, Japan; 3) National Hospital Organaization Yamato Mental Medical Center, Nara, Japan; 4) Simon Fraser University Vancouver Cabada; 5) Laboratory of Disease Control, Tsukuba Primate Research Center, National Institute of Biomedical Innovation Tsukuba, Japan; 6) Tohoku University, Sendai, Japan; 7) Riken, Yokohama, Japan.

Background: Alzheimer's disease (AD) is a common complex disease and characterized by the accumulation of Abeta in brain. To identify Abeta modifier gene(s) efficiently, we postulated that combining distinct mouse strains and model mice with transcriptomics would have a synergistic effect and minimize the drawbacks encountered in approaches like genome-wide association studies (GWAS), human transcriptome studies and mouse quantitative trait locus (QTL) analysis. Methods: To avoid detecting secondarily affected genes by Abeta, we used non-Tg mice in the absence of Abeta pathology and selected candidate genes differently expressed in the certain strain. We also prepared AD model (APP-Tg) mice with mixed genetic backgrounds. APP Tg mice were crossed onto three different strains, C57BL6 (B6), SJL and DBA/2 (DBA). The levels of Abeta in the cortex were measured by ELISA. The levels of mRNA in the hippocampus were measured using Illumina Mouse-Ref-8 (40 arrays) and QPCR. SNPs in mice with mixed genetic background were analyzed by TaqMan assay. We also measured the total and splice variant E levels of KLC1 in human brain (14 control and 10 AD) and lymphocyte (17 control and 47 AD). The functional analysis of the identified gene is performed using transfected neuroblastoma cells. Results: Abeta accumulation levels in mice with DBA rich genetic backgrounds were drastically lower, less than one third of those in SJL or B6 rich genetic background (e.g. APP Tg mice whose 75% genetic background came from DBA had only 15% Abeta40 and 32% Abeta42 of those with 69% SJL and 31% B6 genetic background.) To identify the gene(s) which control Abeta accumulation, we performed transcriptomics in the mouse brains. QPCR analysis confirmed the correlation between the levels of identified gene and Abeta (Pearson's product-moment correlation R2=0.33, p<0.0001). Genomic analysis in identified gene also supported these findings. Because the transcriptome of brain is considerably conserved between species, we continued with our transcript analysis approach to examine whether the findings in mice can be translated to humans. We observed that the mRNA levels of identified gene were elevated in AD brain (+30.7%, p=0.0096).

1187W

Expression of neurotransmitter receptors and regulators genes in peripheral blood of a drug-naïve first-episode psychosis sample before and after antipsychotic treatment. *M.L. Santoro, V.K. Ota, C.S. Noto, A. Gadelha, P.N. Silva, M.I. Melaragno, M.A.C. Smith, Q. Cordeiro, R.A. Bressan, S.I. Belangero.* Federal University of Sao Paulo, Sao Paulo, Brazil.

The psychotic symptoms in Schizophrenia usually manifest in the late adolescence and early adulthood. The investigation of patients during their first episode of psychosis (FEP), particularly before they take any antipsychotic medication, could be especially helpful to disentangle the biological underpinnings of disease onset, progression and the antipsychotic effects. In this way, our aim was to carry out an expression profiling study comparing: 1) drug-naïve FEP patients (n=10) and healthy subjects (n=9); and 2) drug-naïve FEP patients before and after (n=10) 8 weeks of risperidone treatment. Blood samples were collected and RNA was isolated and reverse-transcribed to cDNA. To verify gene expression we used PCRarray technology, which assesses expression of 84 neurotransmitter receptor and regulatory genes plus five housekeeping genes. We considered as a detectable level of expression those genes with ΔCt < 35 - HKG (housekeeping genes) geometric mean of each group sample. We were able to identify three groups of genes: A, genes not expressed in any group (14 genes); B, genes not expressed in some groups group (10 genes); and C, genes expressed in all groups (55 genes). For group C, we performed Mann-Whitney U test (FEP×Control) or the Wilcoxon test (FEP before × after treatment) and Bonferroni correction for multiple comparisons, and for group B we used Fisher exact test, in order to find genes that are expressed only under a specific condition. After Bonferroni correction, none of the associations remained for the group C of genes (p > 0.009). Concerning group B, we found that PROKR1 expressed preferentially in drug-naïve FEP patients compared to the FEP patients after treatment (p=0.02), and with a tendency between FEP and controls (p=0.07). Our results suggest that treatment with risperidone may decrease expression levels of PROKR1, returning to undetectable expression in FEP patients after treatment such as the control group. However, this gene has not been investigated in psychiatric disorders, and further studies are needed to investigate the role of the prokineticins in antipsychotic treatment. These present data are relevant in terms of cost and effectiveness to studies of gene expression. PROKR1 was expressed only in FEP group, but not in others, indicating that its expression could be regulated by the psychosis condition or risperidone treatment.

1188T

The cerebral glucose transporter SLC45A1 is mutated in individuals with non-syndromic intellectual disability and epilepsy. *M. Srour[1,2], N. Shimokawa[3], F.F. Hamdan[1], S. Dobrzeniecka[4], G.A. Rouleau[5], C. Poulin[2], J.L. Michaud[1].* 1) Sainte-Justine Hospital Research Center, Montréal, H3T 1C5, Canada; 2) Division of Pediatric Neurology, Montreal Children's Hospital-McGill University Health Center, Montreal, H3H 1P3, Canada; 3) Department of Integrative Physiology, Gunma University Graduate School of Medicine, Gunma 371-8511, Japan; 4) CHUM Notre-Dame Hospital Research Center, Montreal, H2L 2W5, Canada; 5) Montreal Neurological Institute, McGill University, Montreal, H3A 2B4, Canada.

We have identified a consanguineous Lebanese family with two sisters affected with moderate non-syndromic intellectual disability and epilepsy of unknown etiology. SNP genotyping was performed in the two affected sisters as well as in their unaffected sibling. We identified 7 regions larger than 1Mb that were homozygous in the two affecteds but not in the unaffected sister. Whole exome sequencing was performed in one affected individual. The exome dataset was analyzed by looking for homozygous rare (<0.5% minor allele frequency) coding and splicing variants found in the candidate homozygous regions. We excluded variants that did not segregate with the phenotype in the family or that were found in 95 ethnically-matched controls. Of the remaining variants, only one missense variant in *SLC45A1* (c.C629T/ p.A210V) was predicted damaging by Polyphen-2 and mutation taster. *SLC45A1* encodes a glucose transporter that has only recently been characterized. Slc45a1 is highly expressed in the brain of mice, is induced after hypercapnia and mediates glucose uptake along the pH gradient. We show that *Slc45a1*-transfected COS-7 cells carrying the same c.C629T (p.A210V) mutation exhibited a 40-50% decrease in intracellular glucose uptake, suggesting that the mutation affects the function of the Slc45a1 transporter. Mutations in another cerebral glucose transporter, *GLUT1*, are implicated in neurologic disease and are well known to result in epilepsy with variable degrees of intellectual disability, or movement disorders. Our results suggest the possibility of specific treatment in the form of a ketogenic diet for the family, and implicate for the first time *SLC45A1* in human disease. Identification of additional similarly affected families will be necessary to establish *SLC45A1* as a neurodevelopmental disease gene. Collaborations towards this effort are welcome.

1189F

Adducin function is essential for sustained changes in neuronal activity upon learning. *V. Vukojevic[1,2], F. Peter[1,2], P. Demougin[1,2], N. Hadziselimovic[1,2], J.-F de Quervain[2,3,4], A. Papassotiropoulos[1,3,4], A. Stetak[1,3,4].* 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 2) University of Basel, Department Biozentrum, Life Sciences Training Facility, Klingelbergstrasse 50/70, 4056 Basel, Switzerland; 3) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 4) University of Basel, University Psychiatric Clinics, Wilhelm Klein-Strasse 27, 4055 Basel, Switzerland.

Identifying molecular mechanisms that underlie learning and memory is one of the major challenges in neuroscience. Previously, we have reported that genetic variability of the human ADD1 gene is significantly associated with episodic memory performance in healthy young subjects. Taken the advantages of the nematode Caenorhabditis elegans, we have shown that α-adducin is required for consolidation of synaptic plasticity, for sustained synaptic increase of AMPA-type glutamate receptor (GLR-1) content and altered GLR-1 turnover dynamics. ADD-1, in a splice- form- and tissue-specific manner, controlled the storage of memories presumably through actin-capping activity (Vukojevic et al, 2012). In the present work, we further investigated the effects of adducin dependent changes on neuronal activity during synaptic transmission. With the help of genetically encoded calcium indicators (GECIs) we visualized the neuronal activity in vivo in defined neuronal populations. Specifically, we monitored calcium currents in neurons crucial for avoidance behavior, especially in AVA and RIM neurons that controll activation and inhibition of reversals. Taken together, the method implemented enabled us to get the insight of function and integration at the level of a single neuron. Moreover, we were able to caught in vivo, a single neuronal cell in learning and memory formation. Remarkably, in add(lf) mutants the learning phase induced changes in AVA and RIM neuronal activity fail to be consolidated after 30 minutes. Therefore we were also capable to investigate in vivo the role of neurons in memory storage. These findings further support the role of ADD-1 in the stabilization of synapses, changes in GLR-1 dynamics and finally modulations of neuronal activity patterns. All together, our results suggest that the lack of adducin in AVA interneuron has consequences on synapse remodelling and changes of neuronal activity that are functionally reflected also in other members of the motor network.

1190W

Influence of Alzheimer's disease genes on cognitive decline: the Guangzhou Biobank Cohort Study. *S.S. Cherny[1], H.S. Gui[1], L. Xu[2], P.C. Sham[1], C.Q. Jiang[3], T.H. Lam[2], B. Liu[3], Y.L. Jin[3], T. Zhu[3], W.S. Zhang[3], G.N. Thomas[4], K.K. Cheng[4].* 1) Psychiatry & Centre for Genomic Sciences, Univ Hong Kong, Pokfulam, Hong Kong; 2) Department of Community Medicine, School of Public Health, University of Hong Kong, Hong Kong; 3) Guangzhou No. 12 Hospital, Guangzhou, China; 4) Public Health, Epidemiology, and Biostatistics, University of Birmingham, Birmingham, UK.

Mild cogntive impairment is a reliable predictor of the future onset of clinical dementia. In Hong Kong, the prevalences of very mild dementia and mild dementia in people aged 70 years and older were estimated to be 8.5% and 8.9%, respectively, in a population-based sample. For those subjects with a clinical diagnosis of dementia, Alzheimer's disease (AD) was the most common likely cause (73.5%), with 22.4% having vascular dementia (VaD), and 3.9% dementia with symptoms of Parkinson's disease. In addition to the impact on the patient and their family members, the economic cost of AD has been estimated to range from around $48,000 to $585,000 per patient per year around the world. To identify risk factors involved in cognitive decline, we selected 1325 extreme cognitive decline subjects and 1083 matched controls from Guangzhou Biobank Cohort Study (GBCS) for DNA genotyping at 30 known AD associated SNPs. Full information maximum likelihood (FIML) regression was adopted to analyse quantitative cognitive change scores, while multiple logistic regression was used to investigate the genetic effect of those variants on dichotomous phenotype. No allelic association was found by individual variant analysis. At the level of genotypic association, not only did we confirm that the APOE $\varepsilon 4$ homozygote can significantly predict cognitive decline ($p < 0.05$), but also revealed carriers of the ACE rs1800764_C allele are more likely suffer decline than non-carriers, especially in the samples without college education. However, these effects together only explain 1.3% of the phenotypic variance, and suggest that AD risk variants/genes are only minor predictors of cognitive decline in these Chinese samples.

1191T

Fine Mapping and Association Analysis of Candidate Genes for Autism Spectrum Disorder and Language Impairment in the NJLAGS Sample. *A. Hare[1], A. Seto[1], J. Flax[1], M. Azaro[1], S. Buyske[2], C. Bartlett[3], L. Brzustowicz[1].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Department of Statistics, Rutgers University, Piscataway, NJ; 3) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and The Ohio State University, Columbus, OH.

Over the past decade, the New Jersey Language and Autism Study (NJLAGS) has collected families that contain one individual with autism and another individual with Specific Language Impairment (SLI) but not autism. This is the first study of its kind to investigate the share genetics between autism and SLI. Using a comprehensive neuropsychological testing battery, three categorical phenotypes: language impairment (LI), reading impairment (RI), and social impairment (SRS-DT), and two quantitative phenotypes: social impairment (SRS-QT) and obsessive-compulsive behaviors (YBOCS) were developed. Autism proband scores were included in the quantitative phenotypes and all categorical phenotypes included autism diagnosis as impaired for language, reading, or social skills, respectively. A previous study identified linkage in these families to 13q21.2 (YBOCS), 14q32.31 (SRS-QT), 15q25.1 (LI), 15q26.2 (SRS-DT), and 16p12.3 (RI). As genome-wide association did not reveal strong evidence for association, Ingenuity Pathway Analysis (IPA) was used to select candidate genes for fine mapping analysis. Four groups were analyzed in IPA: communication impairment (LI + RI), social impairment (SRS-DT + SRS-QT), restricted/repetitive behaviors (YBOCS), and an overall autism model (LI + RI + SRS-DT + SRS-QT + YBOCS). Genes in each analysis group were given a score that corresponded to the number of relevant functions identified by IPA. Seven of the highest-ranking genes were selected for association analysis from the LI, RI, and SRS-DT linkage regions: AKT1, JAG2, PTPN9A, SEMA7A, NTRK3, FES, and SCCN1B. Each gene was fine mapped using an oligonucleotide ligation assay and was analyzed for association using the KELVIN framework. Each gene was analyzed for association to its respective phenotype with and without the inclusion of autism diagnosis and for autism diagnosis only. JAG2 yielded the strongest evidence for association (PPLD|L = 5%) when autism diagnosis was included. When autism diagnosis was excluded, NTRK3 yielded the strongest evidence for association (PPLD|L =7%). For autism diagnosis only, no evidence for association was detected. The lack of strong evidence for association provides support for a role of rare variants in autism susceptibility and their role in the shared genetics between autism and SLI.

1192F

***TRPM8* variants that protect from migraine make individuals less susceptible to cold pain.** *M.A. Kaunisto[1,2], E.J. Holmström[2], V. Anttila[1,3,4,5], M. Kallela[6], E. Hämäläinen[1,5], A. Stubhaug[7], M. Wessman[2,1], C.S. Nielsen[8], E. Kalso[9,10], A. Palotie[1,4,5].* 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Folkhälsan Research Center, Helsinki, Finland; 3) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 6) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland; 7) Division of Emergencies and Critical Care, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 8) Division of Mental Health, Norwegian Institute of Public Health, Oslo, Norway; 9) Institute of Clinical Medicine, University of Helsinki, Finland; 10) Pain Clinic, Department of Anaesthesia and Intensive Care Medicine, Helsinki University Central Hospital, Finland.

Several common variants predisposing to migraine have recently been identified. The most consistently associating variants lie within or around *TRPM8*. TRPM8 is a ligand-gated cation channel that plays a key role in sensing cold pain. Although the function of this channel has been intensively studied, no evidence of *TRPM8* variants affecting cold pain sensitivity exists.

To elucidate the putative role of *TRPM8* in cold pain, we genotyped two migraine-associated SNPs, rs10166942 and rs17862920, in two study cohorts. First of these, the BrePainGen study cohort consists of 1,000 Finnish patients undergoing breast cancer surgery and tested for experimental pain before surgery. From the second cohort, the population-based Tromsø Study, 2,369 Norwegian individuals tested for experimental cold pain were studied. Cold pain was measured in both cohorts by immersing the hand to circulating +3ºC water for the maximum time tolerated by the study subjects. Pain intensity was assessed at regular intervals with a 0 to 10 Numerical Rating Scale (NRS). Association between *TRPM8* SNPs and NRS values as well as time to withdrawal was examined with linear regression analysis. Sex, age, BMI and anxiety status were included as covariates. The primary phenotype was the NRS value reported soon after the beginning of the test (at 15 s or 13 s, depending on the cohort).

Meta-analysis combining the results of these two cohorts showed that both of the SNPs were associated with cold pain sensitivity at the specified time-point (p=0.001, beta=-0.24 for rs10166942 and p=0.004, beta=-0.3 for rs17862920, assuming an additive model). The carriers of the minor alleles reported less intense cold pain than the non-carriers which is in line with the fact that the minor alleles of these SNPs are known to protect from migraine. The mean NRS of the homozygous rs17862920 T-allele carriers was 1.2 scores lower, compared to the rest of the subjects.

When the later time-points were examined, it became evident that the effect was only seen in the beginning of the test. Furthermore, there was no evidence of association between the SNPs and cold pain tolerance. This suggests that the difference is in the early sensory processing of cold pain. Based on these results it seems possible that rather than being migraine specific susceptibility variants these SNPs may have a role in more general pain sensitivity mechanisms.

1193W

The Role of CNTNAP2 in a Developmental Language Disorder and Potential Mechanisms. *N. Li[1], S.L. Wolock[1], S.A. Petrill[2], J.F. Flax[3], A.S. Bassett[4], L.M. Brzustowicz[3], C.W. Bartlett[1].* 1) The Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and Department of Pediatrics, The Ohio State University, Columbus, OH; 2) Department of Psychology, The Ohio State University, Columbus, OH; 3) Genetics, Rutgers University, Piscataway, NJ; 4) Department of Psychiatry, University of Toronto, Toronto CA.

Background: The purpose of this study was to assess CNTNAP2 in a sample of extended pedigrees ascertained for specific language impairment (SLI). SLI is a developmental failure to learn vocabulary and grammar at the expected rate despite normal cognition, speech patterns and hearing. CNTNAP2 is associated with different behavioral disorders including SLI and autism, as well as with normal language development. However, studies to date have focused on SNPs across the gene without regard to functional potential. Here, we performed three studies to examine the potential function of CNTNAP2 in language development. Methods: STUDY 1: We used dense SNPs in 21 nuclear and extended SLI families from the U.S. and Canada. Nine language and reading phenotypes were tested for linkage/association. STUDY 2: We conducted a meta-analysis of 798 subjects from five published human brain expression Quantitative Trait Loci (eQTL) studies with SNPs imputed to the 1000 Genomes reference haplotypes. STUDY 3: We performed follow-up analysis of SNPs in 106 human brain samples (frontal lobe) from our lab. Results: STUDY 1: Using the quantitative trait analysis of a standardized language assessment, we observed a multipoint posterior probability of linkage (PPL) of 85% (MOD=4.7). Analysis of a categorical written language impairment diagnosis yielded a PPL of 40%, and similar results occurred with the quantitative trait underlying the categorical written language impairment trait. There was no linkage with intelligence in this sample. STUDY 2: Two SNPs directly underneath the multi-point linkage peak were also found to be eQTLs for CNTNAP2 expression in frontal cortex with posterior probability of linkage disequilibrium (PPLD) values from 10%-39%. STUDY 3: Follow-up analysis using qPCR on an in-house collection of brain samples also indicated three SNPs that were potentially eQTLs for CNTNAP2. Conclusions: While CNTNAP2 is associated with multiple neurocognitive phenotypes, we observed no evidence that CNTNAP2 influences general intelligence. Furthermore, the linkage peak in our samples appears to be driven by one or more eQTLs within CNTNAP2 that influence expression in a genotype dependent way. Studies to isolate the causal SNP(s) that actually influence expression levels are warranted.

1194T

Association of Language Impairment with Language, Reading, and ADHD candidate genes. *S. Smith[1], J. Gayan[2], M. Rice[3].* 1) Dept Pediatrics, Univ Nebraska Med Ctr, Omaha, NE; 2) Bioinfosol, Sevilla Area, Spain; 3) Dept. Speech, Language, Hearing, Univ Kansas, Lawrence, KS.

Specific language impairment is a significant developmental disorder in language acquisition occurring in about 7% of 6-year-olds in the US. Estimates of heritability vary based on the severity, with heritabilities greater than 0.5 for children who had been seen by a speech pathologist. We previously reported association of LI with SNPs in a Reading Disability candidate gene, KIAA0319, and with a language candidate gene, FOXP2. The current study expands the coverage of these and additional genes, with concentration on SNPs in regulatory regions. We genotyped 422 SNPs in 23 candidate genes in 293 people from 62 nuclear families with a proband with SLI. Probands were between 3 and 12 years old and were ascertained through school-based speech pathologists, with diagnosis confirmed with standardized testing appropriate for the child's age. Their siblings between ages 2 and 17 were also recruited and tested. All children were given an additional battery of 9 tests covering reading and language abilities. Quantitative association analysis was done with the PBAT program implemented by SVS (Golden Helix) testing the hypothesis of association in the presence of linkage since these were candidate genes with some previous evidence for relationship with LI or a related disorder. Several genes and phenotypes showed association with multiple SNPs. The greatest evidence for association was in the expected genes FOXP2, CNTNAP2, and KIAA0319, but additional evidence was found with DRD2, BDNF, and DBH. The strongest results were with FOXP2. The most significant SNPs were concentrated at the 5' end of the gene, with the haplotype rs923875.rs7785744.rs73716355 (C:T:G) at a nominal p=0.000074 for the Gray Oral Reading phenotype. For CNTNAP2, the haplotype rs12703865.rs9640492.rs10952672 (A:G:A) showed a nominal p=0.0013 for a measure of vocabulary growth. For KIAA0319, the strongest haplotype was not in the same region that is generally associated with reading: rs807533.rs2760179.rs807509 (G:G:C) p=0.00098 for a reading phenotype; however, another haplotype was located within the putative reading disability region: rs3087943.rs2294689.rs2143340 (G:C:T) with the Gray Oral reading phenotype, p=0.0037. We also replicated association with PCDH9 on chromosome 13 with 3 SNPs, all with the TEGI grammar phenotype: rs2324908 (p=0.0084), rs4883774 (p=0.0094), and rs9564308 (p=0.0094). An additional 21 families has been genotyped on the same platform and will provide additional data.

1195F

Molecular delineation of the 1p36.3 locus and candidate gene resequencing in 363 patients with polymicrogyria. *S. Beijnsberger[1], G.L. Carvill[1], S.F. Berkovic[2], I.E. Scheffer[2], W.B. Dobyns[3], H.C. Mefford[1].* 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA 98195, USA; 2) Epilepsy Research Center and Department of Medicine, University of Melbourne, Austin Health, Australia; 3) Pediatrics and Neurology, University of Washington, Seattle, WA, 98195, USA.

Polymicrogyria (PMG) is characterized by an abnormal folding pattern of the cortex, which results in an excessive number of gyri and abnormal cortical lamination. PMG occurs in several forms and causes a wide spectrum of phenotypes, ranging from minimal neurologic manifestations to severe intellectual disability and refractory epilepsy. Mutations in a small number of genes have been shown to cause PMG, including *TUBA1A*, *TUBB5*, *TUBB2B*, *TUBB3* and *WDR62*. Furthermore, copy number variants (CNVs) at several genomic loci have been consistently linked to PMG, including the 1p36.3 locus. The 1p36.3 critical interval for PMG was previously localized to the terminal 4.8Mb of 1p. Using genome-wide array-CGH we recently detected an interstitial 4Mb *de novo* 1p36.3 deletion in a patient with bilateral perisylvian PMG. Importantly, this deletion partially overlaps with the known 1p36.3 critical interval and allowed us to refine this PMG locus to a 2.3Mb region (chr1: 2.5-4.8Mb) encompassing 12 genes. We selected six candidate genes from this region for targeted resequencing, based on gene function and expression in the brain: *ACTRT2*, *AJAP1*, *ARHGEF16*, *TP73*, *TPRG1L* and *WRAP73*. Using molecular inversion probes for targeted capture followed by multiplexed next generation sequencing we are screening 363 patients with PMG, for mutations in these genes. To date we have sequenced 118/363 probands with 90% of the target base pairs sequenced at >25X required for reliable variant calling. We have identified eight non-synonymous variants in four candidate genes (*TP73*, *WRAP73*, *ARHGEF16* and *AJAP1*) that are not present in 6500 control exomes. Segregation analysis of these variants to determine whether these variants arose *de novo* and are therefore likely causative for PMG is underway. This targeted resequencing approach and segregation analysis will be performed for the remaining 245 probands to identify *de novo* mutations in our candidate genes. Furthermore, we are expanding our targeted resequencing approach to include candidate genes from additional CNV loci and exome sequencing studies. In conclusion, we have reduced the size of the 1p36.3 PMG region more than twofold, and present both candidate genes and variants for identification of the causative PMG gene at this locus.

1196W

Application of customized CGH-array for mGluR genes in Attention Deficit/Hyperactivity Disorder (ADHD). *R. Pellegrino, C. Kao, L. Vazquez, D. Hadley, A. Kini, J. Glessner, H. Hakonarson.* CAG, Children s Hospital of Philadelphia, Philadelphia, PA.

The purpose of the project is to perform higher-resolution genotyping by aCGH on subjects with autism and ADHD to determine the prevalence of CNVs in genes within the metabotropic glutamate receptor (mGluR) network. Recently,Center for Applied Genomics(CAG)has found genes in this network were enriched for CNVs in patients with ADHD and autism (data not published). The analysis was based on genotyping on Illumina SNP arrays(550–610K). The CNV analyses revealed that ~10% of individuals with ADHD and ~2–3% of individuals with autism have CNVs within genes in the mGluR network. However, the resolution of the analyses was limited by the spacing of the probes in these SNP arrays (~4–6kb between probes)and potentially important CNVs that were missed due to lack of sensitivity of the arrays. Therefore, we designed a custom array for comparative genomic hybridization (CGH) with dense coverage (~300–400bp between probes) within genes in the mGluR network as well other genes implicated in ADHD and other mental disorders. The ADHD and autism cohort within CHOP that were used in the previous analyses will be systematically screened again with the custom CGH array for mGluR network mutations. Importantly, those subjects with these mGluR network mutations could represent a distinct subset where the etiology of their disorder arises from mGluR dysfunction. This potential discovery could lead to responsive therapeutic with mGluR modulators.

1197T

***FOXG1* is implicated in abnormal brain development in two unrelated cases with 14q aberrations.** *G. Macintyre[1], K. Schlade-Bartusiak[1], M.L. Martínez-Fernández[2], M.L. Martinez-Frías[2], R. Arteaga[3], P. O'Brien[4], M.A. Ferguson-Smith[4], D.W. Cox[1].* 1) Medical Genetics, University of Alberta, Edmonton, Alberta, Canada; 2) Research Centre for Congenital Anomalies (CIAC), CIBER de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain; 3) Neuropediatrics service, Hospital Universitario Marques de Valdecilla, Santander, Spain; 4) Centre for Veterinary Science, University of Cambridge, Cambridge, UK.

We have characterized two cases of *de novo* 14q rearrangements on the proximal long arm of chromosome 14, each associated with abnormal brain development. The first is a *de novo* inversion in a male, who presented at 6 months with microcephaly, hypotonia, epilepsy, developmental and psychomotor delays, and partial agenesis of the corpus callosum. Weight and length were normal. At 11 months, the patient exhibited partial, generalized seizures. At 20 months, developmental delay was evident, weight 10.5Kg (25th centile), length 81cm (25th centile) and occipital frontal circumference (OFC) = 43cm (<3rd centile). Karyotype and FISH analyses revealed a *de novo* paracentric inversion: 46,XY,inv(14)(q12q21.3). The proximal break lies between genes, while the distal break is in a gene-rich area and may cause a gene disruption. Candidate genes for disruption include *KLHDC1*, *KLHDC2*, *SOCCAG1* and *ARF6*. However, lymphoblast gene expression analyses indicated altered expression of several other genes compared to controls, including genes adjacent to the proximal breakpoint at 14q12: *FOXG1*, position unchanged, is upstream of the proximal breakpoint and is up-regulated in patient lymphoblasts, while *ARHGAP5* is displaced to a more distal location on chromosome 14 as a result of the inversion, and is down-regulated. FOXG1, originally known as Brain Factor 1 (BF1), is essential to normal mammalian brain development and *FOXG1* gene deficits have been associated with corpus callosum defects, as well as an atypical Rett syndrome. Of additional interest, is the altered regulation of *ARHGAP5*, encoding RhoGAP5/p190B, as p190B deficits have previously been associated with altered neuronal migration, as well as major forebrain and corpus callosum defects in mice. The second case involves the posthumous CGH characterization of a 4.2Mb deletion within 14q12 (46,XY,del(14)(q12q12)). The patient exhibited cyanosis and hypotonia, with agenesis of the corpus callosum. Several genes were found to be deleted including *NOVA1*, *FOXG1*, and *PRKD1*. Our data from two unrelated cases support published findings from various authors that identify the transcription factor, FOXG1, as a protein essential for the normal development and function of the brain.

1198F

Uncovering Obsessive-Compulsive Disorder risk genes using a high-resolution genome-wide CNV approach. *M.J. Gazzellone[1,2], S.M. Shaheen[3], B. Li[3], R.J. Schachar[3], A.C. Lionel[1,2], B. Thiruvahindrapduram[1], M. Uddin[1], K.D. Fitzgerald[4], C.R. Marshall[1,2], D.R. Rosenberg[5,6], N. Soreni[7], G.L. Hanna[4], P.D. Arnold[3,8], S.W. Scherer[1,2].* 1) The Centre for Applied Genomics and Program in Genetics and Genomic Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 3) Department of Psychiatry and Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI, United States; 5) Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI, United States; 6) The Children's Hospital of Michigan, Detroit, MI, United States; 7) Department of Psychiatry and Behavioural Neurosciences, Faculty of Health Sciences, McMaster University, St. Joseph's Healthcare, Hamilton, Ontario, Canada; 8) Department of Psychiatry and Institute of Medical Science, Toronto, Ontario, Canada.

Background: Obsessive-Compulsive Disorder (OCD) is a neuropsychiatric condition characterized by persistent unwanted thoughts (obsessions), ritualized actions and repetitive behaviors (compulsions), and excessive anxiety. It is common (2% lifetime prevalence) and often presents during childhood. Family and twin studies suggest that genetic factors underlie the pathology of OCD, particularly when the symptoms begin in childhood. To date, genetic investigations have focused primarily on candidate gene, genetic linkage, and genome-wide association approaches, but there have been no published copy number variation (CNV) studies. Findings reported to date explain only a fraction of the genetic architecture of OCD. Our approach seeks to build upon the foundation established in studies of other neuropsychiatric conditions which posit that rare inherited and *de novo* CNVs may elevate disorder risk. We aimed to establish whether this type of variation contributes to the pathogenesis of OCD. **Methods:** We undertook a genome-wide CNV scan using two high-resolution microarrays: the Affymetrix CytoScan HD array and the Illumina OMNI 2.5M Quad array (both featuring around 2.5 million probes). In stage one of the project, DNA was obtained from 100 affected children and their parents and run on the CytoScan HD array. We then conducted a second scan of unrelated probands on either the CytoScan HD array (50 probands) or the OMNI 2.5M Quad array (112 probands). We identified high-confidence CNVs by requiring their identification by two or more CNV detection algorithms. Rare variants were identified by comparing the stringent CNV calls from our case cohorts to stringent calls obtained from population-based controls genotyped on the same array (873 CytoScan HD controls and 2,988 OMNI 2.5M Quad controls). **Results:** Rare CNVs overlapping genes previously implicated in neuropsychiatric disorders, and absent in the controls, were uncovered in the OCD probands. These include exonic deletions or duplications of genes involved in neuronal migration (i.e. *ASTN2*), synaptic function and signal transmission (i.e. *NLGN1* and *NLGN4X*), and postsynaptic scaffolding in glutamatergic synapses (i.e. *DLGAP1* and *DLGAP2*). In addition, we identified new candidate OCD genes involved in synaptic function or plasticity (i.e. *BTBD9*). Our findings suggest that rare copy number changes contribute to OCD risk and that genes expressed at the synapse may play a role in the onset of OCD when perturbed.

1199W

Family structure is predictive of genetic architecture in autism spectrum disorders. *K. Schmitz-Abe[1,2,4,5,6], M. Chahrour[1,2,3,4,5,6], T. Yu[1,2,3,4,5,6,7], S. Hill[1,2,3,4,5,6], G. Sanchez-Schmitz[8,9], J. Partlow[1,2,3,4,5], B. Barry[1,2,3,4,5], B. Mehta[1], S. Servattalab[1], A. Ngoc Lam[1], C. Walsh[1,2,3,4,5,6], K. Markianos[1,2,5].* 1) Division of Genetics, Department of Medicine, Children's Hospital Boston, Boston, Massachusetts, USA; 2) Manton Center for Orphan Disease Research, Children's Hospital Boston, Boston, Massachusetts, USA; 3) Howard Hughes Medical Institute, Children's Hospital Boston, Boston, Massachusetts, USA; 4) The Autism Consortium, Boston, Massachusetts, USA; 5) Departments of Pediatrics and Neurology, Harvard Medical School, Boston, Massachusetts, USA; 6) Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, USA; 7) Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; 8) The Division of Infectious Diseases, Boston Children's Hospital; 9) Harvard Medical School, Harvard University Boston, MA 02115, USA.

Analyses of large autism datasets have provided statistical and functional evidence for the role of rare point mutations and transmitted and de novo copy number variants (CNVs), and offer crucial insights into the diverse genetic mechanisms that can lead to Autism Spectrum Disorders (ASDs). Here we present CNV analysis for a cohort of 183 consanguineous families with one or more children affected with ASD. We provide new insights into the genetic architecture of ASDs as our cohort is uniquely enriched for recessive loss of function variants. We follow up findings and draw comparisons with additional large ASD and control datasets: the Simons Simplex and the Autism Genetic Resource Exchange (AGRE) collections (2,670 affected individuals; 9681 total individuals). Comparing across these cohorts, we demonstrate that ascertainment can lead to selection of different underlying genetic mechanisms causing ASDs. These differences are reflected in metrics such as the affected male:female ratio and the relative contribution of de novo CNVs versus inherited homozygous deletions. Specifically, we find that de novo CNVs play a significant role in non-consanguineous families with a single affected child ($p=0.04$), but a lesser role in multiplex families, and they are no more common in ASD cases than controls in multiplex consanguineous families. In contrast, we present the strongest statistical evidence ($p=0.013$) to date that homozygous deletions, are a major contributor to ASD disease burden in consanguineous families, contributing to as much as 5–10% of cases.

1200T

Susceptibility Loci for Pigmentation and Melanoma In Relation to Parkinson's Disease. *J. Dong[1], J. Gao[1], M. Nalls[2], X. Gao[3], X. Huang[4], J. Han[5], A. Singleton[2], IPDGC. IPDGC[6], H. Chen[1].* 1) NIEHS, durham, NC; 2) National Institute on Aging, Bethesda, Maryland; 3) Harvard School of Public Health, Boston, Massachusetts; 4) Pennsylvania State University-Milton S. Hershey Medical Center, Hershey, Pennsylvania; 5) Channing Laboratory, Harvard Medical School, Boston, Massachusetts; 6) International Parkinson's Disease Genomics Consortium (IPDGC).

Objective: To examine potential associations between genetic susceptibility loci for pigmentation or melanoma and the risk for Parkinson's disease (PD). Methods: We examined 13 SNPs from previous GWAS studies on pigmentation or melanoma in relation to PD risk among 808 PD cases and 1,623 controls in the Parkinson's Genes and Environment (PAGE) Study and then a broader selection of 360 SNPs among 5,333 cases and 12,019 controls from the International Parkinson's disease Genomic Consortium (IPDGC). In the PAGE study, we also examined the colors of hair, eye, or skin and melanoma in relation to PD risk. All participants were self-reported Caucasians. Results: As expected, many of these SNPs were associated with pigmentation phenotypes among PAGE participants. However, neither these SNPs nor pigmentation phenotypes were associated with PD risk after adjusting for multiple comparisons. In PAGE, 18 PD cases (2.2%) and 26 (1.6%) controls also had melanoma, resulting in an odds ratio of 1.3 (95% confidence interval: 0.7-2.4). The IPDGC analysis confirmed that none of the pigmentation or melanoma SNPs was associated with PD risk. Conclusions: Our study did not show any association between pigmentation or melanoma related SNPs and the risk for PD.

1201F

Genetic Risk Factors in Utah Pedigrees at High Risk for Suicide. *H. Coon[1], R. Pimentel[2], K.R. Smith[2,3], C. Huff[4], H. Hu[4], L. Jerominski[1], J. Hansen[1], M. Klein[5], W.B. Callor[6], J. Byrd[6], A. Bakian[1], S. Crowell[1,7], W.M. McMahon[1], E. McGlade[1], D. Yurgelun-Todd[1], P. Renshaw[1], T. Grey[6], D. Gray[1].* 1) Dept Psychiatry, Univ Utah, Salt Lake City, UT; 2) Pedigree & Population Resource, Huntsman Cancer Inst., Univ Utah, Salt Lake City, UT; 3) Dept Family Consumer Studies, College Soc Behav Sci, Univ Utah, Salt Lake City, UT; 4) Dept Epidemiol MD Anderson Cancer Center, Univ Texas, Houston, TX; 5) Core Research Facilities, Health Sci Center, Univ Utah, Salt Lake City, UT; 6) UT State Office Medical Examiner, UT Dept Health, Salt Lake City, UT; 7) Dept Psychology, Univ UT, Sale Lake City, UT.

We used unique population-based data resources to identify 22 high-risk extended pedigrees that show clustering of suicide twice to three times that expected from age-sex-cohort adjusted incidence rates that also incorporate pedigree structure. We have studied genetic risk factors in 5 of these large pedigrees each with 17–51 related suicide decedents, 5-9 of which have previously-collected DNA. These decedents were genotyped with the Illumina HumanExome BeadChip. Genotypes were analyzed using the Variant Annotation, Analysis, and Search (VAAST) program package that computes likelihoods of risk variants using the functional impact of the DNA variation, aggregative scoring of multiple variants across each gene, and pedigree structure. We prioritized variants that were: 1) shared across pedigree members, 2) rare in publicly-available sequence data from 1,358 controls, and 3) screened against 258 other Utah suicides not in the pedigrees to eliminate potential false positives. Sequence variants were prioritized statistically and then implicated genes were screened for previous disease associations and functional relevance. Findings included membrane protein genes and several genes with neuronal involvement and/or known associations with psychiatric conditions. Genes implicated in particular pedigrees may be associated with significant co-morbid psychiatric or medical conditions and/or demographic attributes unique to that pedigree. While the study is limited to variants included on the HumanExome BeadChip, these findings warrant further exploration, and demonstrate the utility of this high-risk pedigree resource.

1202W

Association study of BDNF with completed suicide in the Japanese population. *A. Hishimoto[1,2], W. Ratta-apha[1], K. Shiroiwa[1], I. Sora[1].* 1) Dept. of Psychiatry, Kobe Univ Grad Sch Med, Kobe, Japan; 2) Dept. of Psychiatry and Behavioral Science Albert Einstein College of Medicine, Bronx, NY. USA.

Brain-derived neurotrophic factor (BDNF) is involved in neuronal survival, brain plasticity, and neuronal development. Impairment of BDNF has been implicated in the pathophysiology of psychiatric disorders and suicide. Previous reports have provided evidence for an association between a Val66Met single nucleotide polymorphism (SNP) in BDNF and suicidal behavior among various psychiatric disorders. The present study was conducted to investigate the association between BDNF and subjects who completed suicide. The allelic and genotypic distributions of eight SNPs found in BDNF were determined in 307 subjects who completed suicide and 380 normal subjects using a Taqman probe assay. The genotypic and allelic distributions of the selected SNPs did not differ between the completed suicide and normal subjects. In addition, we performed a meta-analysis of the BDNF Val66Met SNP in suicidal behavior and completed suicide using data from the present study as well as seven previously published studies. These results showed that the Met allele of the BDNF Val66Met SNP tended to associate with suicidal behaviors in the Asian population but not with the completed suicide subjects.

1203T

Rare and common variants near CHRNB3-CHRNA6 are associated with cocaine dependence. *B. Sadler[1, 2], G. Haller[1], N. Saccone[1], L. Bierut[1], A. Goate[1], Study of Addiction, Genetics and Environment (SAGE) Collaborators.* 1) Washington University School of Medicine Department of Psychiatry 660 S. Euclid Ave. Saint Louis, MO 63110; 2) Arizona State University School of Human Evolution and Social Change 900 S. Cady Mall Tempe, AZ 85281.

In the U.S.A., cocaine is the second most frequently abused illicit drug. Although there have not yet been large-scale GWAS for cocaine dependence, but the CHRNA3-A5-B4 nicotinic receptor gene cluster on chromosome 15, identified by GWAS to be associated with nicotine dependence, also harbors variants associated with decreased risk for cocaine dependence. Variants within the CHRNA6-B3 gene cluster on chromosome 8 were found to affect nicotine consumption in several GWAS and thus represent intriguing candidate genes for the study of drug dependence. Using genotype data from a GWAS of the Study of Addiction: Genetics and Environment (SAGE) dataset including 2714 European Americans (605 subjects with non-zero scores on the DSM-IV cocaine symptom count and 2109 subjects with scores of zero on the DSM-IV cocaine symptom count), we tested for association of CHRNA6-B3 SNPs with DSM-IV cocaine dependence symptom count. The top SNP for symptom count was rs9298626 ($\beta = 0.528$, $p=7.0\times10{-}4$), although multiple SNPs passed the multiple test correction. To determine whether there was evidence for multiple independently associated variants at the locus contributing to risk for this phenotype, the most significant SNP was added as a covariate in a linear regression model. We find evidence for more than one independent signal within this locus in European Americans for cocaine dependence symptom count. Further, when using only exposed but not addicted individuals as controls, the signal remained. This signal appears to be related to the previously observed association with nicotine dependence in the region. These data suggest that the CHRNA6-B3 locus contains multiple independent variants affecting risk for cocaine dependence symptom count and other drug addictions.

1204F

De novo deletion of 1q21.1 and Xq28 duplication in a family with developmental delay. *K. Ha, H. Cho, L. Layman, H. Kim.* Georgia Regents University, Augusta, GA.

Distal 1q21.1 microdeletion syndrome is a rare contiguous gene deletion disorder with de novo or autosomal dominant inheritance patterns. MECP2 duplication syndrome is an X-linked neurodevelopmental disorder characterized by intellectual disability, global developmental delay, and other neurologic complications including seizures later in life. Patients with these syndromes share some phenotypes such as mild dysmorphic features, cardiac abnormalities and mental retardation. Since these two syndromes are genetically unrelated, it has not been reported for these two different genetic syndromes to occur in the same family. Here we describe two siblings carrying a chromosome 1q21.1 microdeletion and chromosome Xq28 duplication. Using a comparative genomic hybridization (CGH) array, we identified a 1.24 Mb distal heterozygous deletion at 1q21.1 from position 146.5-147.78 Mb resulting in the loss of 12 genes and a 68 Kb deletion within intron 2 of AUTS2 gene at 7q11.22 in a girl with hypothyroidism, short stature, sensory integration disorder, and soft dysmorphic features including cupped ears and a unilateral ear pit. Protein kinase, AMP-activated, beta 2 non-catalytic subunit (PRKAB2) regulating cellular energy metabolism, and Gap junction alpha-5 protein (GJA5), a gap channel protein, might have contributed the short stature, chronic constipation and sensory integration disorder in this patient. We also characterized a 508 Kb duplication at Xq28 encompassing MECP2 in the younger brother referred for hypotonia, poor speech, cognitive and motor impairment. In addition to MECP2, this duplication encompassed several other genes including L1CAM and FLNA that previously reported in other patients with a Xq28 duplication. The parental CGH analysis revealed that the 1q21.1 deletion in the elder sister is de novo, but the Xq28 duplication in the younger brother was inherited from the mother with a normal phenotype, suggesting a recessive pattern of genes involved in this region. We have described the candidate genes and phenotype-genotype relationships in two different chromosome regions of these two syndromes.

1205W

The first familial null mutation of an autism susceptibility gene in Autism Spectrum Disorder. *N. Lambert[1,3], V. Wermenbol[2], B. Pichon[1], S. Acosta[3], J. van den Ameele[3,4], C. Perazzolo[3], D. Messina[2], F. Musumeci[2], B. Dessars[1], A. De Leener[1], M. Abramowicz[1], C. Vilain[1].* 1) ULB Center of Human Genetics, Hôpital Erasme, ULB, Brussels, Belgium; 2) Department of Paediatric Neurology, Hôpital Erasme, Université Libre de Bruxelles (ULB), Brussels, Belgium; 3) Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire (IRIBHM), Université Libre de Bruxelles (ULB), Brussels, Belgium; 4) Department of Neurology, Ghent University Hospital, Ghent, Belgium.

The autism spectrum disorder results from interactions of genetic and environmental factors. Many genes have been described as candidate gene for autism susceptibility. The gene we studied is a candidate gene for autism susceptibility and is implicated in neurodevelopment and social brain circuit, but no mutation of this gene has ever been described. Here, we describe the first case of a familial mutation of this gene, consisting of an interstitial genomic deletion removing exons 12 through 15, causing a frameshift and premature stop codon, with evidence of nonsense-mediated mRNA decay. The heterozygous mutation was associated with autism in one patient and language and social impairment in a sibling. Our observations delineate the phenotypic spectrum associated with a clearly defined, very likely complete loss of function mutation of this gene. Incomplete penetrance in this family was consistent with this gene as a partial susceptibility gene for ASD. Implication of this gene in normal and pathological brain development opens new perspectives for understanding the pathophysiology of autism and for eventual therapeutical clues.

1206T

Use of a custom-designed array-CGH to identify small CNVs in nonsyndromic Autism Spectrum Disorders. *C.M. Ribeiro[1], E.S. Moreira[1], A.L.B. Martins[2], A.G. Morales[1], S.G. Ferreira[1], V.N. Takahashi[1], D.P. Moreira[1], K. Griesi-Oliveira[1], A.C.F. Conte[3], C. Rosenberg[1], E. Vadasz[4], M.R. Passos-Bueno[1].* 1) Departamento de Genetica e Biologia Evolutiva, Instituto de Biociencias, University of Sao Paulo, Sao Paulo, Brazil; 2) UNESP, Sao Jose do Rio Preto, Sao Paulo, Brazil; 3) Laboratorio de Genetica, Hospital de Base, Faculdade de Medicina de Sao Jose do Rio Preto, Sao Paulo, Brazil; 4) Instituto de Psiquiatria, Hospital das Clinicas, Faculdade de Medicina, University of Sao Paulo, Sao Paulo, Brazil.

Autism Spectrum Disorders (ASD) are highly heterogeneous and, despite the enormous efforts to identify their underlying genetic causes, at least 70% of the cases are idiopathic. We have hypothesized that a proportion of these cases might be caused by small CNVs not detected through conventional microarray platforms or NGS. To address this hypothesis, we designed an array-CGH with high density of probes targeted to the coding exons and to the UTR of 269 genes involved in biological processes potentially significant to the pathophysiology of ASD. A total of 508 ASD Brazilian patients and 550 Brazilian healthy individuals were included in the study after a signed informed consent. The IB-USP Ethics Committee approved this research. All patients were diagnosed by psychiatrists using DSM-IV, ICD-10 and ASQ, and all were negative for CNVs at 15q11-13, 16p11 and 22q13 through MLPA. Exclusion criteria comprised autism-related syndromes, metabolic disorders, facial dysmorphisms, congenital malformations, or maternal exposition to known teratogenic drugs/ infectious agents during pregnancy. The number of individuals harboring at least one CNV and the number of CNVs per subject were significantly different between 103 cases and 200 controls ($p<0.0001$). These findings also applied to the separate analyses of the microdeletions and microduplications ($p\leq0.0002$), with odds ratios related to the number of alterations per individual between 2.92 and 3.32. Considering the subjects with two or more alterations, the differences between cases and controls were even greater (for microdeletions, $p=0.000005$; OR=28.74 [95% CI 3.7-223.1]). Forty-five out of 103 patients (43.7%) presented at least one CNV smaller than 1 kb which disrupted one of the genes in our customized array, a rate significantly higher than the proportion (21.5%) found in controls ($p=0.00009$). Finally, we detected 10 ASD-specific CNVs (five in the UTR) in 11.6% of the patients. These results are different from those usually obtained through conventional screenings with genome-wide platforms and indicate that the use of this high-resolution custom-designed array can increase the chance of detecting alterations enriched in ASD. This study also suggests that disruption of non-coding DNA may contribute to the risk factors for ASD, corroborates the association of *SHANK2*, *DIAPH3*, *GCH1*, *GRM5* and *MARK1* with nonsyndromic ASD, and reveals *SNAP29*, *SLC17A6*, *PRKC*α, *MBD2* and *GAD2* as new candidates for these disorders.

1207F

Linkage Disequilibrium Mapping of the 13q21 Specific Language Impairment Locus Using Epistasis Analysis Models. *S.L. Wolock[1], N. Li[1], S.A. Petrill[2], J.F. Flax[3], A.S. Bassett[4], L.M. Brzustowicz[3], C.W. Bartlett[1,5].* 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Department of Psychology, The Ohio State University, Columbus, OH; 3) Department of Genetics, Rutgers University, Piscataway, NJ; 4) Department of Psychiatry, University of Toronto, Toronto, Canada; 5) Department of Pediatrics, The Ohio State University, Columbus, OH.

Background: We present a fine-mapping study of the 13q21 region linked to Specific Language Impairment (SLI), a neurodevelopmental failure to develop normal vocabulary and grammar despite otherwise normal cognition and ability. In previous work, we mapped the SLI3 locus to 13q21 using a written language impairment phenotype using five extended families from Canada. We replicated the locus using nuclear and extended families from the United States. In refining the localization, we found that a coding SNP in BDNF associated with memory greatly increased localization and evidence for the locus when included as part of a gene-gene interaction with the unidentified 13q21 risk alleles. Here, we performed additional mapping of the region to identify the gene responsible for the linkage signals. Methods: We assessed linkage/association of Illumina SNP array genotypes with eight phenotypes using the posterior probability of linkage (PPL) and the posterior probability of linkage disequilibrium (PPLD) metrics. We conducted two analyses: a baseline analysis and an analysis that incorporated BDNF genotypes from a coding SNP associated with memory into a gene-gene (epistatic) interaction model. The posterior probabilities from the two models are on the same scale and can be directly compared to assess the evidence that a SNP has an epistatic interaction. Results: Several SNPs within an LD block showed association with the categorical written language impairment diagnosis used to map this locus in the original genome scans (maximum PPLD = 41%). We also observed weaker association with the quantitative trait underlying the written language impairment trait (maximum PPLD = 17%). In both cases, modeling the coding SNP in BDNF showed an epistatic effect. The implicated region overlaps ATXN8OS, an anti-sense transcript of the KLHL1 gene. ATXN8OS contains a tri-nucleotide repeat that, when expanded, may increase the risk for a form of spinocerebellar ataxia. It has additionally been associated with several neuropsychiatric phenotypes. Conclusion: While additional work to elucidate the role of this antisense transcript is ongoing, the findings suggest a role for BDNF-ATXN8OS interaction in the etiology of SLI.

1208W

Evidence for the involvement of MIR185 and its target genes in the development of schizophrenia. *AJ. Forstner[1,2,19], FB. Basmanav[1,2,19], M. Mattheisen[3,4], AC. Böhmer[1,2], MV. Hollegaard[5], E. Janson[6], E. Strengman[6], L. Priebe[1,2], F. Degenhardt[1,2], P. Hoffmann[1,2,7], S. Herms[1,2,7], W. Maier[8], R. Mössner[8], D. Rujescu[9], RA. Ophoff[10,11], S. Moebus[12], PB. Mortensen[13], AD. Børglum[14,15], DM. Hougaard[5], M. Rietschel[16], A. Zimmer[17], MM. Nöthen[1,2], X. Miro[17], S. Cichon[1,2,7,18].* 1) Institute of Human Genetics, University of Bonn, Germany; 2) Department of Genomics, Life and Brain Center, Bonn, Germany; 3) Department of Biomedicine, University of Aarhus, Denmark; 4) Department of Genomics Mathematics, University of Bonn, Germany; 5) Section of Neonatal Screening and Hormones, Statens Serum Institute, Copenhagen, Denmark; 6) Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, The Netherlands; 7) Division of Medical Genetics, University Hospital Basel and Department of Biomedicine, University of Basel, Switzerland; 8) Department of Psychiatry, University of Bonn, Germany; 9) Department of Psychiatry, University of Halle-Wittenberg, Halle, Germany; 10) Center for Neurobehavioral Genetics, University of California Los Angeles, USA; 11) Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Center, Utrecht, The Netherlands; 12) Institute of Medical Informatics, Biometry and Epidemiology, Essen, Germany; 13) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; 14) Department of Biomedicine, Human Genetics, and Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark; 15) Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark; 16) Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University, Germany; 17) Institute of Molecular Psychiatry, University of Bonn, Bonn, Germany; 18) Institute of Neuroscience and Medicine INM-1, Research Center Juelich, Germany; 19) These authors contributed equally to this work.

Background: The 22q11.2 deletion syndrome (22q11.2DS) is a phenotypically complex syndrome caused by a hemizygous microdeletion in chromosomal region 22q11.2. It occurs approximately 1 in 2,000 - 4,000 births and about 30% of carriers develop schizophrenia, making this syndrome one of the strongest known genetic risk factors for schizophrenia. Most of the deletions at 22q11.2 are either 1.5 or 3 megabases in size spanning 35 and 60 known genes, respectively. Research has yet to confirm which genes within the deletion region are of importance in terms of this schizophrenia-risk effect. The minimal 1.5 megabase deletion region contains the gene MIR185, which encodes microRNA 185. This has two validated targets (RhoA, Cdc42), both of which have been associated with altered expression levels in schizophrenia. **Methods:** The role of MIR185 in schizophrenia was investigated by: 1) monitoring miR-185 expression in embryonic and adult mouse brains; and 2) investigating the role of common and rare variants at this locus in humans. The latter approach involved three steps. Firstly, gene-based analyses were performed for common variants in MIR185 and its target genes using Schizophrenia Psychiatric Genomics Consortium genome wide association data. Secondly, the MIR185 gene was resequenced in 1,000 schizophrenia patients and 500 controls of German origin to investigate the role of rare variants. Thirdly, promising variants were followed up by genotyping an additional 3,530 patients and 4,018 controls. **Results:** In situ hybridization in mice revealed miR-185 expression in brain regions implicated in schizophrenia and other neuropsychiatric manifestations of 22q11.2DS. Gene-based tests revealed no association between schizophrenia and MIR185 at either the gene-level or the level of individual SNPs. However, three target genes were associated with schizophrenia after correction for multiple testing. Resequencing identified two rare patient-specific novel variants directly flanking MIR185. However, follow-up genotyping provided no further evidence for their involvement in schizophrenia. **Conclusions:** The miR-185 expression patterns in mice and the genetic association results for the three miR-185 target genes suggest that miR-185 and its down-stream pathways may be implicated in the development of schizophrenia in both 22q11.2DS patients and idiopathic cases. MicroRNA-mediated dysregulation is therefore a possible etiological mechanism in schizophrenia.

1209T

Identification of novel mutations in *SLC20A2* and *PDGFRB* in a Brain Bank series of IBGC. *M. Sanchez-Contreras[1], M. Baker[1], N. Finch[1], A. Nicholson[1], A. Wojtas[1], A.J. Strongosky[2], D.B. Calne[3], S. Calne[3], A. Jon Stoessl[3], J.E. Allanson[4], D.F. Broderick[5], M. Hutton[1], D.W. Dickson[1], O.A. Ross[1], Z.K. Wszolek[2], R. Rademakers[1].* 1) Department of Neuroscience, Mayo Clinic, Jacksonville, FL; 2) Department of Neurology, Mayo Clinic, Jacksonville, FL; 3) Pacific Parkinson's Research Center and National Parkinson Foundation Center of Excellence, Neurology, University of British Columbia, Vancouver, B.C., Canada; 4) Eastern Ontario Regional Genetics Program, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 5) Department of Radiology, Mayo Clinic, Jacksonville, FL.

Idiopathic Basal Ganglia Calcification (IBGC) is characterized by bilateral calcification of the basal ganglia and presents clinically with a spectrum of neuropsychiatric and motor syndromes including dystonia, parkinsonism, tremor and chorea. Mutations in 2 genes, *SLC20A2* and *PDGFRB*, have recently been shown to cause familial forms of IBGC. We set out to identify and study mutations in *SLC20A2* and *PDGFRB* in a series of pathologically confirmed IBGC cases from the Mayo Clinic Florida Brain Bank. Methods: 27 cases were included based on the presence of significant basal ganglia calcification upon pathological examination. All coding regions of *SLC20A2* and *PDGFRB* were sequenced. For *SLC20A2*, copy number analysis was also performed using real-time quantitative PCR. Confirmatory assays for a genomic deletion in *SLC20A2* were performed by PCR based on the coordinates obtained from whole-genome sequencing (WGS) in one patient. For *PDGRFB*, we generated wild type and mutant forms and performed transient expression in HeLa cells to evaluate its tyrosine kinase function. Upon stimulation with PDGF-BB, the autophosphorylation of wild-type and mutant PDGFRB at multiple tyrosine residues was assessed by western blot. Results: In our IBGC brain bank series, we identified two novel coding variants in *SLC20A2*. One novel variant was predicted to induce a nonsense mutation in exon 3 (p.S113X) and the second mutation consisted of a 563,256 bp genomic deletion with precise breakpoints defined by WGS. This deletion co-segregated with the disease in a large Canadian family and affected multiple genes. The combined deletion of *SLC20A2* and *THAP1* in this family likely contributed to the early onset and predominant dystonia. Additionally, we identified one novel coding variant in exon 15 of *PDGFRB* predicted to induce a missense mutation in the kinase domain (p.R695C). Our functional studies indicate that *PDGRFB* mutations disrupt the autophosphorylation of PDGFRB. Conclusion: Genetic mutations in *SLC20A2* and *PDGFRB* were found in 3 cases of our IBGC series. The identification of a novel mutation and a complete deletion of *SLC20A2* confirm an important role for *SLC20A2* haploinsufficiency in IBGC. Our work further supports a pathogenic role for PDGFRB in IBGC. Moreover, abnormal signaling through PDGFRB and deficient phosphate transport through SLC20A2 may be involved in the same pathogenic pathway leading to calcification in IBGC.

1210F

Extreme genetic heterogeneity among rare copy number variation in autism spectrum disorders. *M. Barbosa[1], A.G. Chiocchetti[2], D. Haslinger[2], R. Waltes[2], E. Duketis[2], T. Jarczok[2], M. Sachse[2], F. Poustka[2], A. Voran[3], E. Huy[3], F. Degenhardt[4], S. Herms[4,5], P. Hoffmann[4,5], J. Meyer[6], M.M. Biscaldi-Schäfer[7], K. Schneider-Momm[7], B.G. Schimmelmann[8], S. Kupferschmid[8], B. Herpertz-Dahlmann[9], S. Gilsbach[9], S. Moebus[10], S. Cichon[4,5], S.M. Klauck[11], D. Pinto[1], C.M. Freitag[2].* 1) Department of Psychiatry and Genetics and Genomic Sciences, Seaver Autism Center, Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, NY, USA. Address: 1470 Madison Avenue, 10029, NY, USA; 2) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe-Universität Frankfurt am Main. Address: Deutschorden-Str. 50, 60528, Frankfurt, Germany; 3) Department of Child and Adolescent Psychiatry and Psychotherapy, Saarland University, Homburg. Address: Kirrberger-Str Building 90.2, 66421 Homburg / Saar, Germany; 4) Institute of Human Genetics, Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. Address: Sigmund-Freud-Str. 25, 53127, Bonn, Germany; 5) Division of Medical Genetics, University Hospital and Department of Biomedicine, University of Basel, Switzerland. Address: Mattenstr. 28, 4058 Basel, Switzerland; 6) Department of Neurobehavioral Genetics, University of Trier, Germany, Address: Johanniterufer 15, 54290 Trier, Germany; 7) Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital Freiburg, Address: Hauptstr. 8, 79104 Freiburg, Germany; 8) University Hospital of Child and Adolescent Psychiatry, Bern; Address: Bolligenstrasse 111, 3000 Bern, Switzerland; 9) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital RWTH Aachen, Address: Neuenhofer Weg 21 in 52074 Aachen, Germany; 10) Heinz-Nixdorf-Recall Study, University Duisburg-Essen, Germany, Address: Hufeland-Str. 55, 45122 Essen, Germany; 11) Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany. Address: Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.

Autism spectrum disorders (ASD) are a group of complex neurodevelopmental disorders characterized by impairment in social interaction, language and communication deficits and repetitive behavior. The etiology for ASD has only been identified in 10-20% of cases screened so far. The aim of this study is to identify copy number variants (CNVs) with a high pathological potential for ASD and to clarify the complex genetic heterogeneity of the disorder in a well-characterized ASD cohort. A cohort of 716 German patients with ASD and their relatives (in a total of 2114 subjects) were assessed using the ADI-R and ADOS diagnostic tools. The control group is composed of 1320 German subjects ascertained through the Heinz Nixdorf Recall Study. Both cohorts have been genotyped on the Illumina Human OmniExpress 730K-SNP-array. For copy number variation (CNVs) analysis, we used 4 calling algorithms: PennCNV, QuantiSNP, GNOSIS, CNVPartition. A total of 1967 ASD-samples passed stringent quality control, 673 being probands (572 males and 101 females). Of these, 394 were from trios, 42 from quads and 20 from large families. Using also data from an additional 5000 European controls, priority was given for rare CNVs (*de novo* or inherited). A variety of *de novo* CNVs was detected in cases, from large aberrations (e.g. a novel 13q14.2 5Mb-deletion) to rare pathogenic CNVs (e.g. a Potocki-Lupski 17q12-duplication-syndrome). Interestingly, inherited CNVs along with second genomic hits were seen in four subjects (e.g. a maternal *NRXN1* intragenic deletion and *de novo* 22q11.2-duplication), giving support to a multiple-hit hypothesis. Also, we found CNVs that are often related to a remarkably variable neurobehavioral phenotype such as the 3q29-deletion syndrome, which can present as intellectual disability, ASD, psychosis, anxiety, hyperactivity and/or aggressiveness. Further experimental validation and network modelling of genes intersected or disrupted by CNVs are ongoing.

1211W

Different homozygous mutations in *Italic Text*PARKIN*Italic Text* gene are responsible for early onset Parkinson's disease in Arab-Israeli families that live in close vicinity. *S. Ben-Shachar[1], Z. Afawi[2], R. Masalha[3], S. Badarny[4,5], T. Neiman[1], A. Bar-Shira[1], A. Orr Urtreger[1,6].* 1) Genetic Institute, Tel Aviv Sourasky Med Ctr, Tel-Aviv, Israel; 2) Center for Neuroscience, Ben-Gurion University of the Negev, Be'er Sheva, Israel; 3) Department of Neurology, Soroka Medical Center, Ben Gurion University of the Negev, Be'er Sheva, Israel; 4) 4.Department of Neurology, Carmel Medical Center, Haifa, Israel affiliated to Rappaport Faculty of Medicine -Technion, Haifa, Israel; 5) Rappaport Faculty of Medicine -Technion, Haifa, Israel; 6) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Bold Text Introduction:**Bold Text** Parkinson's disease (PD) affects about 2% of the population older than 65 years. Multiple genetic, environmental and age-related factors influence the risk for PD. Mutations in the *Italic Text*PARKIN, PINK1, DJ1*Italic Text*, and *Italic Text*ATP13A2*Italic Text* genes cause many of the less common autosomal recessive PD (ARPD). A founder mutation in *Italic Text*PARKIN*Italic Text* gene, c.101delA, has been previously detected in Arab-Israeli families with ARPD. We aimed to detect the genetic etiology for this disease in consanguineous families from this origin. **Bold Text**Methods:**Bold Text** Six consanguineous families, with a positive family history of PD and at least two affected members, concordant with ARPD, were recruited. The known *Italic Text*PARKIN*Italic Text* founder mutation was screened, followed by homozygous mapping using SNP based array. Analysis was completed by Sanger sequencing of *Italic Text*PARKIN*Italic Text* coding region. **Bold Text**Results:**Bold Text** In only one family, the known *Italic Text*PARKIN*Italic Text* c.101delA founder mutation was detected in an homozygous state. In two of the other five families, loss of heterozygocity (LOH) was detected at the *Italic Text*PARKIN*Italic Text* locus. A novel homozygous mutation, c.C996G>A (p.C332X), was detected in one family, segregating with the phenotype. In another family a homozygous deletion of exon 4 was detected. Interestingly, this homozygous deletion was previously described in a few families with early onset ARPD in Japan. Full sequence of *Italic Text*PARKIN*Italic Text* coding region in the families that did not demonstrate LOH at this locus did not detect homozygous/compound heterozygote mutations. We also detected LOH of *Italic Text*PINK1*Italic Text* gene in two families: in the family that carry the *Italic Text*PARKIN*Italic Text* exon 4 deletion, and in another family that demonstrated additional LOH in the *Italic Text*ATP13A2*Italic Text* gene locus. None of the six families showed LOH in the *Italic Text*DJ1 locus. **Bold Text**Conclusions:-**Bold Text** Three different homozygous mutations in *Italic Text*PARKIN*Italic Text* were detected in 3/6 of consanguineous Arab-Israeli families with PD. Of note that the different homozygous alleles were found in families that are living in a same geographic region. These results further demonstrate the high mutation rate in the *Italic Text*PARKIN*Italic Text* gene. Other, yet unknown, genes likely to be associated with ARPD in this population.

1212T

Transcriptome outlier analysis implicates schizophrenia candidate genes as harboring rare functional variants of large effects. *J. Duan[1], W. Moy[1], J. Freda[1], E.I. Drigalenko[2], H.H.H. Göring[2], A.R. Sanders[1], P.V. Gejman[1], Molecular Genetics of Schizophrenia (MGS) Collaboration.* 1) Dept Psychiatry, Northshore Univ Healthsystem/Univ of Chicago, Evanston, IL; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

Multiple rare and large copy number variants (CNVs) are associated with high risk for schizophrenia (SZ), and some common risk variants have been uncovered by GWAS. However, a substantial part of the genetic risk still remains unexplained and disease mechanisms are largely unknown. To uncover additional loci of large effect, we propose to sequence targets derived from transcriptome analyses. Most gene expression studies rely on the study of average abundances in case-control samples. Although useful, this approach assumes substantial etiologic homogeneity, and may miss rarer genetic effects. We have employed an alternative approach to identify expression outliers (i.e., expression distribution extremes) where the tails of the distribution are enriched for cases. We hypothesize that such case-enriched expression outliers are caused by rare coding variants (e.g., nonsense or splice site mutations) or regulatory mutations of large effect on gene expression. We analyzed an RNAseq dataset comprised of lymphoblastoid cell lines from 312 SZ cases and 322 controls, none of which carry a known SZ-associated CNV. We calculated the Z-scores of the expression for each gene for the whole sample. Expression outliers were defined as genes with abundances beyond a predefined standard deviation cutoff (2SD or 3SD). We identified 828 expression outlier genes with 2SD-tails (401 lower and 427 upper) enriched for SZ cases. In these genes we observed enrichment of brain-expressed genes, SZ-risk CNV-spanned genes, and genes within CNVs associated with neurodevelopmental disorders. Noteworthy outlier genes: *KCTD13* that drives the 16p11.2 CNV phenotype, *DGCR8* at 22q11.21 deletion region, and *TCF3* belonging to the same gene family as the SZ-associated *TCF4*. We did not find global outlier burden differences in SZ cases and controls (i.e., total number of outlier genes/individual). However, SZ cases showed higher outlier burden for genes spanned by the above SZ-risk CNVs and a trend towards higher outlier burden for genes within CNVs associated with neurodevelopmental disorders. These results suggest that genes identified by outlier expression analyses are relevant to SZ pathogenesis. However, given the moderate sample size and multiple statistical tests, we consider our results exploratory. We are currently sequencing exons and regulatory sequences of top ranking outlier genes and will report the functional variants contributing to aberrant mRNA expression in SZ cases.

1213F

Molecular genetic study on a Japanese family with amyotrophic lateral sclerosis. *Y. Ichikawa[1], H. Ishiura[1], J. Mitsui[1], S. Endo[1], Y. Takahashi[2], A. Ishii[3], M. Watanabe[3], A. Tamaoka[3], M. Murata[2], K. Doi[4], J. Yoshimura[4], S. Morishita[4], J. Goto[1], S. Tsuji[1].* 1) Dept Neurology, Univ of Tokyo, Tokyo, Japan; 2) Dept Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 3) Dept Neurology, Univ of Tsukuba, Ibaraki, Japan; 4) Dept Computational Biology, Univ of Tokyo, Tokyo, Japan.

Background: Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder characterized by degeneration of motor neurons in the brain and spinal cord. Although the majority of ALS patients are sporadic, approximately 10% of ALS cases are familial (FALS). In the Japanese population, mutations of *SOD1*, *FUS*, *TARDBP*, and *C9orf72* account for approximately 60% of the cases of FALS, and the causative genes remain to be elucidated in 40 % of the cases of FALS. **Objective:** This study is aimed to search for a causative gene for a Japanese autosomal dominant FALS family. **Subjects and Methods:** Three patients were affected in this family. The age of onset of one patient was around 60 and those of the other two patients were in their 70s. Concerning the clinical presentations, one was progressive bulbar palsy; the others were Aran-Duchenne type. Genomic DNAs were extracted from nine individuals including two affected patients. Genotyping was carried out using Affymetrix SNP Array 6.0 and linkage analysis was performed using allegro and SNP HiTLink. To explore the causative mutations, we conducted an exome analysis of the index patient. **Results:** The linkage analysis revealed the highest multipoint parametric LOD score of 1.8 in 13 regions on chromosomes 1, 4, 5, 6, 8, 9, 10, 11, 12, and 14 spanning 189.5Mb in total. Although, the loci of the causative genes for dominant FALS including *SETX*, *ANG*, and *TARDBP* were included in the candidate region, the mutations of these genes were not detected by the exome analysis. **Conclusion:** The results of the analyses suggest that ALS in the present family is caused by a mutation of a novel gene.

1214W

Association between serotonin transporter variant, 5-HTTLPR, and hoarding traits in a population-based pediatric sample. *V. Sinopoli[1], A. Dupuis[2], C. Burton[1,2], J. Crosbie[1,2], J. Shan[2], R. Schachar[1,2], P. Arnold[1,2].* 1) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 2) The Hospital for Sick Children, Toronto, ON, Canada.

Obsessive-compulsive disorder (OCD) is an anxiety disorder characterized by recurring, intrusive obsessions, and compulsive behaviors aimed at alleviating the resulting anxiety. It is a heterogeneous disorder that is both genetically and phenotypically complex, which complicates our ability to identify the heritable components of OCD and the genes that are responsible. At the phenotypic level, studies consistently reveal specific symptom dimensions within the disorder. At the genetic level, the promoter region of the serotonin transporter gene, 5-HTTLPR, has been implicated in OCD, but findings have been inconsistent. Statement of Purpose: The aim of this study is to examine a functionally triallelic variant in 5-HTTLPR and its genetic association with obsessive-compulsive (OC) traits, as a whole, and specific subgroups of OC traits in a population-based sample of children. Methods: We have identified OC traits in 7,500 Caucasian children, from the Ontario Science Centre in Toronto, and have selected the top and bottom 10% of our sample for comparison. Following collection and extraction of DNA, we directly genotyped the 5-HTTLPR polymorphism and classified variants based on functionality. We examined the association between 1) genetic presentation of the variant and high versus low OC traits and 2) genetic presentation of the variant and high versus low OC traits within each factor-based subgroup we have identified. Results: No significant difference is seen between the high OC trait group and the low OC trait group, as a whole. A significant difference in allelic and genotypic distribution is seen between individuals with high and low OC trait extremes specifically in the hoarding factor group. The LA/LA genotype occurs more often with low extreme hoarding traits (P-value 0.004). Likewise, the LA allele occurs more often with low extreme hoarding traits, compared to any other allele (P-value 0.005). Conclusion: Our findings reveal a significant difference in genotypic and allelic distribution only between individuals with high versus low extreme hoarding scores. This illustrates a need to consider specific symptom dimensions within OCD, when examining serotonergic and other genetic variants. By reducing phenotypic heterogeneity, our approach promises to more accurately facilitate identification of genetic risk factors in OCD.

1215T

Disruption of a large intergenic noncoding RNA, *FLJ42709*, is associated with developmental delay and language disorder. *S. Fan[1], L. Dukes-Rimsky[1], S. Ladd[1], B.R. DuPont[1], C. Skinner[1], L. Wang[2], K.B. Clarkson[1], A.K. Srivastava[1, 2].* 1) J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC.

Large intergenic noncoding (linc) RNAs are emerging as key regulators of diverse cellular processes including neurodevelopment. However, the function of the majority of these lincRNAs remains uncharacterized. We identified a paternally inherited balanced translocation t(5;12)(q15;q15) in a 2 year and 7-month-old male patient with developmental delay, speech delay, significant expressive language delay, and congenital infantile left eye esotropia. Fluorescence *in situ* hybridization and whole genome sequencing defined the chromosomal breakpoints and revealed the disruption of a previously uncharacterized lincRNA gene, *FLJ42709* at 5q15. The breakpoint at 12q15 disrupted no gene. Further studies by PCR amplification and sequencing of the breakpoint junctions revealed the identical breakpoints in the patient's father who had dyslexia and stutters. We analyzed quantitative expression of several genes flanking each chromosomal breakpoint and found no change in their expression levels than control individuals. In addition, we looked for disease causing variants other than the translocation breaks using the Ingenuity Variant Analysis tool (Ingenuity Systems Inc, CA). No potential disease causing variant was found. Thus, we further characterized *FLJ42709* and confirmed expression of at least three major transcribed isoforms of *FLJ42709*. We further determined that all three isoforms were expressed abundantly in human fetal brain and in most other tissues tested. To infer the physiological function of *FLJ42709*, we looked for genes that are co-expressed with *FLJ42709*. On the basis of probe-to-probe correlation coefficient calculated from an integrated set of 2,968 microarray expression profiles of healthy human tissue samples, we found that the expression of *FLJ42709* was largely correlated with genes enriched with Gene Ontology terms including synaptic transmission, cell-cell signaling, neuron projection, and synapse. Recently, several genomic deletions at 5q15 encompassing *FLJ42709*, *NR2F1*, *FAM172A*, *POU5F2*, and *MIR2277* genes have been reported in patients with developmental delay, hypotonia, strabismus/estropia, deafness and significant delay in gross motor skills, cognitive and communication skills. All together, our findings raised the possibility of a likely role for the lincRNA gene *FLJ42709* in neurodevelopmental disorders.

1216F

Gene-based association analysis of Alzheimer's disease risk gene CD33 with brain amyloid-β burden and microglial activity measured by PET. *K.T. Nho[1,2], S. Shanker[1], S. Kim[1,2,3], S.L. Risacher[1,3], V.K. Ramanan[1,10], L. Shen[1,2,3], K.K. Yoder[1,3], P.S. Aisen[4], R.C. Petersen[5], M.W. Weiner[6,7], B.C. McDonald[1,3,8], E.F. Tallman[1,3], G.D. Hutchins[1,3], J.W. Fletcher[1], M.R. Farlow[3,8], B. Ghetti[2,3,9], A.J. Saykin[1,2,8,10] for the Alzheimer's Disease Neuroimaging Initiative (ADNI).* 1) Radiology and Imaging Sciences, Indiana University, Indianapolis, IN; 2) Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; 3) Indiana Alzheimer Disease Center, Indianapolis, IN; 4) Department of Neurosciences, University of California, San Diego, San Diego, CA; 5) Mayo Clinic, Rochester, MN; 6) Departments of Radiology, Medicine and Psychiatry, University of California, San Francisco, San Francisco, CA; 7) Department of Veterans Affairs Medical Center, San Francisco, CA; 8) Department of Neurology, Indiana University School of Medicine, Indianapolis, IN; 9) Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN; 10) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN.

Background: Large-scale GWAS have identified and confirmed *CD33* significantly associated with late onset Alzheimer's disease (LOAD) (Naj et al., 2011). A recent study showed *CD33* is expressed in microglial cells in the human brain and inhibits microglial uptake of amyloid-β (Griciuc et al., 2013). Our aim was to perform gene-based association analyses to investigate the relationship of *CD33* variants with amyloid-β (Aβ) burden measured using florbetapir positron emission tomography (PET) imaging (Ramanan et al., 2013) (Analysis 1) and microglial activity measured using [^{11}C]PBR28 PET imaging (Yoder et al., in press) (Analysis 2). **Methods:** 550 participants in the ADNI (Alzheimer's Disease Neuroimaging Initiative) cohort with florbetapir PET scans were used in Analysis 1, and 26 participants in the IMAS (Indiana Memory and Aging Study) cohort with [^{11}C]PBR28 PET scans were used in Analysis 2. The endophenotypes used in Analysis 1 were mean regional standardized uptake value ratio (SUVR) values from 5 regions adjusted for age, gender, diagnosis, *APOE* ε4 status, and *BCHE* genotypes (Ramanan, et al. 2013) and in Analysis 2 were mean PBR SUV from 8 regions adjusted for *TSPO* genotype. SNPs within ±10 kb of the *CD33* gene after imputation using MACH were included. Gene-based association analyses were performed using a dominant genetic model and SNPs were considered significant if they had a $p<0.05$ on linear regression analyses and independent if $r^2 \leq 0.5$. Permutation was used to correct for multiple SNPs. **Results:** From our analyses, we found that the *CD33* gene was significantly associated with Aβ SUVR ($p<0.0044$) and marginally PBR SUV ($p<0.072$) in many regions. In Analysis 1, rs12971800 located at the 3' end of the gene was identified with the minor allele associated with higher Aβ SUVR and accounted for ~2% of the phenotypic variation. In Analysis 2, another SNP (rs273645) located at the 3' end of the gene was identified with the minor allele associated with lower PBR SUV. **Conclusions:** *CD33* was associated with PET imaging measures associated with AD, suggesting *CD33* apprears to contribute to known neuropathologic features of AD such as amyloid burden and inflammation. This is relevant in light of the previous findings and warrants further investigation in independent and larger samples.

1217W

Association of ADH and ALDH variants with alcohol-related phenotypes in a Native American community sample. *Q. Peng[1], I.R. Gizer[2], O. Libiger[1], C. Bizon[3], K.C. Wilhelmsen[4], N.J. Schork[1,5], C.L. Ehlers[6].* 1) Scripps Translational Science Institute, La Jolla, CA; 2) Department of Psychological Sciences, University of Missouri, Columbia, MO; 3) RENCI, University of North Carolina, Chapel Hill, NC; 4) Department of Neurology and Genetics, University of North Carolina, Chapel Hill, CA; 5) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; 6) Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, La Jolla, CA.

Previous studies have shown that alcohol dehydrogenase (ADH) genes and aldehyde dehydrogenase (ALDH) genes may affect the risk of development of alcohol dependence, and that polymorphisms within these genes may differentially affect risk depending on the ethnic group evaluated. We considered variations in the ADH and ALDH genes in a large study investigating risk factors for substance use in a Native American (NA) population. We ultimately tested for associations between alcohol dependence, severe alcohol use and alcohol withdrawal symptoms in the genomic regions around the ADH and ALDH genes. Seven hundred and eight (708) NAs, from extended pedigrees, participated in the study. Blood derived DNA was sequenced using Illumina low-coverage whole genome sequencing (WGS), as well as genotyped using an Affymetrix Exome1A chip. All subjects were assessed using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) in order to make DSM-IV diagnoses. To control for admixture and familial relatedness a variance component approach that incorporated ancestry admixture estimates was used in the association tests. In additional analyses we examined the ancestral backgrounds of the genomic regions under investigation and their influence on the phenotypes. False discovery rates were used to set significant p values. Seventy-seven (77) ADH variants showed significant evidence of association with a severe alcohol use phenotype. These significant variants spanned across the entire 7 ADH gene cluster region on chromosome 4q, however, the majority of them resided in ADH4-ADH6-ADH1A-ADH1B, in both genic and inter-genic regions. Four significant associations (three in ADH and one in ALDH2) were also observed with alcohol dependence symptoms for variants. These studies suggest that WGS combined with ancestry and admixture analyses can identify significant variants associated with alcohol dependence phenotypes in the regions of the major alcohol metabolizing enzymes. (Supported by AA10201, DA030976...).

1218T

Characterization and Replication of Chromosome 15q25 Nicotine Dependence Signals in an Old Order Amish population. *J.D. Backman[1], A.D. Million-Mrkva[1], L.E. Hong[3], M.A. Ehringer[4], B.D. Mitchell[2], A.R. Shuldiner[2], K.A. Ryan[2], K.L. Nugent[3], S.H. Stephens[2].* 1) Epidemiology and Human Genetics, University of Maryland at Baltimore, Baltimore, MD; 2) Endocrinology, Diabetes, and Nutrition; University of Maryland School of Medicine, Baltimore, MD; 3) Psychiatry, University of Maryland School of Medicine, Baltimore, MD; 4) Institute for Behavioral Genetics, University of Colorado, Boulder, CO.

The *CHRNA5/CHRNA3/CHRNB4* locus on chromosome 15q25 encoding neuronal nicotinic acetylcholine receptor genes has been associated with nicotine dependence and smoking behaviors in multiple populations. This region includes three statistically distinct loci that have been associated with various smoking-related phenotypes tagged by three SNPs: rs16969968, rs578776, and rs588765. This analysis focuses on a replication of previous studies in a unique population, namely the Old Order Amish (OOA) of Lancaster County. The OOA is a closed founder population with exceptional genetic and environmental characteristics lending themselves to the study of complex diseases. As virtually no Amish women reportedly smoke, we restricted analysis to the 896 OOA males, who ranged in age from 18 to 96 years. Smoking behaviors, obtained by questionnaire, included regular tobacco use (yes/no), quantity used, and age of initiation. Smoking rates are low in Amish men (20%), and among smokers the intensity of smoking is low; between 1 to 3 cigarettes per day for those who reported quantity. Direct genotypes were not available for the three SNPs of interest and so we substituted highly correlated ($r^2>0.8$) SNP proxies that captured the genetic information at each locus. A fourth SNP in this cluster (rs1948) was also evaluated that had previously been associated with age of smoking initiation. SNP associations were tested using a regression approach under an additive genetic model that adjusted for age and family structure. Regular smoking was significantly associated, and in the expected direction, for all loci (odd ratios ranging from 1.16 - 1.34, all $p < 0.003$). Interestingly, frequency of the 'risk' allele (A) for rs16969968, normally 0.39 in Caucasian populations (CEU), is 0.56 in the OOA population. Neither age of initiation nor quantity smoked was significantly associated with any of these 4 SNPs. We observed that the highly replicated, statistically distinct loci in the *CHRNA5/CHRNA3/CHRNB4* gene cluster were associated with smoking behaviors in the Amish, with similar effect sizes. It is striking that these loci stand out given the low smoking rates and low intensity of smoking in this population.

1219F

Phospholipid Flippase ATP8A2 is essential for Normal Visual and Auditory Function and Photoreceptor and Spiral Ganglion Cell Survival. *X. Zhu[1,2], Y. Zhou[1,2], X. Zhu[1,2].* 1) The Sichuan Provincial Key Laboratory for Human Disease Gene Study and The Institute of Laboratory Medicine, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, Sichuan, China; 2) Sichuan Translational Medicine Hospital, Chinese Academy of Sciences, Chengdu, Sichuan, China.

ATP8A2 is an important P4-ATPase phosphatidylserine flippase, which is required for proper maintenance of neuron and axon function. Atp8a2 is highly expressed in the brain, spinal cord, testes and retina. In the retina, ATP8A2 is primarily expressed in the light-sensitive photoreceptor cells and localized in the outer segment compartment. Although mutations in ATP8A2 have been reported to cause mental retardation in humans and spinal motor neuronal degeneration in the wabbler-lethal (wl) mouse, the physiological and cellular role of ATP8A2 in sensory systems have not been investigated. In this study we have analyzed the retina and cochlea of Atp8a2 mutant wl mice in order to determine the role of ATP8A2 in visual and auditory systems. ATP8A2-deficient mice had shortened photoreceptor outer segments, a significant reduction in rod and cone photoresponses, and decreased photoreceptor viability. The ultrastructural organization of photoreceptor outer segments appeared normal. The auditory brainstem response threshold was significantly higher in the ATP8A2-deficient mouse. The inner and outer hair cells appeared normal, but a significant degeneration of spiral ganglion cells was apparent. Our studies indicate that ATP8A2 plays a crucial role in photoreceptor and spiral ganglion cell function and survival by maintaining phospholipid composition and asymmetry and contributing to vesicle trafficking in these neurons.

1220W

Mutations in Secretin receptor may be related to autism spectrum disorder. *K. Kojima, T. Yamagata, A. Matsumoto, Z. Yang, S. Nakamura, E. F. Jimbo, M. Y. Momoi.* Pediatrics, Jichi Medial University, Shimotsuke-shi, Tochigi, Japan.

Secretin is a peptide hormone released from duodenum to stimulate exocrine secretion from pancreas. It is also produced in the brain and functions as a neuropeptide hormone in the central nervous system. Secretin receptor (SCTR) is a member of the type G-protein coupled receptor family, and expressed in the brain. In the rat brain during the early postnatal period, an active synaptogenesis period, expression levels of secretin and SCTR mRNA are higher than in the adult brain. (Tay, et al. Neuroscience letters. 2004) Since SCTR-deficient mice show abnormal social and cognitive behaviors, furthermore, synaptic plasticity in the hippocampus is impaired (Nishijima, et al.HMG.2006), the involvement of secretin to the pathogenesis of autism spectrum disorder (ASD) is suggested. To investigate the contribution of secretin and its pathway to ASD, we screened the SCTR gene for mutations in ASD patients. (Subjects and methods) We analyzed 197 DNA samples from Japanese ASD patients with the parent's informed consent and Caucasian patients from the Autism Genetic Resource Exchange (AGRE). Each exon and adjacent intron of SCTR was amplified by PCR and subjected to direct sequencing. Base changes detected in ASD samples were screened in 126 Japanese and Caucasian control samples. (Results) We detected three missense mutations in the SCTR gene that were not detected in controls, the R2C in one Caucasian male patient, the P90L in one Japanese male patient and the A245T in one Caucasian male patient. We also found another base substitution in the SCTR gene, the F411S in two Caucasian patients, but also in one control. (Discussion) These mutations are conserved regions. The R2C mutation in the SCTR is located in signal peptide domain and the P90L mutation is located in hormone binding domain. These mutations can have effect to alter the secretin binding to the receptor and its function. These findings suggested that secretin pathway play an important role in the pathogenesis of ASD.

1221T

Hypermethylation of the enolase gene (ENO2) in autism: A candidate biomarker for screening. *N. Zhong[1, 2, 3], Y. Wang[2].* 1) Department of Human Genetics, New York State Institute for Basic Research in Developmental Disabilties, Staten Island, NY; 2) Institute of Children Health Care, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China; 3) Peking University Center of Medical Genetics, Beijing, China.

It has been hypothesized that dysregulation of brain-expressed genes is the major predisposing underlying mechanism for autism. This dysregulation may be mediated by differential methylation of CpG sites within gene promoters, resulting in the unbalanced production of certain proteins, which could be candidate biomarkers and which could be used for early clinical screening of autism. Peripheral blood from autistic and control children were analyzed by methylated-DNA immunoprecipitation (MeDIP) chips to discover differential methylation. One neuron-specific gene, ENO2, was found to be hypermethylated in the autistic samples. This difference was validated by bisulfite sequencing PCR (BSP). The differential expression of ENO2 gene was further analyzed with real-time quantitative polymerase chain reaction (RT-qPCR) and by ELISA. We obtained blood samples from 101 age and sex matched autistic and control subjects. Five pairs were randomly selected for MeDIP chip analysis. A total 475 genes were found to be differentially methylated comparing the five pairs. This included 228 different in promoter regions and 247 different in CpG islands. The hypermethylation of ENO2 within the promoter region was confirmed by BSP to be present in 15.8% (16/101) of the autistic samples. The mean ENO2 RNA levels in these 16 autistic samples was reduced by about 70% relative to that in controls. The average level of ENO2 protein expression in the 16 autistic samples (15.15±3.52 μg/L) was about half of that in the controls (33.78±8.18 μg/L). These findings suggest that reduced ENO2 expression may be a biomarker for a subset of autistic children.

1222F

Application of causal inference methods to investigate APOE variation and cognitive impairment in multiple sclerosis. *E. Elboudwarej[1], P. Yousefi[2], L. Shen[3], M.F. George[1], H. Quach[1], F.B.S. Briggs[1], A. Bernstein[4], C. Schaefer[3], L.F. Barcellos[1,3].* 1) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, UC Berkeley, Berkeley, CA; 2) Division of Environmental Health Sciences, School of Public Health, UC Berkeley, Berkeley, CA; 3) Division of Research, Kaiser Permanente, Oakland, CA; 4) Palm Drive Hospital, Petaluma, CA, USA.

Previous studies have shown that apolipoprotein E (APOE) variation, particularly the ε4 allele, is not associated with overall risk of developing multiple sclerosis (MS) or disease severity measured by the MS Severity Score (MSSS). Cognitive deficits affect ~70% of MS cases and greatly impact daily activities, employment and relationships. However, cognitive issues vary among MS cases and do not strongly correlate with physical disability or disease course. APOE is a widely studied gene in relation to cognition, and recent studies suggest a possible link between ε4 and greater cognitive dysfunction in MS. Here, a causal framework was used to assess the effect of APOE ε4 genotype on cognitive impairment. Subjects were recruited from the Kaiser Permanente Medical Care Plan, Northern California Region (877 MS cases and 576 controls). Cognitive status was determined using a 37 point scale validated telephone interview cognitive status (TICS-M) assessment tool, and accounted for level of education. As expected, when compared to control individuals, MS case cognitive scores were lower ($p=4\times10$-4). The G-computational formula simple substitution estimator (SS) and target maximum likelihood estimation (TMLE) were used to estimate the effect of APOE ε4 carrier status on cognitive score in MS cases. Non-parametric models were adjusted for the following covariates to help identify the true causal relationship between genotype and outcome: gender, depression, genetic ancestry, family history of MS, MSSS, age of onset, age at cognitive assessment, and previous use of disease modifying therapies. The machine learning algorithm, SuperLearner in R, was used to select the prediction model with the lowest cross-validated risk, for implementation of SS and TMLE. Results were also compared to those obtained from traditional parametric linear regression modeling. Using the G-computational formula SS, MS cases carrying at least one ε4 allele scored higher on the cognitive test, though this difference was very modest (β =0.27; 95%CI: 0.12, 0.76). Results based on TMLE showed no evidence for an APOE ε4 effect on cognitive impairment in MS cases (β =0.30; 95%CI: -2.40, 3.77) and were consistent with linear regression analyses (β =0.24, 95%CI: -0.49, 0.97). Overall, our results based on causal inference methods do not support a role for APOE ε4 in cognitive impairment observed in MS.

1223W

Social skills impairments in girls with Turner syndrome. *M. Inbar-Feigenberg[1,2], D. Grafodatskaya[2], S. Choufani[2], BHY. Chung[1,3], LJ. Roberts[2], C. Russell[4], W. Roberts[4], J. Hamilton[5,6], R. Weksberg[1,2,6].* 1) Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, ON, Canada; 3) Centre of Reproduction, Growth & Development, Department of Pediatrics & Adolescent Medicine, The University of Hong Kong, Hong Kong; 4) Autism Research Unit, Hospital for Sick Children, Toronto, ON, Canada; 5) Division of Endocrinology, Department of Pediatrics, Hospital for Sick Children, Toronto, ON, Canada; 6) Department of Pediatrics, University of Toronto, Toronto, ON, Canada.

Introduction: Turner syndrome (TS) is one of the most common sex chromosome abnormalities caused by complete or partial monosomy of the X chromosome, with a prevalence of ~1/2000 female live births. Individuals with TS present with short stature, gonadal dysfunction and other systemic malformations. A specific neuro-cognitive profile has been reported, sometimes including impaired social cognition. Autism and autism spectrum disorders (ASD) are reported in 5% and 25% of TS patients respectively. Skuse et al, 1997 reported that females who inherit their single X-chromosome from their father have better social skills than females who inherit it from their mother. The authors hypothesized that an imprinted locus on the X-chromosome is relevant to social functioning. Hypothesis: Females with TS demonstrate parent of origin- specific differences in social cognition. Methods and Results: We recruited 28 individuals with TS (age 3–18 years) and their parents at the Pediatric Endocrinology Clinic. We collected buccal samples from the proband and both parents. In addition, parents completed two social skills questionnaires for their daughters, one used originally by Skuse et al and the Social Responsiveness Scale (SRS) that assesses social awareness, social cognition, social communication, social motivation, and autistic mannerisms. SRS total scores fall into three categories: normal (T score ≤59); mild to moderate, consistent with mild ASD (T score=60–75); and severe, consistent with autism (≥76). We compared the scores for groups of girls with TS carrying a single maternal vs paternal X chromosome, as well as a group with karyotypes other then XO. In 14/28 patients (50%) scores were> 60. In 7/28 (25%), scores were in the mild/moderate ASD range and in 7/28 (25%), scores were in the severe autism range. Score differences for the sub-scales of SRS showed higher scores for autistic mannerisms. A good correlation between Skuze et al and the currently SRS was found ($R^2 = 0.80$). We did not find any correlation in social skill measures with parent of origin of the X chromosome in our small TS cohort. Conclusions: We found the rate of autistic features in TS to be significantly higher than previously reported. These data have significant implications for genetic counseling. We suggest that individuals with TS be routinely screened for ASD for early identification and initiation of behavioral interventions.

1224T

Examining the role of the 1q deletion critical region gene *SCCPDH* in early brain development in zebrafish. *E.A. Burke[1,2], W.A. Gahl[2], C.F. Boerkoel[1].* 1) NIH Undiagnosed Diseases Program Translational Laboratory, Bethesda, MD; 2) Section on Human Biochemical Genetics, Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

Submicroscopic terminal 1q deletion results in a syndrome consisting of severe psychomotor delays, aphasia, hypotonia, microcephaly, corpus callosum abnormalities, and facial dysmorphism. Though previous studies have narrowed the critical region for this phenotype to 1q43-1q44, the deleted genes responsible for the syndrome remain unknown. We hypothesized that a potential contributor to this syndrome is *SCCPDH*, which lies within the 1q44 critical region and is expressed in the developing human and zebrafish brain. Knockdown of the zebrafish homologue using morpholinos caused a dose-dependent loss of brain tissue at the midbrain-hindbrain boundary, decreased eye size, and hydrocephalus. Each phenotype was rescued by co-injection of wild type *SCCPDH* mRNA in over 80 percent of embryos. Overexpression of wild type mRNA in zebrafish resulted in an increased incidence of cyclopia and abnormal forebrain development. Through *in situ* hybridization, *SCCPDH* morphants were found to widely overexpress *pax2* in the optic primordia in a manner similar to embryos injected with *shh* (sonic hedgehog) mRNA. These results, along with gene expression analyses, suggest a potential link between SCCPDH and Shh-related signaling. Therefore, we conclude that SCCPDH expression is highly regulated during brain development and that one mechanism by which it may affect brain development is through modulation of Shh signaling.

1225F

Investigation of the role of serotonin pathway in the etiology of Tourette's syndrome. *N. Sun[1,2], Z. Pang[5], L. Deng[1,2], RA. King[6], M. Sheldon[1,2,3], JC. Moore[1,3], RP. Hart[3,4], M. Konsolaki[1,2], GA. Heiman[1,2], JA. Tischfield[1,2,3].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Piscataway, NJ; 3) RUCDR Infinite Biologics NIMH Stem Cell Resource, Piscataway, NJ; 4) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ; 5) Child Health Institute of New Jersey, New Brunswick, NJ; 6) Yale Child Study Center, Yale University School of Medicine, New Haven, CT.

Tourette's syndrome (TS) is a childhood onset neurodevelopmental and genetic disorder characterized by chronic motor and phonic tics. Previous studies suggest the etiology of TS is the result of interactions between multiple genetic and environmental factors, but specific genetic factors contributing to TS causation are poorly understood. A recent publication documented that an I425V mutation in the serotonin transporter gene (SLC6A4) is found in certain subsets of TS patients, suggesting the involvement of the serotonin pathway in the etiology of TS. It is unclear how this mutation leads to altered expression of genes in neurotransmitter pathway(s). We collected blood samples from all five family members of a two-generation pedigree from the New Jersey Center for Tourette Syndrome Sharing Repository. Four individuals from this family, diagnosed with TS or/and Obsessive-Compulsive Disorder (OCD), carry the I425V mutation. We established induced pluripotent stem cell (iPSC) lines from each individual in this family. In order to develop a neuron-based TS model in vitro, we induced iPSC lines into neural stem cells (NSCs) and further differentiated the NSCs into electrophysiologically functional neurons. By performing whole cell patch clamp analysis on these neurons, we are assessing whether or not the I425V mutation leads to electrophysiological changes. In addition, we are conducting gene expression analyses of NSCs and neurons to identify any differentially expressed genes in neurogenesis pathways involving serotonin which may lead to the discovery of epigenetically-modified genes involved in disease causation and development.

1226W

Distinguishing Autism Spectrum Disorders From Other Developmental Delays Using Blood RNASeq. *S. LETOVSKY[1], M. Causey[1], M. Aryee[2], J. Skoletsky[1], C. Proulx[1], F.R. Sharp[3], I.N. Pessah[3], R. Hansen[3], J. Gregg[3], I. Hertz-Picciotto[3].* 1) SYNAPDX, Woburn, MA; 2) Massachusetts General Hospital, Boston, MA; 3) MIND Institute, University of California, Davis, CA.

There is an unmet need for objective biomarkers to assist clinicians in the early diagnosis of childhood neurodevelopmental disorders. The aim of this study was to assess whether blood gene expression measured using next generation RNA sequencing (RNASeq) could provide a biomarker to distinguish children on the autism spectrum from children with other conditions that might present in the same clinical setting. The CHARGE (CHildhood Autism Risks from Genetics and the Environment) study recruited children between the ages of 2 and 5, some of whom were diagnosed on the autism spectrum, and others with other developmental delays. Subjects were grouped based on thresholds of the ADOS, ADI-R, Vineland and Mullens test into autism spectrum disorder (ASD) and other developmental delay (DD) groups to approximate the clinical use case of a secondary screen for autism in children suspected of neurodevelopmental disorders. RNASeq was performed on RNA from blood samples acquired in PAXgene tubes. 174 ASD and 96 DD samples passed final QC, for a total of 270 samples. Sequence data were processed through the Tuxedo RNASeq pipeline. Samples were divided into a training set (n= 153) and a holdout set (N=117), each of which was repeatedly subsampled to achieve gender and age balance between the ASD and DD groups. On each iteration, informative features were selected by t-test and a support vector machine classifier was trained on a balanced subsample of the training set and tested on a balanced subsample of the holdout set; AUC's (area under the ROC curve) were averaged across iterations. The mean AUC for the holdout set was 65.6 +/– 2.9%. When a 90% sensitivity threshold was selected on the classifier risk score, the mean specificity was 25.3, with 95% CI [13.6, 40.6%]. Gene categories found significant by ranksum test on the t-statistic include RNA processing, cell cycle, immune and inflammation-related GO categories. To our knowledge this represents the first report of a classification signature for ASD vs. DD using blood RNASeq. A gene expression signature with moderate AUC has potential clinical utility as a sensitive assay for identifying children at risk for ASDs within a population that is already suspected of neurodevelopmental disorders. Planned followup studies include a multicenter clinical study to further refine and validate a blood-based assay.

1227T

A 3.1 Mb deletion at 2q22.1-q22.3 including NXPH2 gene in a boy with intellectual disability. *N. Yoshihara[1], K. Kurosawa[1], I. Oohashi[1], Y. Kuroda[1], H. Osaka[2], S. Yamashita[2], M. Iai[2], K. Takano[2], M. Okuda[2], M. Takagi[2], R. Anzai[2], T. Wada[3].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Pediatric Neurology, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Division of Medical Genetics, Kyoto University, Kyoto, Japan.

[Introduction] The synapses play a crucial role in neural transmission and complicated brain function. NXPH2 gene, expressed in brain and kidney, encodes Neurexophilin 2, which is a glycoprotein, working as signaling molecules with Neurexin 1, cell surface receptors at synapses. [Case] The patient is 4-year-old boy, without any familial history nor abnormal birth history. He was referred to our medical center for investigation of his intellectual disability (IQ 63). He did not show autistic signs or symptoms. He had a distinctive facial appearance with full cheeks and widely spaced eyes, low-set-ears and narrow mouth. MRI and EEG were normal. Array-comparative genomic hybridization revealed a 3.1Mb deletion at 2q22.1-q22.3 including NXPH2 and a part of LRP1B gene, which is known as a tumor suppressor gene. [Discussion] Recently, more reports are published suggesting the link between genes coding synaptic proteins, including the Neurexin family and autism, intellectual disability or schizophrenia. Previously, only 3 cases of the deletion including NXPH2 were reported. All of those regions include other several genes and all of those patients presented with external malformation and severe autism or intellectual disability. The present case had a minimum deletion including NXPH2. These results indicated that intellectual disability of this patient results from haploinsufficiency of NXPH2.

1228F

Mutation spectrum in the dystrophin gene disclosed by MLPA in 181 Vietnamese Duchenne/Becker muscular dystrophy patients. *K.V. Tran[1], H.N. Do[1], T.H. Tran[1], M.H. Ta[1], A.T.P. Le[1], D.C. Vu[2], V.T. Ta[1], M. Matsuo[3].* 1) Center for Gene and Protein Research, Hanoi Medical University, Hanoi, Viet Nam; 2) Department of Medical Genetics, Metabolism & Endocrinology, National Hospital of Pediatrics, Hanoi, Vietnam; 3) Department of Medical Rehabilitation, Faculty of Rehabilitation, Kobegakuin University, Kobe, Japan.

Duchenne/Becker muscular dystrophy (DMD/BMD) the most common X-linked muscular dystrophy is caused by mutation in dystrophin gene. Deletion and duplication in the dystrophin gene account for 60-70% of mutation. Multiplex ligation-dependent probe amplification (MLPA) is the most powerful and convenient method to identify exon deletions or duplications in the dystrophin gene because of its overall gene coverage. The present investigation was designed to detect mutation in the dystrophin gene in 181 unrelated Vietnamese Becker/Duchenne patients using MLPA analysis. Among the 181 cases, deletions and duplications encompassing one or more exons were identified in 105 (58%) or 12 (6.6%) cases, respectively. Deletions were found to cluster in the proximal (14.3%) and central hotspot regions (72.4%); 14% were observed to have gross deletions and 1.2% had deletion out of hotspot regions (exon 61-67). The deletion patterns were categorized into 48 patterns. Deletion of exon 48-50 or 45-50 where the most common pattern was deletion of exon 48-50,which was found in 11 cases (10%). Single-exon deletion was found in 14 cases (13%) by MLPA. Further examination disclosed that one of them was not an exon deletion but a single-nucleotide change (c.2227C>T) leading to a nonsense mutation. Outliers from the reading frame rule were 11 DMD (10.4 %). Remarkably, 25 and 14 cases were found treatable by exon 51 and 53 skipping, respectively. From these findings, the largest mutation database of Vietnam dystrophinopathy was established.

1229W

Underlying Genetic Etiologies in Epilepsy Patients:A retrospective large single centre study. *A. Guerin[1], E. Imhof[1], M. Zak[2], S. Mercimek-Mahmutoglu[1].* 1) Division of Clinical and Metabolic Genetics, Departmant of Pediatrics, The Hospital for Sick Children, Toronto, Canada; 2) Division of Neurology, Department of Pediatrics,The Hospital for Sick Children, Toronto, Canada.

Background: Epilepsy is abnormal neuronal electrical activity in the brain causing recurrent seizures. About 0.5-1% of the general population are affected by epilepsy. 25-30% are classified as intractable. Despite extensive investigations and neuroimaging studies, an underlying etiology is not determined in more than 50% of patients. Methods: All epilepsy patients were included into this retrospective review followed in the Neurology and Clinical and Metabolic Genetics Clinics at The Hospital for Sick Children between 2004-2013. Electronic patient charts were reviewed. Information was entered into an excel database. Genetic etiologies were subdivided into chromosomal, defined genetic syndromes, inherited metabolic disorders and neurogenetic syndromes. Results: There were 1298 epilepsy patients. There was no identifiable underlying cause in 688 patients (53%). In 610 patients (47%) there was an identifiable cause. In the latter group, chromosomal disorders, detected by karyotype and microarray, comprised 43/610 patients (7%). Identifiable genetic syndromes, e.g. Marfan syndrome, Down syndrome, Williams syndrome, represented 67/610 patients (11%). Neurogenetic causes, e.g. Dravet syndrome and Tuberous sclerosis complex were found in 74/610 patients (12.1%). Inherited metabolic disorders were represented by 45/610 cases (7.4%) the two most common causes being lysosomal storage disorders and mitochondrial encephalopathy. Conclusion: Collectively, genetic etiology was the most common underlying cause in 37.5% epilepsy patients. It is very likely that there is an underlying genetic cause in the 53% epilepsy patients with no known etiology. Prospective studies using whole genome sequencing will help us to unravel genes causing epilepsy, which will guide us to find new treatment for intractable epilepsies.

1230T

A useful approach for the genetic diagnosis of infectious agents using next-generation sequencing. *Y. Sakiyama[1], Y. Higuchi[1], H. Tanabe[1], J. Yuan[1], S. Nozuma[1], A. Yoshimura[1], A. Hashiguchi[1], Y. Okamoto[1], R. Hirano[1], E. Matsuura[1], R. Okubo[1], T. Iwasaki[2], F. Matsuda[3], S. Izumo[4], H. Takashima[1].* 1) Department of Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan; 2) Department of Dermatology, Graduate School, Tokyo Medical and Dental University, Tokyo, Japan; 3) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto Japan; 4) Center for Chronic Viral Diseases, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan.

Background: We found four patients with slowly progressive encephalomyelitis with cognitive dysfunction in the same geographical area. Histopathologic analysis of biopsied brain tissue revealed aggregation of periodic acid-Schiff-positive macrophages, and an agent with distinctive features was observed around the vessels in all patients. However, culture polymerase chain reaction (PCR) and serologic assays for a wide range of infectious agents were uninformative. **Methods:** Transcriptome analysis was performed using a next-generation sequencer (MiSeq®, Illumina) on cDNA extracted from the biopsied brain samples of two patients using a paired-end sequencing approach. To identify nonhuman cDNA, the Burrows-Wheeler Aligner (BWA) was first used against a human genome (UCSC: hg19) and human RefSeq (release 54, mRNA) to roughly remove most human sequences. The remaining unmapped reads were subsequently analyzed by BLASTN against the nucleotide database from NCBI. To detect DNA fragments of the pathogen more efficiently, the PAS-stained pathogens located in the perivascular spaces of the brain of the two patients were separately isolated from the section by laser microdissection (LMD). **Findings:** Among the cDNA sequences obtained from the affected brain tissue of two patients, a few reads displayed striking homology to a gene present in a microorganism, which we considered the infectious agent causing the present disease according to morphological observation. Furthermore, additional DNA sequences from the LMD-isolated tissue of the two patients were extremely homologous to the genome of the same microorganism. **Interpretation:** We consider that unbiased high-throughput sequencing using LMD-isolated tissue can be a useful tool for the detection of infectious agents if the morphological features are already revealed by microscopic observations.

1231F

Decreased tryptophan metabolism: the biochemical fingerprint of Autism Spectrum Disorders. *L. Boccuto[1], C.F. Chen[1], A.R. Pittman[1], C.D. Skinner[1], H.J. McCartney[1], K. Jones[1], B.R. Bochner[2], R.E. Stevenson[1], C.E. Schwartz[1].* 1) Greenwood Genetic Center, Greenwood, SC; 2) Biolog, Hayward, CA.

Autism spectrum disorders (ASDs) are common neurodevelopmental conditions whose biological basis has been largely undetermined. Although the ASD prevalence has risen to 2% of the school-aged US population and the conditions have a tremendous impact on society and families, the current diagnosis is solely based on the analysis of the complex behavioral phenotype and is typically not performed before the age of 3 years. We analyzed the metabolic profile of lymphoblastoid cell lines from 87 patients with ASDs, 50 with non-ASD neurodevelopmental disorders, and 78 normal individuals, using Biolog Phenotype MicroArrays (PMs). This assay is designed to evaluate the production of energy (NADH) in cells via diverse metabolic pathways. The ASD cohort showed reduced energy production when tryptophan was the sole energy source. The abnormality in tryptophan metabolism correlated with both syndromal and non-syndromal ASD, independent of the genetic background of the individual. This metabolic alteration was not observed in lymphoblastoid cell lines from the 78 normal patients or from the 50 patients with non-ASD intellectual disability, schizophrenia or conditions similar to syndromal ASD except for the behavioral traits. Additionally, analysis of data from a previous gene expression study of 10 patients with ASDs found abnormal levels for some genes involved in tryptophan metabolic pathways. Preliminary results indicate that the decreased tryptophan metabolism is detectable even in fresh blood leukocytes from patients with ASDs, suggesting that the assay might be further simplified and optimized into a quick and reliable screening tool, applicable at any age. Tryptophan is an essential amino acid and a precursor of serotonin, quinolinic acid, and kynurenic acid which are involved in neurodevelopment and synaptogenesis. Quinolinic acid is the structural precursor of NAD+, a critical energy carrier in mitochondria and the serotonin branch of the tryptophan metabolic pathway generates NADH. Additionally, the levels of quinolinic and kynurenic acid are strongly influenced by the activity of the immune system. Therefore, decreased tryptophan metabolism may alter brain development, neuroimmune activity and mitochondrial function. Our finding of decreased tryptophan metabolism appears to provide a unifying biochemical basis for ASDs and an initial step in the development of treatment or prevention protocols.

1232W

Genetic evaluation of hereditary spastic paraplegia. *S. Schirmer[1], Z. Kohl[2], T. Rödl[1], M. Becher[1], J. Winkler[2], U. Hehr[1,3].* 1) Center for Human Genetics , Regensburg, Germany; 2) Division of Molecular Neurology, University Hospital Erlangen, Erlangen, Germany; 3) Department of Human Genetics, University of Regensburg, Regensburg, Germany.

Hereditary spastic paraplegias (HSPs) comprise a clinically and genetically heterogeneous group of neurodegenerative disorders with progressive degeneration of the corticospinal tract. Patients with complicated forms show additional clinical findings, e.g. thin corpus callosum, cognitive impairment or peripheral neuropathy overlapping with a wide spectrum of other underlying genetic conditions. Currently 56 HSP loci have been assigned with causal mutations identified in 37 genes. Genetic testing by conventional Sanger sequencing in the diagnostic setting usually is restricted to the genes, most frequently affected. We here report the results of genetic testing for HSP by linkage analysis, Sanger sequencing and MLPA over the last 10 years. A subgroup of 20 patients received extensive neurologic workup in our outpatient clinic for movement disorders (cohort 1). In the overall cohort of 176 patients with suspected pure and complicated HSP causal mutations were identified in SPG4 (n=39), SPG3a (n=8), SPG7 (n=2), SPG11 (n=2), SPG20 (n=2) and SPG31 (n=2). To further increase the diagnostic yield we now evaluated a virtual gene panel for massive parallel sequencing of currently 30 HSP genes (step 1) based on the Illumina TruSight™Exome assay, which for further diagnostic workup of mutation negative patients allows subsequent evaluation of additional candidate genes (step 2) associated with other monogenic forms of spastic movement disorders. We validated step 1 on a cohort of 12 HSP patients with known mutations. Sample preparation was processed with the TruSight™Exome Enrichment kit, the Library was sequenced as a 150bp paired-end run with the MiSeq Reagent Kit v2. Reads were aligned to the human reference genome (UCSC hg19, NCBI build 37.1) and evaluated with an in house workup including the DNASTAR Lasergene®SeqMan Pro™ and DNASTAR ArrayStar® software. With our step 1 protocol we reached an average coverage of 58x with a minimal coverage of more than 20x for 91% of all 460 analyzed targets. All 12 expected HSP causing mutations were reliable detected including nonsense (n=2), missense (n=7) and frameshift (n=1) mutations, one 5bp deletion and one 9bp duplication. Clinically selected families of cohort 1 with suspected complicated HSP and more than one affected patient are currently evaluated with the step 1 and step 2 panel in order to further delineate the genetic spectrum of rare autosomal recessive complicated HSP.

1233T

Functional Analysis of the Autism and Intellectual Disability Gene PTCHD1 Reveals interactions in Synapses and Involvement in the Hedgehog Pathway. *K. Mittal, K. Sritharan, B. Degagne, J. Vincent.* Molecular Neuropsychiatry and Development Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.

This study is focused on investigating the functional aspects of a recently identified gene, PTCHD1, and how its disruption leads to Autism Spectrum Disorder and/or Intellectual Disability. Firstly, we sought to identify PTCHD1 splice variants contributing different coding sequences that might have higher or more specific expression in brain. We identified a new transcript skipping exon 2. This new transcript is predicted to encode a 542 amino acid protein in comparison to the 888 amino acid protein encoded by the PTCHD1 long isoform. It has just 4 transmembrane domains, and encodes a 62 kDa protein as compared to 12-transmembrane domains in the long isoform and a 101 kDa protein. We also found the presence of an additional exon upstream of exon 1. The quantitative expression analysis demonstrates that expression of the two new PTCHD1 transcripts is highest in human cerebellum as compared to the brain sub-regions and other tissues. The biological relevance of the new isoforms remains to be investigated. Initial expression data, indicates that these additional PTCHD1 transcripts are expressed chiefly in the brain, thus, these transcripts and encoded isoforms may be more relevant to autism and ID. PTCHD1 is structurally similar to the Hedgehog (Hh) signaling pathway receptors PTCH1 and PTCH2. The Hh pathway plays an important role in embryonic development and adult stem cell functioning. To establish the involvement of PTCHD1 in Hh pathway, expression analysis was done with Hh pathway genes and putative PTCHD1 partners. The quantitative expression analysis with over expression of PTCHD1 revealed increased levels of neuroligin and neurexin transcripts, which indicates either interaction with proteins at the synapse or a regulatory effect on these genes. Immunoprecipitation studies indicate interactions between PTCHD1 and Postsynaptic Density protein 95 (PSD-95). PSD-95 determines the size and strength of synapses in the postsynaptic density of neuronal excitatory synapses, and this putative interaction could be important to elucidate disease etiology. Preliminary results suggest sub-cellular localization of PTCHD1 in cilia, suggesting similar function as other Patched-related proteins. We hypothesize that PTCHD1 localization to primary cilia could inhibit the Hh pathway by excluding Smoothened and also allows cilia to function as chemo sensors for the detection of extracellular Shh, similar to PTCH1, during neuronal development and synapse formation.

1234F

GABRB2 in Schizophrenia. *H. Xue.* Division of Life Science, Hong Kong University of Science and Technology, Hong Kong, China.

Deciphering the molecular basis of schizophrenia is essential to effective management of this devastating mental disorder. Over the past decade, my research group has focused on the basic research on schizophrenia etiology through the discovery and characterization of a schizophrenia-associated gene - GABRB2, coding for GABAA receptor β2 subunit. The association between schizophrenia and single nucleotide polymorphisms (SNPs) in introns 9 and 10 of GABRB2, first reported by my group, has been cross-validated for multiple ethnic groups. Functional impacts of the schizophrenia associated non-coding SNPs in GABRB2 have been demonstrated at both mRNA and protein levels, viz. genotype-dependent alterations in mRNA expression and splicing, and effects of genotypes on isoform ratios and electrophysiological attenuation of GABAA receptors. Through extensive molecular genetics, population genetics and evolutionary genetics characterizations, GABRB2 has been shown by us to be under strong positive selection, active recombination as well as genomic imprinting, likely contributed to by a human lineage-specific insertion of an AluY transposable element. Their work on epigenetic regulation of GABRB2 revealed its developmental control and disruption in schizophrenia. Most recently, they have also extended GABRB2 association from psychotic disorders to social cognitions. Our work has thus improved current understanding of schizophrenia at molecular level centered at GABRB2, which represents at present one of the best characterized schizophrenia candidate genes.

1235W

Association between emotional memory and genetic polymorphisms on histone deacetylase 5. *F.R. Hartmann[1], D.J.-F. de Quervain[2,3], A. Papassotiropoulos[1,2,4], A. Heck[1].* 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Basel, Switzerland; 2) University of Basel, Psychiatric University Clinics, Basel, Switzerland; 3) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Basel, Switzerland; 4) University of Basel, Department Biozentrum, Life Sciences Training Facility, Basel, Switzerland.

Histone acetylation is a key epigenetic mechanism that enhances gene transcription by adding acetyl groups to the histone tail. Histone deacetylases (HDACs) remove acetyl groups, thus leading to a less accessible chromatin structure and a supressed gene expression. Several studies have shown that inhibition of HDACs enhances synaptic plasticity and memory formation. For example, HDAC inhibition modulates memory in the hippocampus via CREB (cyclic-AMP response element binding) binding protein. In mice, HDAC inhibition during memory consolidation of an object recognition task enhances formation of long-term memory. Histone deacetylase 5 (HDAC5) is a member of class II HDACs, which are characterized by shuttling between the nucleus and the cytoplasm and play an important role in transcription regulation, cell cycle progression and developmental events. HDAC5 has been associated with behavioural adaptations to emotional stimuli (cocaine addiction and chronic stress) in mice. However, the role of HDAC5 in learning and memory formation in humans is still unclear. We therefore investigated the role of the HDAC5 gene for episodic memory in two samples of healthy young participants (n=1'293 and n=983), who performed the identical picture-based episodic memory task using stimuli from the International Affective Picture System (IAPS). The phenotype of interest was episodic memory performance, measured by the number of correctly recalled pictures after 10 minutes delay. Genotype data were available for five single nucleotide polymorphisms (SNPs) mapping on HDAC5 (Affymetrix chip 6.0). We observed significant associations between two SNPs and memory performance particular for emotional stimuli in both samples, with carriers of the minor-allele showing worse memory performance (p(combined sample) = $7.2*10^{-5}$ and $7.8*10^{-5}$; respectively). Our results suggest a role for HDAC5 in emotional memory formation.

1236T

***CMYA5*, a candidate gene for schizophrenia: Expression in the brain and the effect of a functional variant on binding.** *A. Hsiung, R. Shiang, X. Chen.* Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA.

Schizophrenia is a devastating psychiatric disorder with a prevalence of approximately 1% and is characterized by delusions, hallucinations and deficits of cognition and emotion. Although many DNA variants are identified to associate with schizophrenia, the functional consequences of such variants that result in increased risk are unknown. In collaboration with a large group of investigators, we have found strong evidence that the *CMYA5* gene is associated with schizophrenia in a two-stage study using more than 33,000 subjects (Chen et al., 2010). The non-synonymous single nucleotide polymorphism (SNP), rs10043986, is a functional variant that changes the highly conserved proline residue 4063 to leucine in the *CMYA5* protein, myospryn. Current studies of *CMYA5* focus on its role in skeletal and cardiac muscles, and no studies of *CMYA5* in the brain and neuronal cells have been reported, the tissue most likely to be affected in patients with schizophrenia. We hypothesize that myospryn is expressed in the brain and this SNP affects the binding properties of myospryn to its binding partner, desmin intermediate filament. Using reverse transcription PCR, myospryn and desmin transcripts are shown to express in mouse brain regions. Western blotting analysis also confirms the protein expression of myospryn and desmin in the same brain regions. Next, we investigated whether the two variants of rs10043986 would change myospryn binding to desmin using a yeast two-hybrid assay (Y2H) and validating with surface plasmon resonance (SPR). The results show that both alleles of myospryn bind to desmin. A quantitative Y2H assay shows that the Pro allele has significantly weaker binding compared to the Leu allele, providing evidence that rs10043986 results in a functional change of myospryn's binding to desmin. The SPR result shows that the Pro allele has higher equilibrium dissociation constant than the Leu allele in the steady-state analysis, confirming the differential binding. In conclusion, expression of *CMYA5* in the brain provides evidence that it has function in the brain and is important for the schizophrenia pathophysiology, and rs10043986 affects the binding properties of myospryn to desmin and is a potential causal variant.

1237F

Association of NTRK2 and emotional processing in healthy young subjects. *K. Spalek[1], A. Heck[2], D. Coynel[1], D.J.-F. de Quervain[1, 4], A. Papassotiropoulos[2, 3, 4].* 1) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Basel, Switzerland; 2) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Basel, Switzerland; 3) University of Basel, Department Biozentrum, Life Sciences Training Facility, Basel, Switzerland; 4) University of Basel, Psychiatric University Clinics, Basel, Switzerland.

The neurotrophin tyrosine kinase receptor 2 (NTRK2), also known as 'tropomyosin-related kinase' receptor B (TrkB), is one of two receptor types binding neurotrophins with high affinity. Its activation plays a central role in cell survival, differentiation and synaptic plasticity as well as neurotransmitter release. Several studies provide support for a relationship of NTRK2 and psychiatric disorders, e.g. depression, schizophrenia, addiction, eating and anxiety disorders. Expression studies in patients with psychiatric disorders and animal models of psychiatric diseases report reductions of TrkB expression in the brain. Furthermore, injection of TrkB agonist in mice' hippocampus induced long-term activation of TrkB and antidepressant effects whereas the injection of TrkB antagonist decreased TrkB activity and increased anxiety- and depression-like behaviours. In general, emotional dysregulation is a common denominator of many psychiatric disorders, and it can be assumed that NTRK2 is involved in this process. Based on the evidence for an influence of NTRK2 on emotional processing in psychopathology and the lack of studies investigating the role of NTRK2 in healthy young subjects, we investigated genetic associations between NTRK2 single nucleotide polymorphisms (SNPs) and emotional processing, measured by arousal ratings of emotional (negative and positive) and neutral pictures from the International Affective Picture System (IAPS) in a sample of 1'161 healthy young subjects. Our results show a significant association between two NTRK2 SNPs and emotional processing of positive pictures. Specifically, major allele carriers rated positive pictures as more arousing compared to non-carriers. We were able to replicate this association result in an independent sample of 822 healthy young subjects who performed the identical picture rating task. Our findings suggest a role of NTRK2 in emotional processing in healthy young subjects and might add useful information for the understanding of its role in psychopathology.

1238W

PRICKLE1 interaction with SYNAPSIN I reveals a role in Autism Spectrum Disorders. *L. Paemka[1], V.B. Mahajan[1], J.M. Skeie[1], L.P. Sowers[1], S.N. Ehaideb[1], P. Gonzalez-Alegre[1], T. Sasaoka[2], H. Tao[3], A. Miyagi[4], N. Ueno[4], S. Wu[1], B.W. Darbro[1], P.J. Ferguson[1], A.A. Pieper[1], J.K. Britt[1], J.A. Wemmie[1], D.S. Rudd[1], T. Wassink[1], H. El-Shanti[5], H.C. Mefford[6], G.L. Carvill[6], J.R. Manak[1], A.G. Bassuk[1].* 1) University of Iowa, Iowa City, IA; 2) Kitasato University School of Medicine, Japan; 3) Hospital for Sick Kids, Toronto, Canada; 4) National Institute for Basic Biology, Japan; 5) Shaffalah Medical Genetics Center, Doha, Qatar; 6) University of Washington, Seattle, Washington.

The frequent comorbidity of Autism Spectrum Disorders (ASDs) with epilepsy suggests a shared underlying genetic susceptibility; and several genes, when mutated, can contribute to both disorders. Recently, PRICKLE1 missense mutations were found to segregate with ASD, however the mechanism by which mutations in this gene might contribute to ASD is unknown. To elucidate the role of PRICKLE1 in ASDs, we carried out studies in Prickle1+/- mice and Drosophila, yeast, and neuronal cell lines. We show that mice with Prickle1 mutations exhibit ASD-like behaviors. To find proteins that interact with PRICKLE1 in the central nervous system, we performed a yeast two-hybrid screen with a human brain cDNA library and isolated a peptide with homology to SYNAPSIN I (SYN1), a protein involved in synaptogenesis, synaptic vesicle formation, and regulation of neurotransmitter release. Endogenous Prickle1 and Syn1 co-localize in neurons and physically interact via the SYN1 region mutated in ASD and epilepsy. Finally, a mutation in PRICKLE1 disrupts its ability to increase the size of dense-core vesicles in PC12 cells. Taken together, these findings suggest PRICKLE1 mutations contribute to ASD by disrupting the interaction with SYN1 and regulation of synaptic vesicles.

1239T

Altered dose-response relationships in addiction genetics. *G.R. Uhl[1,2], J. Drgonova[1], F.S. Hall[1].* 1) Molecular Neurobiology, NIDA IRP NIH, Baltimore, MD; 2) .Depts of Neurology, Neuroscience and Mental Health, Johns Hopkins Medical Institutions.

Classical genetic studies document substantial heritable influences on DSM substance dependence. Linkage and genome wide association studies identify only modest-sized effects at any locus. Clusters of nearby SNPs within selected genes, including those that encode the cell adhesion molecules PTPRD, CDH13 and NrCAM, do display 10-2 > p > 10-8 associations with dependence in many more independent samples than expected by chance. For each of these genes, 5' SNPs display nominally-significant associations with levels of mRNA expression in postmortem brain samples. A validated model for abuse liability, conditioned place preference, can provide dose-response information about the rewarding (and aversive) properties of abused substances in mouse models, providing typical inverted U dose-response relationships. Unexpectedly, in mice with altered expression of each of these genes, we have identified altered dose-response relationships for cocaine-conditioned place preferences. Reduced expression of either CDH13 or PTPRD provides a leftward shift in dose response relationships for the ascending limb of the 'inverted U' dose response relationship that is characteristic of cocaine reward. Conversely, reduced expression of NrCAM appears to depress the descending limb of the dose response relationship for cocaine conditioned place preference. Control behavioral tests fail to provide evidence for confounding explanations for these results. Each of these observations receives significant additional support from human datasets that include association with altered dose response relationships from stimulant doses administered in laboratory settings, association with doses of self administered coffee or cigarette consumption, and association with self reported doses of alcohol required to exert subjective effects. It is important to consider dose-response relationships for both rewarding and aversive features of abused substances in genetic, nosologic and therapeutic approaches to addictions. (Financial support: NIDA-IRP/NIH/DHHS).

1240F

Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *A. Griciuc[1], A.R. Parrado[1], A. Serrano-Pozo[2], A.N. Lesinski[1], C.N. Asselin[1], K. Mullin[1], B. Hooli[1], S.H. Choi[1], L. Bertram[3], B.T. Hyman[2], R.E. Tanzi[1].* 1) Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA; 2) Alzheimer's Disease Research Laboratory, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA; 3) Neuropsychiatric Genetics Group, Max-Planck Institute for Molecular Genetics, Berlin, Germany.

Preservation of cognitive abilities is one of the major medical challenges of the 21st century. Here, we describe a novel pathway that regulates the clearance of amyloid beta (Aβ) in the aging brain, and might play a pivotal role in the pathogenesis of Alzheimer's disease (AD) [1]. We have previously shown that mutations in the CD33 gene confer increased risk for AD. CD33 is a sialic acid-binding immunoglobulin-like lectin that regulates innate immunity but has no known roles in the brain. Our extensive pathological analyses uncovered a prominent microglial expression of CD33 in the aging human brain; moreover, the number of CD33-positive microglia is markedly increased in AD relative to age-matched control brains. The minor (T) allele of the CD33 single nucleotide polymorphism (SNP) rs3865444, which confers protection against AD, is associated with marked reductions in CD33 protein levels, CD33 microglial expression and levels of insoluble Aβ42 in AD brain. Furthermore, the numbers of CD33-immunoreactive microglia closely correlate with the levels of insoluble Aβ42 levels and plaque burden in AD brain. Experiments using microglial cell cultures indicate that CD33 inhibits the uptake of Aβ42 by microglia, and requires its sialic acid-binding domain to mediate this effect. Mouse genetics experiments further indicate that CD33 deletion markedly decreases the levels of insoluble Aβ42 and the amyloid plaque burden in APP/PS1 transgenic mice. To gain further insight into the regulation of CD33 expression in the aging human brain, we are currently characterizing 5 additional CD33 mutations (including both protective and risk variants) that we have recently identified. Collectively, our results suggest that CD33 activity in microglial cells promotes Aβ pathology and inhibition of CD33 function could represent a novel therapy for AD.
[1] Griciuc et al. (2013), Alzheimer s Disease Risk Gene CD33 Inhibits Microglial Uptake of Amyloid Beta. Neuron, 78(4):631-43.

1241W

Involvement of non-Hsa21 genes and MicroRNAs provide etiological basis for abnormal phenotypes in pathogenesis of Down syndrome. *A. Pathak, S.R. Phadke.* Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh , India.

Down syndrome (DS), the most frequent genetic disorder leading to mental retardation is caused by partial or complete triplication of human chromosome 21 (Hsa21). The differential expression of genes located on extra chromosome 21 is generally assumed responsible for phenotypic abnormalities but this gene dosage hypothesis has not been fully assessed on genome-wide basis. The expression patterns of genes related to phenotypic abnormalities may provide insights into their roles in pathogenesis of DS. To analyze the differential gene expression and understand the molecular mechanism underlying pathogenesis of DS, we performed global gene expression profiling in blood samples of 14 DS and 4 normal subjects using human whole transcriptome microarray. The microarray analysis revealed total of 624 genes (195 upregulated and 429 down regulated) were differentially expressed in DS patients as compared to control. Out of the genes present on chromosome-21, a total of 210 genes were differentially expressed ranging from 1.5 to 5 fold compared to normal individuals. Genes involved in physiological pathways such as apoptosis regulation, cell cycle regulation, signal transduction, cell maturation, and immunity showed dysregulation. Several genes localized on chromosome-21 such as APP, SOD1, DYRK1A, COL6A1 showed differential expression and the levels were conserved across all DS subjects. Interestingly, several non chromosome-21 genes such as RCAN3 (chromosome 1), ANK3 (chromosome 10), CDK17 (chromosome 12) etc., having roles in cardiogenesis, signal transduction and differentiation of neurons showed conserved levels of expression across the DS subjects. Further to investigate the role of microRNAs (miRNAs) in regulation of gene expression, global miRNA profiling was performed in 4 Down syndrome patients and 1 control using Affymetrix miRNA 3.0 array. Several Hsa21 miRNAs like miR-99a, let-7c, miR-125b-2, miR155, miR-802 showed overexpression effecting the regulation of genes involved in DS pathogenesis. The gene dosage hypothesis on chromosome-21 may partially explain the neurological and other symptoms but our results substantiate the involvement of genes localized across different chromosomes in pathogenesis of DS. Further, identification of miRNAs involved in DS pathogenesis may lead to identification of new therapeutic targets for DS. Our data lead to more systematic and improved understanding of molecular mechanism underlying the pathogenesis of the disease.

1242T

Induced pluripotent stem cell derived neural cells and Parkinson's Disease-associated rare variants. *K. Belle, J.M. Van Baaren, A. Liu, J.A. Vance, D.M. Dykxhoorn.* Human Genetics and Genomics, Univeristy of Miami, Miami, FL.

Parkinson's disease (PD) is a neurodegenerative disorder of the central nervous system. Late in life PD patients experience motor-dysfunction and cognitive impairments such as dementia, progressing in severity with age. PD has been linked to genetic mutations (familial, idiopathic) as well as environmental exposure, though the causal reasons are different, the end phenotype is the same. While many of these genes have basic roles in many different cell types and throughout the life cycle, the disease manifests itself specifically in pars compacta of the substantia nigra. Symptoms are first noticed late in life and progress until death. The advent of iPSC technology provides researchers with a safe and relevant way to investigate and treat genetic disorders in a patient specific manner; a small tissue sample from an affected patient is used to produce appropriate cell type for study. To date, animal models have been used as models for PD. These animal models have been very useful, but it is difficult to use them to accurately recreate a complex disease, as their genetic backgrounds differ so greatly from that of humans. The study of PD in affected human cells will lead to better understanding of the genes and pathways involved in PD. Our research centers around modeling PD in vitro, with symptomatic cell types derived from actual patient disease cells, and age-matched controls, (to differentiate between effects related solely to age versus those related to PD). The goal of this research is to provide cellular models for PD in vitro. To further investigate PD-related gene expression and morphologies, we have devised a system using shRNA-mediated silencing of PD-associated genes, and expression of rare variants, in aging neural cells derived from controls and affected patient. Gene expression and morphological changes are collected with respect to patient disease status, as well as sample age (days in culture and passage number). This system of co-knockdown/expression, allows us to further study the effects of rare variants on different genetics backgrounds (familial affected, idiopathic affected, and control). It is believed that this system will exacerbate cellular phenotypes related to the disease. These types of experimentation will help us obtain insight into the pathways involved in the disease. Further, successful iPSC models of PD could be useful in testing potential therapeutic interventions.

1243F

More than ER stress: molecular mechanism for misfolded PLP1 that impacts subcellular dynamics and clinical severity of Pelizaeus-Merzbacher disease. *Y. Numata[1, 2], T. Morimura[3], S. Nakamura[1], E. Hirano[1], S. Kure[2], Y. Goto[1], K. Inoue[1].* 1) Mental Retardation and Birth Defect Research, NCNP, Kodaira, Tokyo, Japan; 2) Dept. Pediatrics, Tohoku Univ. School of Med, Sendai, Japan; 3) Unit for Neurobiology & Therapeutics, MNRC, Shiga Univ. of Medical Science, Otsu, Japan.

Involvement of endoplasmic reticulum (ER) stress and the subsequent unfolded protein response (UPR) has been implicated in pathogenesis of multiple human inherited diseases, including Charcot-Marie-Tooth disease (CMT) and Pelizaeus-Merzbacher disease (PMD). PMD-causing mutant PLP1 accumulates in the ER and induces ER stress. However, the link between the wide clinical severity of PMD and the cellular response induced by mutant PLP1 remains largely unknown. Here we identified that misfolded mutant proteins also impact global subcellular dynamics and cellular environment. We found that expression of mutant PLP1 in HeLa cells, MO3.13 oligodendrocytic cells, and primary oligodendrocytes depletes major ER chaperones with a KDEL (Lys-Asp-Glu-Leu) motif from the ER. This can be detrimental to cells because accumulation of misfolded proteins in the ER generally up-regulate ER chaperones to alleviate ER stress. The PLP1 mutants also induce fragmentation of the Golgi apparatus (GA). These organelle changes are more prominent in cells expressing severer mutant, potentially associated with disease phenotype. Similar changes are also observed in CMT-causing MPZ mutant that triggers ER stress, suggesting a common molecular mechanism among these disease genes. Moreover, PLP1 mutants inhibited global trafficking of secretary and membrane proteins, possibly leading to deleterious cellular environment. Notably, we found that KDEL receptor, which is critical in retrotransport of ER chaperons, was also dislocated in the ER, possibly causing the depletion of ER chaperons. The fact that inhibition of GA to ER trafficking by brefeldin A induced similar cellular phenotypes supported this hypothesis. Altogether, we propose that misfolded disease-causing mutant proteins not only induce ER stress and trigger UPR, but also cause depletion of ER chaperones, GA fragmentation and protein trafficking deficit, further contributing to the pathogenesis of inherited ER stress-related diseases and their disease severity.

1244W

Potential regulatory functions of late-onset Alzheimer's disease associated variants. *S.L. Rosenthal[1], M.M Barmada[1], X. Wang[1], F.Y. Demirci[1], O.L. Lopez[2,3], M.I. Kamboh[1,3].* 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA; 3) Alzheimer's Disease Research Center, University of Pittsburgh, Pittsburgh, PA, USA.

Late-onset Alzheimer's disease (LOAD) is a multifactorial neurodegenerative disorder resulting in loss of cognitive and executive function and ultimately death. It currently affects over 5 million individuals in the United States alone. To date, ten loci (*APOE, BIN1, CLU, EPHA1, PICALM, CD2AP, CD33, ABCA7, MS4A4A/MS4A6E*, and *CR1*) have been implicated as risk loci for LOAD. Despite a number of associated variants located in these loci, the identified associations show only modest effect sizes, with *APOE* as the sole producer of a considerable odds ratio. Given its complex nature, LOAD is a disease about which we stand to gain substantial insight by examining how these variants may alter transcription and subsequently, pathology and phenotype. This study aims to effectively utilize the newly developed RegulomeDB (http://regulomedb.org) to achieve a better understanding of the potential regulatory function of previously published genome-wide significant SNPs in these nine loci (excluding *APOE*), as well as other suggestive GWAS loci (total 31 top SNPs). We used Broad Institute's SNP Annotation and Proxy search (SNAP - http://www.broadinstitute.org/mpg/snap/) tool to identify 140 SNPs in linkage disequilibrium ($r^2 \geq 0.8$) with the 31 previously published LOAD-associated SNPs. We then examined all 171 SNPs for potential regulatory function using RegulomeDB and found 14 with scores of 2b or lower, indicating a high likelihood of affecting gene expression. Half of these SNPs were located in the *MS4A* region, including the SNP with most evidence for regulatory function (rs667897, RegulomeDB score=1b). Data included in RegulomeDB indicates that this intergenic SNP serves as an eQTL for *MS4A4A* and mediates binding of transcription factors through TCF11:MafG and NFE2L2 motifs. Variants in the regions of *CD2AP, CLU, ABCA7*, and *PICALM* showed low RegulomeDB scores (2a or 2b), suggesting good evidence for affecting gene expression. *SLC28A1* (Regulome DB score=1f) and *HRK/RNFT2* (Regulome DB score=2a), two suggestive LOAD loci, were also among the top 14 SNPs with strong evidence for regulatory function. Notably, all 14 SNPs with suggested regulatory function fall in intronic or intergenic regions. Our results indicate further possibilities for explaining the associations between replicated SNPs and LOAD risk, especially for non-coding variants.

1245T

A functional variation in the *CHRNB3* promoter affects Parkinson's disease risk and smoking. *A. Bar-Shira[1], M. Gana-Weisz[1], Z. Gan-Or[1], E. Giladi[1], N. Giladi[2,3], A. Orr-Urtreger[1,3].* 1) Genetic Institute, Tel Aviv Sourasky Medical center, Tel Aviv, Israel; 2) Movement Disorders Unit, Parkinson Center, Department of Neurology, Tel-Aviv Sourasky Medical Center, Tel-Aviv, Israel; 3) Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel.

Introduction: Parkinson's disease (PD) is affected by various genetic alterations, environmental factors and aging. Interestingly, cigarette smoking is being considered as a protective factor, and it was suggested that nicotine and nicotinic-acetylcholine receptors (nAChRs) may play a role. Since the β3 nAChR subunit is depleted in the striatum of PD patients, we aimed to test the possibility that variations in the *CHRNB3* gene, which encodes this subunit, are involved in PD. **Methods:** *CHRNB3* was sequenced in 100 PD patients. The alteration that was found in the putative promoter of this gene was further analyzed in a cohort of 596 PD patients and 369 controls. Its effect on promoter binding and activity was studied in cellular models using Chromatin immunoprecipitation and Luciferase assays. **Results:** The minor G allele frequency was 0.31 and 0.26 among patients and controls, respectively (p=0.02), and carriers had an OR of 1.33 (95% CI=1.03-1.73) for PD. In addition, the minor allele was strongly associated with smoking in patients, as 48.4% of carriers reported smoking history compared to 32.6% of non-carriers (p<0.0001). The transcription factor Oct-1 binding was almost eliminated in lymphoblasts homozygous for the minor G allele, to only 6.5% percent of the binding in cells with the major A allele. Furthermore, the *CHRNB3* promoter activity was reduced by 70%-96% in cells homozygous for the minor allele. **Conclusions:** The association between the *CHRNB3* variation and smoking in PD patients and its functional role in the promoter may suggest a molecular link between PD, smoking and nAChRs. These findings may also raise the possibility that nicotine treatment in PD should involve genotyping and personal adjustment.

1246F

Identification of microRNA expression quantitative trait loci in the nucleus accumbens of human postmortem brains from alcohol dependent subjects and matched controls. *M. Mamdani, V. Williamson, G. McMichael, B. Riley, K. Kendler, V. Vladimirov.* Virginia Institute for Psychiatric and Behavioral Genetics, VCU, Richmond, VA.

Alcohol dependence (AD) is a chronic addiction disorder with heritable factors accounting for 60% of the risk. While genetic studies have identified numerous AD loci, the mechanisms underlying neuroadaptations to excessive alcohol consumption are unclear. MicroRNAs (miRNAs), a species of non-coding RNA, are abundantly expressed in the brain and predominantly function by down-regulating gene expression. Involved in normal brain development and function, miRNAs are also implicated in several neuropsychiatric disorders; however, only one study has assessed miRNA expression in human AD subjects and in the prefrontal cortex only. Here, we evaluate miRNA expression and function in the nucleus accumbens (NAc), a major addiction-related brain region within the mesocorticolimbic pathway that is involved in drug-seeking motivation and reward. Expression quantitative trait loci (eQTL) are genomic loci that can regulate miRNA and/or mRNA expression; however, no such studies have been conducted in postmortem AD brains. We hypothesize that genetic risk loci for AD will also impact the expression of those miRNAs and mRNAs differentially expressed in the NAc of AD subjects and controls. In our preliminary results we detected 240 differentially expressed miRNAs at p<0.05; after multiple testing correction at a false discovery rate (FDR) <10%, 29 miRNAs remained significant. Within this list were miRNAs reported to be involved in aging/neurodegeneration (hsa-miR-1538, -516, -34c, -487a) and neurodevelopment/function (hsa-miR-371, -154, -1247). Interestingly, hsa-miR-154 is also reported to be associated with mesocorticolimbic pathway modulation in opiate addiction. To better understand the mechanisms contributing to the miRNA expression we identified AD-associated eQTLs affecting miRNA expression. We detected 275 cis-eQTLs (within 1 megabase) for all 29 miRNAs (FDR <10%). Several of these cis-eQTLs are located in functionally relevant sites including transcription factor binding sites, splicing enhancers/silencers, and putative miRNA target sites. To further explore the specific mechanisms driving AD, mRNA expression data derived from the same sample will be integrated with the genetic and miRNA data to **1)** identify miRNA/mRNA AD-specific interactions and **2)** detect eQTLs mediating these interactions. Subsequently, weighted gene co-expression network analyses (WGCNA) will be performed to identify specific gene network modules under the control of AD-implicated miRNAs.

1247W

Behavioral and neurochemical characterization of mutant mice lacking Lphn3, a gene implicated in ADHD and addiction. *D. Wallis[1], C.A. Orsini[2], B. Setlow[2].* 1) Biochemistry and Biophysics, Texas A&M University, College Station, TX; 2) Department of Psychiatry McKnight Brain Institute University of Florida College of Medicine Gainesville, FL 32610.

The latrophilin 3 (Lphn3) gene has been linked to susceptibility to attention deficit hyperactivity disorder (ADHD) and vulnerability to development of addiction. This suggests that this gene may be a genetic biomarker of these disorders and may lead to more selective therapeutic targets for treating them. However, little is known about the function of this gene. To characterize the function of the Lphn3 gene, we generated mutant mice. We then performed neurochemical, pharmacological, and behavioral assays to assess Lphn3 function. Four to 6 week old male Lphn3 mutant mice had significantly higher dopamine and serotonin levels in the dorsal striatum than their wild type (WT) counterparts. Given that elevated striatal dopamine is associated with locomotor hyperactivity in other mouse mutant lines, we evaluated locomotor activity in an open field arena. As expected, Lphn3 mutant mice displayed significantly more horizontal activity than both WT and heterozygous mice. Additionally, when Lphn3 mutant mice were administered acute i.p. injections of cocaine (20 mg/kg), they showed a significant elevation in locomotor activity relative to WT mice. In a separate cohort of male and female mice, we explored the contribution of the Lphn3 gene to reward-seeking behavior by assessing instrumental responding (lever pressing) for food pellet rewards under various fixed ratio (FR) schedules of reinforcement. Mice were first trained on a FR1 schedule for five 30 min sessions, after which they were tested on FR3, FR10, FR20 and FR40 schedules (one schedule/session). Lphn3 mutant mice displayed significantly greater instrumental responding for food than WT mice, particularly under high response ratios. Finally, we evaluated performance on rotarod assays and found a statistically significant difference between WT and null Lphn3 females with mutants having decreased latency to fall. Further, the WT mice were more likely to do a passive rotation on the rod, and the mutants were more likely to fall off. Together, these findings are consistent with a role for Lphn3 in regulating behavior, and show that a loss of gene function results in increased reward-seeking, possibly via enhanced striatal monoamine signaling. Current work is focused on characterizing Lphn3 mutant mice in additional behaviors, such as impulsivity and learning, and investigating the neurobiological consequences of the Lphn3 mutation.

1248T

Induced Pluripotent Stem Cells to Model Tourette's Syndrome. *L. Deng[1,2], N. Sun[1,2], Z. Pang[5], RA. King[6], M. Sheldon[1,2,3], JC. Moore[1,3], RP. Hart[3,4], M. Konsolaki[1,2], M. State[7], GA. Heiman[1,2], JA. Tischfield[1,2,3].* 1) Dept Genetics, Rutgers Univ, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Piscataway, NJ; 3) RUCDR Infinite Biologics NIMH Stem Cell Resource, Piscataway, NJ; 4) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ; 5) Child Health Institute of New Jersey, New Brunswick, NJ; 6) Yale Child Study Center, Yale University School of Medicine, New Haven, CT; 7) Department of Psychiatry , UCSF, San Francisco, CA.

Much of our current knowledge about the central nervous system (CNS) and neural function in patients with neuropsychiatric disorders has been obtained from relatively degraded postmortem brain. The inability to sample live CNS tissues impedes our progress to understand possible alterations in gene function and neuropathological abnormalities that develop during the course of the disorder. The rapid growth of iPSC technology has turned somatic cells into multipurpose basic and clinical research tools and opened new windows for modeling human diseases. Tourette syndrome (TS) is a childhood onset neurodevelopmental and genetic disorder characterized by chronic motor and phonic tics. Recent studies have indicated that the histaminergic neurotransmission pathway plays a role in the etiology of TS (Fernandez TV et al., Biol Psychiatry 2012). A previous study of a densely affected two-generation TS pedigree identified a rare segregating nonsense mutation, Hdc W317X, in the l-histidine decarboxylase (Hdc) gene (Ercan-Sencicek et al., NEJM 2010). The causal connection between a reduction in HDC activity and the symptoms of TS remains unclear. We collected whole blood samples from three members of this family and established induced pluripotent stem cell (iPSC) lines from each individual's T cells. Further differentiation of HDC mutant family's iPSCs into histaminergic neurons will enable us to model TS with regard to the histamine and other brain pathways. We will determine if partial loss of HDC function affects the expression of other genes and the metabolism/phenotype of these neurons. Comparison studies of the neurons from TS patients to the unaffected members within this family may lead to a better understanding of the TS phenotype.

1249F

Forgetting in *C. elegans* is regulated via Musashi (*msi-1*) mediated translational control of the Arp2/3 complex. *N.O. Hadziselimovic[1,2], F. Peter[1,2], P. Hieber[1,2], V. Vukojevic[1,2], P. Demougin[1,2], D.J.F. de Quervain[1,3], A. Papassotiropoulos[1,2,4], A. Stetak[1,2,4].* 1) University of Basel, Department of Psychology, Division of Molecular Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 2) University of Basel, Biozentrum, Life Sciences Training Facility, Klingelbergstrasse 50/70, 4056 Basel, Switzerland; 3) University of Basel, Department of Psychology, Division of Cognitive Neuroscience, Birmannsgasse 8, 4055 Basel, Switzerland; 4) University of Basel, University Psychiatric Clinics, Wilhelm Klein-Strasse 27, 4055 Basel, Switzerland.

In order to maintain a highly flexible nervous system, not only learning and generation of memory but forgetting is also essential to eliminate unnecessary memories to allow adaptation to a constantly changing environment.

Using *C. elegans* as a model organism, we recently performed a targeted candidate-gene based screen and found that the highly conserved RNA-binding protein musashi (*msi-1*) mediates forgetting. MSI-1 is expressed in neurons and tissue specific rescue in *C. elegans* demonstrates that the MSI-1 function is exclusively necessary in the AVA command interneuron. Among the previously identified MSI mRNA binding partners in vertebrate cells, ACTR2 is one subunit of the Arp2/3 protein complex serving as a nucleation core for the branching of the actin cytoskeleton. Using immunoprecipitation and subsequent RT-qPCR, we found that mRNAs of the *arx-1, -2* and *-3* subunits of the Arp2/3 complex associate with MSI-1 *in vivo*. We also show that the protein levels of ARX-1, ARX-2 and ARX-3 are downregulated upon associative learning, that is mediated by the translational inhibitory activity of MSI-1. Using genetic epistasis, we establish a link between MSI-1 and the levels and activity of the Arp2/3 protein complex. We show that activation of the Arp2/3 protein complex through expression of a constitutive active form of *wsp-1* in the AVA interneuron can block forgetting similar to deletion of the *msi-1* gene. The role of *msi-1* on forgetting is also reflected in *msi-1* mutants by the strong and persistent consolidation of GLR-1 containing synaptic size increase induced by associative learning. Finally, we demonstrate that GLR-1 signaling regulates both actin capping through the activation of adducin (*add-1*) and inhibition of the Arp2/3 complex mediated actin branching by *msi-1* regulated translational repression and both mechanisms act in concert to establish the proper memory. Thus, our results demonstrate that MSI-1 induces forgetting and represent a novel mechanism of memory regulation linking translational repression to regulation of the structure and complexity of the actin cytoskeleton in neurons.

1250W

Transcriptional impact of *POLR3A* and *POLR3B* mutations in Pol III-related hypomyelinating leukodystrophies. *K. Choquet[1], S. Durrieu[2], G. Boldina[2], M-J. Dicaire[1], D. Forget[3], K. Guerrero[4], B. Coulombe[3], M. Teichmann[2], G. Bernard[4], B. Brais[1].* 1) Neurogenetics of Motion Laboratory, Montreal Neurological Institute, McGill University, Montreal, Canada; 2) INSERM U869, Institut Européen de Chimie et Biologie, Université Bordeaux Segalen, Bordeaux, France; 3) Institut de Recherches Cliniques de Montréal (IRCM), Montreal, Canada; 4) Pediatric Neurodegenerative Laboratory, Department of Pediatrics, Neurology and Neurosurgery, Division of Pediatric Neurology, Montreal Children's Hospital, McGill University Health Center, Montreal, Canada.

Leukodystrophies are a heterogeneous group of neurodegenerative diseases characterized by abnormal white matter in the central nervous system (CNS). It has been established that mutations in *POLR3A* and *POLR3B* are an important cause of five clinically overlapping hypomyelinating leukodystrophies (HML). *POLR3A* and *POLR3B* encode the two catalytic subunits of RNA Polymerase III (Pol III), which synthesizes over 200 non-coding RNAs including transfer RNAs (tRNAs). We hypothesized that mutations in *POLR3A* and *POLR3B* alter Pol III's transcription efficiency, leading to a decrease in the expression of key transcripts for the development and maintenance of CNS white matter. We designed a custom microarray with probes for all known Pol III transcripts to compare expression levels between fibroblasts of four *POLR3A*-confirmed cases and four controls. A relatively small group of transcripts were found to be downregulated (n=21), of which an even smaller number (n=5) were found to be more expressed in brain compared to other tissues. We further validated their downregulation by performing quantitative RT-PCR and Northern Blots in a larger cohort composed of *POLR3A*- and *POLR3B*-confirmed cases as well as healthy controls. We also isolated Pol III from patients and controls fibroblasts' nuclei and performed in vitro transcription assays to evaluate Pol III's transcriptional efficiency. Deregulated Pol III targets are being further investigated to determine whether impairment of Pol III transcription is indeed central to the pathophysiology of Pol III-related HML. Elucidating the molecular mechanisms responsible for this disease is essential to develop potential therapeutic approaches.

1251T

The NINDS Repository: A unique resource of patient-derived, highly characterized primary fibroblasts and induced pluripotent stem cells for neurodegenerative disease research. *C.A. Pérez[1,4], S. Heil[1,4], S. Gandre-Babbe[2,4], C. Rhoda[2,4], K. Panckeri[2,4], M. Self[1,4], K. Hodges[3,4], M. Shutherland[5], R.A. Corriveau[5], C. Tarn[1,4].* 1) NINDS Repository; 2) Stem Cell Biobank; 3) Cell Culture Laboratory; 4) Coriell Institute for Medical Research, Camden, NJ; 5) National Institute of Neurological Disorders and Stroke (NINDS), Bethesda, MD.

Induced pluripotent stem cells (iPSCs) reprogrammed from patient-derived primary fibroblasts have become an increasingly utilized resource for the study of human disease and have proven especially valuable in studying neurodegenerative disorders for which disease models are difficult to establish. The National Institute of Neurological Disorders and Stroke (NINDS) Repository is a public resource established in 2002 aiming to provide a centralized and open collection of biological samples (DNA, lymphoblastoid cell lines, fibroblasts, iPSCs, biofluids such as plasma, serum, cerebrospinal fluids, and urine) and associated de-identified clinical data from a diverse population of affected patients and normal controls. Since 2011, the NINDS Repository has added to its web-based catalog (http://ccr.coriell.org/NINDS) more than 40 iPSC and 130 fibroblast lines. Most iPSC lines are contributed by investigators from a Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS) and Huntington's disease (HD) NINDS-sponsored consortia funded by the American Reinvestment and Recovery Act (ARRA). Additional iPSC lines have been also submitted by other investigators. To ensure the quality of these valuable resources, all iPSCs and fibroblasts submitted to the NINDS Repository by iPSC Consortia and other investigators undergo rigorous quality assessments (viability, pluripotency, karyotyping, differentiation status, gene expression analysis, sterility) prior to distribution by the NINDS repository. The results are summarized in a Certificate of Analysis displayed on the web-based catalog along with the recommended culturing protocol. The NINDS Repository fibroblast and iPSC collections include mostly cell lines bearing specific genetic mutations associated with PD, ALS, HD, frontotemporal degeneration, or Alzheimer's disease, as well as samples derived from neurologically normal controls. For certain affected individuals, the parental fibroblast, corresponding iPSC line, and whole blood DNA are available. It is important that all related biospecimens from the same individual, and their specific characterization details describing publicly available iPSCs and fibroblasts be documented and clearly presented to potential requestors. The NINDS Repository serves as a unique and effective centralized resource for these iPSCs and fibroblasts and their critical phenotypic data, to basic and applied research investigators worldwide.

1252F

Understanding the Risk Pathway from *GABRA2* to Alcoholism: Effects of Personality, Brain, Development and Gene x Environment Interactions. *M. Burmeister[1,2], S. Villafuerte[1,2], E.M. Trucco[2], M. Heitzeg[2], R.A. Zucker[2].* 1) Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI; 2) Department of Psychiatry, University of Michigan, Ann Arbor, MI.

SNPs within *GABRA2* have long been known to be associated with alcoholism. To understand the path from a genetic variant to complex behavior, alcohol abuse, we used the Michigan Longitudinal Study (MLS), consisting of ~463 families who were recruited >20 years ago with a 3-5 year old boy, his siblings and parents. ~70% of fathers met criteria for alcoholism, plus neighborhood controls. Assessments were every 3 years, teenagers every year. Measures considered here include impulsiveness in adults (NEO-PI-R), fMRI during a modified incentive monetary delay task, life-time alcohol problems, youth externalizing behavior (Young Self Report) and parental monitoring. *GABRA2* genotypes were tested for association by regression analysis. Mediation was tested using bootstrapping procedures in AMOS. Growth mixture modeling (GMM) was used to identify externalizing trajectory classes. Class labels were regressed on GABRA2, parental monitoring, and their interaction using multinomial linear regression via GMM. Most *GABRA2* SNPs were in strong LD and formed two major haplotypes. The G (minor) allele of rs279827 represents the risk haplotype with regard to alcoholism. We identified three trajectory classes of externalizing behavior across adolescence, a low, a developmentally limited, and a high risk class. Parental monitoring but not genotype predicted lower levels of externalizing behavior. We observed a significant *GABRA2* x parental monitoring interaction effect on trajectory class membership: While A-carriers' trajectory class membership was largely unaffected by parental monitoring, those with the risk (G) genotype were affected by parental monitoring, both positively and negatively. These and our previously published results demonstrate that subjects who carry the risk haplotypes of *GABRA2* 1) are more impulsive, which mediates the association of *GABRA2* SNPs with alcohol problems, 2) their brains (especially the insula) react stronger when anticipating a reward, and, 3) when not monitored by parents, are more likely to engage in problem behaviors ('act out') during adolescence, and later become more likely to have alcohol problems. Our study demonstrates how understanding of the path from SNP to final phenotypic outcome can be achieved in smaller, well characterized longitudinal samples, and how genetic findings can be pulled into the existing psychosocial literature, merging studies of nature with nurture to explain some aspects of human behavior.

1253W

Investigation of maternal stress among women who carry the FMR1 premutation who are mothers of a child with fragile X syndrome: involvement of the corticotrophin-releasing hormone receptor 1 gene to impact maternal well-being. *J.E. Hunter[1], A. Lori[1,3], D. Hamilton[1], L. Shubeck[1], A. Abramowitz[2], J.F. Cubells[1,3], S.L. Sherman[1].* 1) Dept Human Genetics, Emory Univ, Atlanta, GA; 2) Dept Clinical Psychology, Emory Univ, Atlanta, GA; 3) Dept Psychiatry and Behavioral Sciences, Emory Univ, Atlanta, GA.

The FMR1 contains a polymorphic CGG repeat which, once unstable, is capable of expansion across generations with maternal transmission. Women who carry a premutation allele of FMR1, defined as an allele with 55-199 repeats, are at risk of passing on a full mutation allele, defined as an allele with >200 methylated repeats, to their offspring. Methylated full mutation alleles result in the intellectual disability disorder, fragile X syndrome (FXS). We recently published a study among mothers of children with FXS who carry a premutation showing evidence that polymorphisms within the corticotrophin releasing hormone receptor 1 gene (CRHR1), a major regulator in the endogenous cortisol response to stress, moderated the impact of maternal stress on the severity of distress experience in social situations as measured by the Social Phobia and Anxiety Inventory (SPAI). The goal of the current study is to follow up on these results to determine the severity and sources of maternal stress associated with raising a child with FXS involved in this interaction. Preliminary results for 25 mothers indicate that the interaction between maternal stress and CRHR1 polymorphisms is associated with both child characteristics (Aberrant Behavior Checklist: Irritability subscale, $p=0.04$; Social Avoidance subscale, $p=0.02$) and the impact of having a child with an intellectual and developmental disability on the family (Questionnaire on Resources and Stress: Parent and Family Problems subscale, $p=0.01$). The results of this study will provide potential targets for maternal interventions to alleviate stress and maternal well-being.

1254T

Advanced paternal age is associated with earlier onset of schizophrenia in the affected siblings of multiplex families. *W.J. Chen[1,2,3,4], S.H. Wang[1,2], C.K. Hsiao[1,2], C.M. Liu[3,4], H.G. Hwu[1,3,4].* 1) Institute of Epidemiology & Preventive Medicine, National Taiwan Univ, Taipei, Taiwan; 2) Genetic Epidemiology Core Laboratory, Center of Genomic Medicine, National Taiwan University; 3) Institute of Brain and Mind Sciences, College of medicine, National Taiwan University; 4) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University.

Introduction: Advanced paternal age has been reported to increase the risk of schizophrenia. Early onset, an important component phenotype for schizophrenia, may disrupt the normal processes of maturation of the brain and cause neurodevelopmental deviance. This study aimed to explore if advanced paternal age is associated with not only increased risk of schizophrenia but also earlier onset of schizophrenia in multiplex families. Methods: A total of 1359 affected siblings with schizophrenia and 308 healthy siblings from 694 multiplex families recruited throughout Taiwan were included for this study. Probands and their first-degree relatives were interviewed using the Diagnostic Interview for Genetic Studies. Offspring's paternal age was calculated as its father's age minus the offspring's age at the time of recruitment. We compared the paternal age of affected siblings with their counterparts of healthy sibling within each family using random-effect models to control for familial dependence. Then we compared the distribution of onset age of schizophrenia for eight paternal age groups (<20, 20–25, 25–30, 30–35, 35–40, 40–45, 45–50, and ≥ 50). Results: Increasing 1 year in paternal age was associated with an increased risk of schizophrenia, with an adjusted odds ratio of 1.12 (95% confidence interval 1.03-1.22). Using paternal age of 20–25 as the reference, advancing paternal age was associated with younger age at onset of schizophrenia in the affected siblings. On average, the onset age was lowered for 1.5 years for the paternal age of 25–30 and 5.5 years for the paternal age of ≥ 50, with a P value of 0.04 for the trend test. Conclusion: Advanced paternal age was associated with an increased risk of schizophrenia in the offspring as well as earlier onset of schizophrenia in the affected siblings. These findings derived from multiplex families of schizophrenia imply that the influence of advanced paternal age may be due to that fathers bequeath more mutations as they age.

1255F

Parents and teachers report on different aspect of children's and adolescent's conduct disorder and hyperactivity/inattention behavior.
X.W. Zhang[1,2], P.C. Sham[2,3], S.S. Cherny[2,3], H.Q. Meng[4], Y.X. Fu[4], Y. Huang[1], T. Li[1]. 1) Psychiatry, West China Hospital, Sichuan University, China; 2) Psychiatry, Hong Kong University, Hong Kong; 3) Genome Research Centre, Hong Kong University, Hong Kong; 4) Mental Health Center, First Affiliated Hospital, Chongqing Medical University, China.

Background: Either conduct disorder or hyperactivity/inattention problem poses considerable burden on health care and education, the co-occurrence bring more pressure and impairments for the children and their family. However, the etiology of overlapping is still unclear, and one of the reasons is assessment inconsistent in different cultural background. Method: Subjects were 433 twin pairs aged between 6 and 16 years from Prospective Twin Registry in Southwestern China, whose parent and teacher completed the Strengths and Difficulties Questionnaires. It both contained estimation on twins' conduct disorder and hyperactivity problem from different point of views. And then used the structure equation model to explore the relationship between conduct disorder or hyperactivity/inattention problem in children and adolescent. Results: Both in the bi-traits twin model and bi-raters twin model, biometric model was the best fitting one. According to this model, conduct disorder and hyperactivity behavior were correlated with phenotypic correlation 0.45-0.65, and genetics factors contributed majority (around 70%) to the covariance both for boys and girls on parent's view, whereas environmental also played important (around 55%) in teacher's view. Meanwhile, the results between parent's and teacher's information were not that consistence. The phenotypic correlation between two informants was less than 0.2. In bi-raters twin model fitting, rater bias model was rejected. in the best fitting model, the common environmental effect always contributed more on variance from teacher's point of view than from parent's point of view, so I alluded that they maybe observe the children in different occasions, or evaluate them without fixed reference group. Conclusion: conduct disorder and hyperactivity behavior are correlated with each other, substantially contributed by genetic factors, but not governed by the common specific underlying phenotype. The information from parents and teacher are disparity, and it suggests us to collect data from more than one informant when using questionnaire to evaluate conduct disorder or hyperactivity/inattention problem in China.

1256W

Association of *APOE* Polymorphism and Stressful Life Events with Dementia in the Pakistani Population.
M. Chaudhry[1], S. Hasnain[1], B. Snitz[2], X. Wang[3], D. Winger[4], L. Wang[4], S. Rosenthal[3], F.Y. Demirci[3], M.I. Kamboh[3]. 1) Department of Microbiology and Molecular Genetics, University of the Punjab, Lahore, Pakistan; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA; 4) Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, USA.

Dementia is a major public health problem worldwide. Alzheimer's disease (AD) is a major form of dementia and the *APOE*4* allele is an established genetic risk factor for AD. Similarly, stressful life events (SLE) are also associated with dementia. The objective of this study was to examine the association of *APOE*4* and SLE with dementia in a Pakistani population, which to our knowledge has not been reported previously. We also tested for an interaction between SLE and *APOE*4* and the risk for dementia in this sample. A total of 176 subjects (61 cases and 115 controls) were recruited for this study. All pre-diagnosed cases and healthy controls were then interviewed to assess cognition, co-morbidities, history of SLE and possible differences in demographics. Blood samples were also drawn and genotyping for the *APOE* polymorphism (E2/E3/E4) was performed. The *APOE*4* and stressful life events were each independently significantly associated with the risk of dementia. The odds ratios (ORs) for *APOE*4* carriers and SLE were 2.81 (95%CI: 1.26-6.21; P=0.011) and 1.008 (95% CI: 1.004-1.012; P=1.15E-05), respectively. The gender stratified analysis revealed that *APOE*4* and SLE were independently associated with dementia in males but not in females. However, we did not find a significant interaction between *APOE*4* carrier status and stressful life events in affecting the risk of dementia (P=0.677). Although the sample size of this study was small, the established association of *APOE*4* with dementia was confirmed the first time in the Pakistani population. Furthermore, SLE was also found to be a significant predictor for dementia in this population. Our study also emphasizes the need to improve mental health facilities for older people in Pakistan and to facilitate future dementia research.

1257T

Association between MAOA and aggressive behavior in adolescents receiving the pharmaceutical treatment lisdexamfetamine dimesylate for ADHD symptoms.
K.A. Nelson, M.S.[1], A. Stoker[2], P. Huizenga, H.T.(ASCP), rLAT[1], S. Weaver, R.N.[1], T. Jung[1], E.A. Ehli, R.N., M.S.[1, 3], T.J. Soundy, M.D.[3], K. Bohlen, PharmD[1, 3], Y. Hu, Ph.D.[1, 3], G.E. Davies, Ph.D.[1, 3]. 1) Avera Institutue for Human Genetics, Avera McKennan Hospital and University Health Center, Sioux Falls, SD; 2) University of South Dakota, Sanford School of Medicine, Sioux Falls, South Dakota; 3) Department of Psychiatry, University of South Dakota, Sanford School of Medicine, Sioux Falls, South Dakota.

Attention deficit hyperactivity disorder (ADHD) is one of the most common childhood neuropsychiatric disorders, affecting 5.29% of children worldwide. ADHD symptoms manifest as severely disruptive behaviors of inattention, hyperactivity, and/or impulsiveness. The pharmaceutical lisdexamfetamine dimesylate (LDX [Vyvanse®]), a prodrug stimulant prescribed for ADHD, is thought to alleviate symptoms by inhibiting the dopamine and norepinephrine reuptake pathways. Monoamine oxidase A (MAOA) is an X-chromosome linked gene that catalyzes the degradation of dopamine and norepinephrine. The number of 30-bp repeats located in the polymorphic MAOA promoter region has been shown to alter the enzymatic degradation of dopamine and norepinephrine. Wildtype alleles consist of 3.5 or 4 repeats (high enzymatic activity alleles). The genotypes of 2, 3, or 5 repeats (low enzymatic activity alleles) have been associated with increased aggression in males. This study looked at adolescents being treated for ADHD symptoms with LDX. Nearly half of the study individuals discontinued the medication due to aggressive behavior. Here, we hypothesized that individuals with the low enzymatic activity MAOA alleles would be the individuals more likely to discontinue the drug due to aggressive behavior than individuals with the higher MAOA enzymatic activity genotypes. The study sample included 73 adolescents averaged 12 (6–18) years in age. The final study sample was 85% male and 90% Caucasian. All individuals were being treated within a behavioral health outpatient facility by licensed child psychiatrists for ADHD symptoms. After the ADHD diagnosis, each individual at one time during their treatment, received LDX medication. Each child's buccal cell DNA was used to perform the genotyping of the MAOA 30-bp polymorphism. We found that individuals with high MAOA enzymatic activity alleles were significantly more likely to discontinue LDX due to aggression than individuals with the low enzymatic activity MAOA alleles (OR=0.3083 [95% CI: 0.11, 0.86] p=0.02). Stimulant medications are the first-line pharmacological treatment for children with ADHD. Although these treatments are very effective for the majority of its users, there are patients who can experience considerable adverse side effects. This study has shown a significant association between the discontinuation of LDX due to aggressive behaviors in adolescents with the high MAOA enzymatic activity alleles (3.5 and 4).

1258F

Psychosocial factors are correlated with gender-sensitive differences in relative telomere length (RTL) in Han Chinese university undergraduates.

O. Yim[1], M. Monakhov[2], X. Zhang[3], P.S. Lai[1], S.H. Chew[2], R.P. Ebstein[4]. 1) Pediatrics, National University Singapore, Singapore; 2) Dep't. Economics, National University Singapore, Singapore; 3) Business School, National University of Singapore, Singapore; 4) Psychology Dep't., National University of Singapore.

Telomeres (TTAGGG repeats) cap the ends of chromosomes and numerous studies suggest that telomere length (TL) is a reliable index of cellular and tissue aging. More recently social stress has been shown to predict decreased TL but much remains unknown about the psychosocial factors, especially in early adulthood, that modulate TL. To provide a better understanding of factors that influence TL and by implication aging and health, leukocyte DNA from 991 NUS Chinese undergraduates (52.4% Female) were measured for Relative TL (RTL) and compared to house-keeping β-Hemoglobin gene using the method of Cawthon (2009). A significant difference (T-Test P = 0.0001 ; 989 d.f) in RTL by gender was observed and women, as previously reported, show greater RTL than male students. For women only, increased risk proneness over moderate gain, measured by incentivized behavioral economic tasks, was associated with shorter RTL (coeff. =-0.0102; P = 0.040). Delay discounting was also associated in women only with reduced RTL (coefficient =-0.0089; P = 0.010). Impatient women have shorter telomeres. We also observed a relationship in women between RTL and scores on the NEO-PI-3 Neuroticism-Anxiety facet (coeff.= 0.0072, p=0.045). Higher anxiety scores predict greater RTL in women. Intriguingly, for males and not females, students from families with incomes below the unofficial poverty line are characterized by reduced RTL (T-test P = 0.022). Again in males only, NEO-PI-3 Agreeableness scores are associated with greater RTL (coeff.=0.0208, p=0.017). Agreeableness is a personality trait suggesting the individual is kind, sympathetic, cooperative, warm and considerate. The current study examined the role of risk attitude, personality and socio-economic status on aging at the cellular level indexed by RTL. These psychosocial and behavioral characteristics appear to affect RTL but differently in men and women. Such gender-sensitive responsiveness in RTL to psychological and social environmental cues, that are already present in early adulthood in a non-clinical population, are suggested to have important implications for differential ageing in men and women over the course of the lifespan.

1259W

A Translational Approach for Cocaine Abuse: From Man to Mouse to Man.

R.F. Clark, E. Solano, M.J. Edlund, N.R. Garge, E.O. Johnson. RTI International, Research Triangle Park, NC.

Three categories of factors contribute to the development of cocaine addiction: 1) environmental factors, including cues, conditioning, and external stressors, 2) drug-induced factors, which lead to neurobiological changes resulting in altered behaviors, and 3) genetic factors, accounting for approximately half the risk of developing an addiction. Drug addiction cannot be adequately addressed solely within a single discipline and instead requires a more transdisciplinary approach.

In our methodology, we first used mouse system genetics to identify genes, gene networks and pathways associated with cocaine dependence. We used in silico systems genetics tools on previously identified candidate genes for cocaine abuse phenotypes in humans to initiate these mouse systems genetics studies. Using homologous mouse genes to these human candidate genes, we identified several mouse genes that were strong candidates for expression-QTL mapping of gene regulation in the hippocampus and nucleus accumbens, including Arrb2, Chrna5, Comt, and Oprm1. Global mapping revealed several networks of genes whose expression was correlated with these genes associated with cocaine addiction. In addition, we identified gene networks associated with multiple cocaine abuse phenotypes in mice, such as differential cocaine response to locomotion in an open field. We were able to assemble several pertinent genetic pathways linked to these genes associated with cocaine addiction.

To build a model of cocaine abuse and addiction, the gene networks and pathways were added to an ontology-based network model of cocaine addiction. This network model was built from the integrated expertise from our team members in diverse domains of knowledge that are suspected to be causally related to cocaine addiction, including known environmental factors, such as drug availability, social stressors, peer support, and environmental exposures. This model can be utilized to provide a framework for future cocaine-addiction studies by supporting interdisciplinary research with graphical representation of complex scientific theories, highlighting potential interactions, facilitating meta-analyses across studies, and supporting statistical analysis and predictive modeling, as well as to translate these findings into the clinical setting for improvement of treatment of cocaine addiction.

1260T

The impact of the metabotropic glutamate receptor and other gene family interaction networks on the autism spectrum disorders.

D. Hadley, Z. Wu, C. Kao, A. Kini, A. Mohamed-Hadley, K. Thomas, L. Vazquez, H. Qiu, F. Mentch, R. Pellegrino, C. Kim, J. Glessner, H. Hakonarson, Autism Genome Project Consortium. The Children's Hospital of Philadelphia, Center for Applied Genomics, 3615 Civic Center Boulevard, Philadelphia, PA 19104.

As multiple defective duplicated genes have been discovered to contribute to the etiology of the autism spectrum disorders (ASDs), we prioritized pathways robustly enriched for structural defects of duplicated genes and their gene family interaction networks (GFINs) using a novel network permutation test. Across 6,742 patients with ASDs relative to 12,544 neurologically normal controls, the metabotropic glutamate receptor (GRM) GFIN was significantly enriched for structural defects (P <= 2.40E-09, 1.8-fold enrichment) in ASDs and previously observed to impact attention deficit hyperactivity disorder (ADHD) and schizophrenia. The MAX dimerization protein (MXD) GFIN was also significantly enriched (P <= 3.83E-23, 2.5-fold enrichment) and known to interact with the MXD-MYC-MAX network of genes and transcription factors that is dysregulated in various types of cancer. Furthermore, the calmodulin 1 (CALM1) gene interaction network was strongly enriched (P <= 4.16E-04, 14.4-fold enrichment) and known to regulate voltage independent calcium-activated action potentials at the neuronal synapse. Taken together, our genome-wide focus on GFINs illustrates the respective contributions of many different biological pathways in the etiology of ASDs.

1261F

The potential use of uncommon variants in GWAS data sets of pedigrees with autism spectrum disorders.

H.Z. Wang[1], H. Qin[1], K. Ahn[2], W. Guo[1], Y.Y. Shugart[1]. 1) Division of Intramural Research Program, National Institute of Mental Health, Bethesda, MD; 2) Childhood Psychiatry Branch, National Institute of Mental Health, Bethesda, MD.

GWAS data sets are often overlooked in rare-variant analysis, for the simple reason that DNA chips only sample a small subset of the genome. However with collapsing methods, rare-variants are linearly combined and analyzed as a group rather than individually, and rare variant analysis can further be enhanced using pedigree based methods, since variants rare in the population are enriched within families. We therefore explored the possibility of analyzing rare variants in GWAS data sets using such methods. To test this hypothesis we used FBAT-Rare, a pedigree-based collapsing method, and collapsed rare variants in genes using a frequency based weighing scheme. We applied this method to the Autism Genome Project (AGP) data set consisting of 2665 trios with autism spectrum disorder (ASD) afflicted offspring. We focused our efforts on CNTNAP2, a gene that encodes a member of the neurexin family which functions in the vertebrate nervous system as cell adhesion molecules and receptors. It is positively identified by multiple research teams, including Anney et al. (2012), the authors of the AGP data itself, to be strongly associated with ASD as well as other neurological development disorders. Using FBAT-Rare, we are currently able to detect an association of $p=6.37*10^{-5}$ for the European-only, high IQ male population for ASD, and $p=2.77*10^{-3}$ for strict autism. We realize the presented results do not meet the standard GWAS-wide significance level. However, our results appear to support the previous findings on CNTNAP2 reported by other investigators. More research work will focus on using several other analytical tools to re-analyze the AGP data set obtained from dbGap.

1262W

Genome-wide scan identifies candidate loci for gene-gene interactions affecting relationships between Aβ42, p-tau, and disease status. *M.T.W. Ebbert[1,2], P.G Ridge[1], K. Boehme[1], M. Bailey[1], C. Cruchaga[3], A. Goate[3], S. Bertelsen[3], C.D. Corcoran[4,5], T. Maxwell[6], J.S.K. Kauwe[1], Alzheimer's Disease Genetic Consortium.* 1) Department of Biology, Brigham Young University, Provo, UT; 2) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 3) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO; 4) Department of Mathematics and Statistics, Utah State University, Logan, Utah; 5) Center for Epidemiologic Studies, Utah State University, Logan, Utah; 6) Human Genetics Center, University of Texas, Houston, Texas.

Alzheimer's disease (AD) is complex and affects 35.6 million people worldwide. Aβ42 and p-tau levels in cerebrospinal fluid are key biomarkers for the presence of pathological features of AD, Aβ plaques and neurofibrillary tangles, respectively. Like any complex disease, however, there are individuals that exhibit brain pathology indicative of AD but never develop clinical symptoms. In fact, more than 25% of individuals exhibit such pathology but show no cognitive decline. Likewise, there is considerable variation in the rate of decline in AD patients. These remarkable observations and other poorly understood attributes of AD are likely the result of undiscovered complex gene-gene interactions. To discover candidate loci for gene-gene interactions driving AD we used genotype data from 2,726 subjects for whom CSF biomarker levels have been measured to test for relationship quantitative trait loci (rQTL) that modify the relationships between (1) Aβ42 and p-tau; and (2) Aβ42 and disease case-control status. These rQTL are likely involved in gene-gene interactions but may be overlooked in typical association studies because they often do not have marginal effects on the individual traits whose relationship they modify. Specifically, we hypothesize that genetic modifications that modify these well-characterized relationships exist and explain, at least in part, variation in disease progression and the numerous individuals exhibiting known AD brain pathology without exhibiting clinical symptoms. We hypothesize that genetic factors that modify these relationships are likely to alter disease risk or age of onset. Scanning for rQTL in genome-wide SNP data we identified 3 genome-wide significant SNPs that may modify the relationship between Aβ42 and p-tau, and 2 genome-wide significant SNPs that may modify the relationship between Aβ42 and case-control status. rs8027714 (intergenic between NPAP1 and SNRPN, $p < 1.7e-09$), rs1036819 (intron of ZFAT and ZFAT-AS1, $p < 4.2e{-}08$), and rs7844573 (intron of PSD3 and exon of RPL35P6 pseudogene, 9.7e–08) may modify the relationship between Aβ42 and p-tau while rs57216348 and rs59251601 (both in intron of CPLX2, $p < 2.881e{-}08$) may modify the relationship between Aβ42 and case-control status. We concluded that the reported SNPs affect the relationships between (1) Aβ42 and p-tau; and (2) Aβ42 and disease case-control status, and that they may play an important role in AD etiology.

1263T

Family-based Genome-wide Association Study of resting EEG identifies UROC1 and neurotransmitter biosynthetic process. *S. Kang, M. Rangaswamy, D. Chorlian, N. Manz, B. Porjesz, COGA collaborators.* State University of New York, Downstate Medical Center, Brooklyn, NY.

Electroencephalography (EEG) measures are highly heritable neuroelectrical correlates of resting brain state, and contribute several markers of risk for alcoholism and related disorders. Increased beta power has been found to be a hallmark of alcoholism and related disorders and also a marker of those at risk, and is an index of neural hyperexcitability. To identify genetic variants associated with EEG, we performed a family-based genome-wide association study using extended multiplex families densely affected by alcohol dependence. The advantage of the design of this study is robustness against population substructure. High beta power (20-28 Hz) of the frontal-central bipolar electrode pairs (F3-C3, Fz-Cz, F4-C4) were calculated using standard Fourier transform methods. Samples from the Collaborative Study on the Genetics of Alcoholism (COGA) were genotyped using the Illumina Human OmniExpress array on 118 families, densely affected by alcohol use disorder. Phenotype data was derived from multivariate linear regression models which were constructed from log transformed high beta power EEGs, controlling for log-transformed age and stratified by gender. After quality control procedures, association testing was done on 1,564 samples assuming an additive model using the generalized disequilibrium test. Pathway analysis was performed using ALIGATOR (Association List Go Annotator) to study groups of genes by testing for overrepresentation of members of those groups within lists of genes containing significantly associated SNPs from the GWAS. We found that a SNP in *UROC1* (rs1687482) was significantly associated with F3-C3 high beta power EEG at a genome-wide significant level ($p = 4.8 \times 10^{-8}$). The most significant individual GO category from the pathway analysis is a neurotransmitter biosynthetic process with significantly associated SNPs in *PAH* ($p = 1.2 \times 10^{-5}$). These results suggest *PAH* accounts for some of the variations in high beta power EEG. *PAH* catalyzes the conversion of phenylalanine to tyrosine, shares physical, structural and catalytic properties with tyrosine hydroxylase and tryptophan hydroxylase that catalyze the rate-limiting steps in the biosynthesis of the neurotransmitters dopamine, noradrenaline and serotonin. These findings underscore the utility of using EEG phenotypes to identify meaningful genetic correlates of resting brain state with pathophysiology of neuropsychiatric conditions.

1264F

Association between rs10520400 and Caudate Nucleus Asymmetry in Alzheimer's Disease and Mild Cognitive Impairment. *M.E. Renteria[1,2], P. Rajagopalan[3], S.E. Medland[1], S.K. Madsen[3], J.L. Stein[3], D.P. Hibar[3], N.G. Martin[1], M.J. Wright[1], P.M. Thompson[3], Alzheimer's Disease Neuroimaging Initiative (ADNI).* 1) Queensland Inst Medical Research, Herston, Queensland, Australia; 2) The University of Queensland, School of Psychology, St Lucia QLD 4072, Australia; 3) Imaging Genetics Center, Department of Neurology, Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA, USA.

In Alzheimer's disease (AD) and mild cognitive impairment (MCI), atrophy of the caudate nucleus appears in MR brain scans as reduction in volume and loss of normal asymmetry in patients compared to healthy controls. In the present study, we extracted caudate nuclei volumes of 731 individuals from the Alzheimer's disease Neuroimaging Initiative (ADNI) cohort. Using caudate asymmetry as a disease-associated quantitative trait, we asked whether genetic factors contribute to such inter-hemispheric differences in rates of neurodegeneration. We conducted a genome-wide association analysis of caudate asymmetry across >500,000 directly genotyped single nucleotide polymorphisms (SNPs) in two groups of patients diagnosed with either MCI (N=354) or AD (N=172), and a group of healthy elderly controls (N=204). SNP rs10520400, located in chromosome 4q34.3, was significantly associated in both patient cohorts ($P=6.26 \times 10{-}05$ in MCI, and $P= 1.77 \times 10{-}05$ in AD; combined $P= 4.51 \times 10{-}09$), but not in healthy elderly controls (P=0.13). Interestingly, rs10520400 had previously been reported as a cis expression QTL (eQTL) for the AGA gene, which participates in glycoprotein metabolism. Deficiencies in the AGA enzyme lead to aspartylglucosaminuria, a Mendelian condition that causes progressive decline in mental functioning. Moreover, AD-related neurodegeneration that leads to loss of normal caudate asymmetry may explain why SNP rs10520400 is specifically associated with AD and MCI patients, but not normal aging.

1265W

Scan for rare copy-number variation in 11,850 Swedish schizophrenia samples suggests novel susceptibility loci and provides evidence for convergence with regions of common variant association.
C. O'Dushlaine[1], J. Szatkiewicz[2], J. Moran[1], K. Chambert[1], A. Kähler[3], C. Hultman[3], P. Sklar[4], S. Purcell[1,4], S. McCarroll[1,5], P. Sullivan[2]. 1) Broad Institute, Boston, MA, USA; 2) UNC, USA; 3) Karolinska Institutet, Sweden; 4) Mount Sinai School of Medicine, USA; 5) Harvard Medical School, USA.

Background: Schizophrenia (SCZ) is a heritable disorder with substantial public health impact. Several rare copy number variants (CNVs) have been implicated in this disorder. However, prior studies indicate that larger samples are necessary for new discoveries to be made.

Methods: We conducted a genome-wide association study for schizophrenia with a Swedish national sample (4,719 cases with SCZ and 5,917 controls post-QC). All subjects from the Swedish Schizophrenia Study were born in Sweden and the schizophrenia (SCZ) cases were identified via the Swedish Hospital Discharge Register. DNA was extracted from whole blood and was genotyped at the Broad Institute using Affymetrix (5.0 and 6.0) and Illumina OmniExpress arrays. We used Birdseye to generate rare CNV calls from GWAS arrays. In addition, this Swedish sample has been well characterized for common SNPs (Ripke, O'Dushlaine, et al), enabling an examination of potential overlap in loci and pathways implicated by both sources of variation.

Results: Our results are consistent with literature reports of an increased burden of CNVs in SCZ cases. Our data confirm several known associations, including 16p11.2 duplications and 22q11.2 deletions, and also suggest a number of novel associations. Intriguingly, pathway analysis implicates several biological pathways that have also been implicated by common SNP polymorphisms, including calcium signaling and FMRP target sets. For events of >100kb, we note modest enrichment in cases for deletions in the post-synaptic density (human core) gene set, and also modest association for duplication and deletion in the NMDAR complex, known to modulate synaptic plasticity. We find evidence for modest enrichment of larger deletions in regions of SNP association.

Discussion: Together, these finding suggest both direct (loci) and indirect (pathway) convergence between these disparate sources of variation, shedding further light on core molecular components driving the etiology of this disorder.

1266T

Genome-Wide Association Study for Domain-Specific Cognitive Function.
H. Milo Rasouly[1], J. Chung[1], R. Au[4], L.A. Farrer[1,2,3,4,5], G. Jun[1,2,3]. 1) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA; 2) Department of Ophthalmology, Boston University School of Medicine, Boston, MA, USA; 3) Department of Biostatistics, Boston University School of Medicine, Boston, MA, USA; 4) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 5) Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

Neuropsychological (NP) tests are commonly used to assess cognitive function and impairment in clinical settings. Recently, several genome-wide association (GWA) studies for general cognitive ability have combined tests representing domain-specific cognitive functions and failed to uncover genome-wide significant associations. We performed heritability and GWAS for performance on several NP test 2,700 individuals from 1,790 families in the Framingham Heart Study (FHS). The three most heritable NP tests were the Boston naming test with no cues (BNT), the Wide range achievement test (WRAT) and the Trails B test (TRB) (0.30 to 0.72, p <10−4) after adjusting for age and sex. Family-based GWA analyses using 346,497 genotyped SNPs were conducted separately for each test after adjustment and normalization from the FHS sample. The most significant associations were discovered with two genotyped SNPs in the *PRDM10* gene (PR domain containing 10) for BNT (best SNP: rs6590429, p =1.81×10-7) and four genotyped SNPs in the *PDE11A* gene (Phosphodiesterase 11A) for WRAT (best SNP: rs1880916, p =3.64 ×10−6). No SNP met the initial cutoff for the significant association level p<10−5 with TRB. The association of 17 SNPs (p<10-5) from the discovery GWA analyses were further evaluated with five additional NP tests. The four *PDE11A* SNPs were significantly associated (p<10-3) with two more tests (best SNP: rs4335982 with Wechsler Adult Intelligence Scale- Similarities p=5.0×10-7), while the association of the *PRDM10* SNPs was limited to the BNT test (other tests: p>10-3). We attempted to replicate our top findings in a family-based cohort, the NIA-LOAD Study 3,828 subjects (1,840 Alzheimer disease cases and 1,988 controls from 2,265 families) using the HapMap 2 imputed data. A family-base approach revealed that the *PDE11A* gene contained 17 SNPs significantly associated (p<10−3) with animals category fluency test (best SNP: rs7605091, p=6.4×10-5), while the top discovery SNP rs1880916 was nominally significant (p=0.05). Rs7605091 was also associated with the episodic memory of immediate and delayed logical memory tests (p<0.01). However, none of *PRDM10* SNPs were strongly associated with these tests (p>0.01). Our results suggest that the *PDE11A* gene is involved in a general cognitive function.

1267F

Williams syndrome: Ribosomal profiling reveals genome-wide alterations in protein synthesis.
R. Weiss[1], L. Dai[2], J.R. Korenberg[2]. 1) Human Genetics, Univ. of Utah, Salt Lake City, UT; 2) Center for Integrated Neuroscience and Human Behavior, Brain Institute, Department of Pediatrics, Univ. of Utah, Salt Lake City, UT.

Williams syndrome (WS) is a copy-number variation (CNV) disease caused by a recurrent 1.5 million base pair hemizygous deletion of 28 genes. The WS phenotype includes a specific cognitive profile and unique personality characteristics. Central to interpreting the genetic and neural contributions phenotype induced by these genomic changes is to go beyond the transcriptional level to elucidate how the CNV-alterations play out at levels closer to cellular functions, ie., to determine abundance of the WS encoded protein(s) and their downstream effects on the abundance and networks of proteins expressed from other regions of the genome. Our prior work showed striking, biologically meaningful perturbations of the transcriptome response to the deletion in WS patient-derived lymphoblastoid cell lines (LCLs) using exon microarrays and RNA-Seq. We have shown that the WS deletion initiates a transcript level cis-effect that perturbs both WS region mRNAs, and mRNAs forming a network of interacting proteins and their associated pathways, including MAPK signaling and in actin cytoskeletal regulation. However, to help close the gap to brain circuitry and cellular function, protein consequences of transcriptional hints must be elucidated both in WS and in emerging data on Autism and Schizophrenia. To do this, we examined the level of protein synthesis in WS-derived LCLs by using ribosome profiling, a quantitative measure of translation that utilizes deep sequencing of ribosome-protected fragments. We analyzed the WS region transcripts by RNA-Seq and confirmed the expected 2-fold decrease in mRNA levels caused by gene dosage. We observed that the ribosome profile data supports a similar decrease at the level of actively translating ribosomes on these mRNAs, indicating that there is no compensation for the gene dosage effect at the level of translation for these mRNAs. However, perturbations in the translational efficiency of mRNAs from genes outside the WS region were observed, and in one case, we observed increased translation of LIMK2, a paralog of the WS-region Lim domain kinase 1 (LIMK1) gene. A profile of protein changes at the synapse is in progress. These data suggest a role for altered translational efficiency of mRNAs in interpreting how CNVs alter the abundance of their encoded protein(s) and their secondary effects on the abundance of proteins expressed from other regions of the genome.

1268W

A genome wide association study of epilepsy in a multi-ethnic cohort.
J.P. Bradfield[1], Z. Wei[2], C. Kim[1], R. Chiavacci[1], F. Mentch[1], W. Lo[3], M.R. Sperling[4], D.J. Dlugos[5], T.N. Ferraro[6], R.J. Buono[6], H. Hakonarson[1,7,8]. 1) Center For Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 3) Department of Pediatrics, The Ohio State University; Nationwide Children's Hospital, Columbus,OH; 4) Department of Neurology, Thomas Jefferson University, Philadelphia, PA; 5) Departments of Neurology and Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; 6) Department of Biomedical Sciences, Cooper Medical School of Rowan University, Camden, NJ; 7) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 8) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Epilepsy is a common complex disorder characterized by recurrent seizures. To search for common variants associated with this disorder, we performed a genome wide association study in Caucasians (1364 cases/ 6419 controls) and African Americans (271 cases/ 2843 controls). Along with an analysis of all epilepsies combined together, we analyzed two different subtypes of the disorder. Genetic generalized epilepsies (GGE) accounted for 770 (656 Caucasian/ 114 African American) of the cases, while non-symptomatic/cryptogenic focal epilepsy (Focal) accounted for 865 (708 Caucasian/ 157 African American) of the cases. Samples were genotyped on a combination of the Illumina HumanHap 550v1, 550v3, 610 Quad, and OmniExpress BeadChip platforms. All samples were imputed to ~38 million variants using imputev2 and the 1KG phase 1 interim reference haplotypes. A logistic regression was used to access significance of each variant and the results from both ethnicities were meta-analyzed together. There was one SNP (rs34018214, p=4.97×10-8) that reached genome wide significance in the GGE Caucasian cohort. This SNP resides in the PADI6 gene on chromosome 1p36.13. Loci with suggestive association (p < 1.0×10-6) were found at 20q12(PTPRT), 2q32.2(GLS) and 13q22.3(EDNRB) for all epilepsies; Focal epilepsy showed suggestive association at 9q31.2(ZFN462), 7p21.1(ITGB8), 19q13.11(NUDT19) and 2p16.1(RTN4); GGE showed suggestive association at 6q22.1-q22.31(SLC35F1), 4q22.1-q22.2(GRID2), 16p12.3(XYLT1) and 1q42.12(CNIH3). The validity of these associations will be dependent on replication in a larger independent cohort but nonetheless there are some interesting potential genes at these loci with association supported by multiple SNPs. Glutaminase (GLS) catalyzes the hydrolysis of glutamine to glutamate and ammonia, GRID2 is an ionotropic glutamate receptor, and CNIH3 regulates the trafficking and gating properties of AMPA-selective glutamate receptors. XYLT1 also lies in an area of chromosome 16 which has been shown to contain copy number variants associated with epilepsy, autism, and schizophrenia.

1269T

Analysis of onset age in Late-onset Alzheimer Disease genome-wide association data identifies novel onset age loci and confirms the predominance of *APOE*. *A.C. Naj[1], G. Jun[2,3,4], C. Reitz[5,6,7], B. Kunkle[8,9], W. Perry[8], Y.S. Park[8,9], G.W. Beecham[8,9], R. Rajbhandary[8], K.L. Hamilton-Nelson[8], L.S. Wang[10], J.S.K. Kauwe[11], E.R. Martin[8,9], J.L. Haines[12,13], R. Mayeux[5,6,7], L.A. Farrer[2,3,4,14,15], G.D. Schellenberg[10], M.A. Pericak-Vance[8,9], Alzheimer's Disease Genetics Consortium.* 1) Department of Biostatistics & Epidemiology, University of Pennsylvania, Philadelphia, PA; 2) Department of Medicine (Genetics Program), Boston University, Boston, MA; 3) Department of Biostatistics, Boston University, Boston, MA; 4) Department of Ophthalmology, Boston University, Boston, MA; 5) Taub Institute on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY; 6) Gertrude H. Sergievsky Center, Columbia University, New York, NY; 7) Department of Neurology, Columbia University, New York, NY; 8) The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 9) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 10) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 11) Department of Biology, Brigham Young University, Provo, UT; 12) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 13) Vanderbilit Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 14) Department of Epidemiology, Boston University, Boston, MA; 15) Department of Neurology, Boston University, Boston, MA.

Late-onset Alzheimer disease (LOAD) risk loci may also contribute to variation in LOAD age at onset (AAO), as do the allelic variants in *APOE*. However, roles in AAO for the confirmed risk loci outside of *APOE* have not been explored. Daw et al. (2000) estimated 4 additional loci with effects as great or greater than *APOE* contribute to AAO variation. We examined variants at ten confirmed LOAD risk loci (*APOE*, *CLU*, *PICALM*, *CR1*, *BIN1*, *CD2AP*, *EPHA1*, *ARID5B*, the *MS4A* region, *ABCA7*, and *CD33*) to determine if they contribute to variation in AAO among 9,160 LOAD cases in 14 datasets from the Alzheimer's disease Genetics Consortium (ADGC). Examining the variants most significantly associated with LOAD risk at each locus, we tested association with AAO using linear modeling assuming additive effects, adjusted for population substructure, and performed a random-effects meta-analysis across datasets. We also examined genetic burden using genotype scores weighted by risk effect sizes to examine the aggregate contribution of these loci to variation in AAO. Preliminary analyses confirmed association of *APOE* regional variation with AAO (rs6857, $P=3.30\times10^{-96}$). Variants at several other LOAD risk loci also demonstrated statistically significant associations with AAO ($P<0.005$), including rs6701713 in CR1 ($P=0.00717$), rs7561528 in BIN1 ($P=0.00478$), rs561655 in *PICALM* ($P=0.00223$). Burden analyses demonstrated that *APOE* contributes to 3.1% of variation in AAO ($R^2=0.220$) whereas the other nine genes contribute to 1.1% of variation ($R^2=0.200$) over baseline ($R^2=0.189$), after excluding study-specific effects. Secondary analyses of genomewide association with AAO performed among 10 ADGC case-control datasets (excluding cohort and family studies due to ascertainment differences) identified several regions with multiple SNPs showing suggestive associations ($P<10^{-5}$). These included chromosome 18q21.33 (rs12956834, $P=1.62\times10^{-6}$) and 9p13.3 (rs17356611, $P=2.49\times10^{-6}$), with one set of signals nearing genome-wide statistical significance in a chromosome 13q33.3 biological candidate gene *MYO16* (rs9521011, $P=7.62\times10^{-8}$). We confirmed the association of *APOE* variants with AAO among LOAD cases, and observe associations with AAO in *CR1*, *BIN1*, and *PICALM*. In contrast to earlier hypothetical modeling, we show that the combined effects of other loci do not exceed the effect of *APOE* on AAO, and if additional genetic contributions to AAO exist, they are likely very small individually.

1270F

Mapping Autism Spectrum Disorder Susceptibility Loci in Hispanic or Latino Populations. *Y.S. Park[1,2], N. Dueker[1], J. Jaworski[1], I. Konidari[1], P.L. Whitehead[1], C.D. Bustamante[3], J.L. Haines[4], M.A. Pericak-Vance[1,2], M.L. Cuccaro[1,2], E.R. Martin[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 2) Dr. John T. Macdonald Dept. of Human Genetics and Genomics, University of Miami Miller Schools of Medicine, Miami, fl, USA; 3) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 4) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN, USA.

Individuals of Hispanic or Latino origin (Hispanics), currently the largest minority group in the US and projected to be nearly 30% of the US population by the year 2050, present a unique challenge in complex disease genetic studies due to their admixed genomic background. Autism spectrum disorder (ASD) is highly heritable disorder with strong evidence for genetic influence. Current epidemiologic evidence suggests that ASD may occur as often as 1 in 88 children. Most ASD studies have primarily focused on individuals of European ancestry, making it difficult to generalize results in other populations. Our study interrogated a total of 253,682 SNPs genotyped in 2,033 individuals in over 550 families from three independent cohorts: HIHG (University of Miami and Vanderbilt University), AGRE (Autism Genetic Resource Exchange) and SSC (Simons Simplex Collection). Families with either the proband or at least one parent defined as self-reported Hispanics were included in the analysis. To our knowledge, this is the largest GWAS targeted to study ASD in Hispanics. These individuals have not been studied collectively in any other GWAS. Genomic ancestry was investigated with EIGENSTRAT principal component analysis (PCA) using HapMap 3 and GOAL (Latinos from South Florida) projects as reference populations. Association with ASD was assessed by PDT for nuclear families. The most significant SNP resided in *Low Density Lipoprotein Receptor-Related Protein 1B (LRP1B)* gene (PDT $p=2.62\times10^{-5}$). A deletion involving the region has been noted in a case report for severe intellectual disability, omphalocele, hypospadias and high blood pressure. Also notable among SNPs with nominal significance was a missense exonic SNP in *Thyroglobulin (TG)* gene (PDT $p=1.67\times10^{-4}$) and an intronic SNP in the same gene (PDT $p=5.20\times10^{-5}$). Interestingly, several genes previously implicated in ASD susceptibility also remained nominally significant in our Hispanic ASD individuals, such as: *CDH7* (PDT $p=1.11\times10^{-3}$), *CNTNAP2* (PDT $p=8.76\times10^{-3}$) and *RBFOX1* (PDT $p=4.23\times10^{-2}$). In conclusion, our study shows that common variants may contribute to the genetic complexity of ASD in Hispanics.

1271W

Genome-wide association study of age at onset of alcohol dependence in large COGA families. *M. Kapoor[1], JC. WANG[1], L. Wetherill[2], S. Bertelsen[1], J. Budde[1], L. Nhung[1], A. Agrawal[1], V. Hesselbrock[3], J. Rice[1], J. Nurnberger[2], L. Bierut[1], T. Foroud[2], A. Goate[1].* 1) Psychiatry, Washington University in St Louis, Saint Louis, MO, USA; 2) Indiana University School of Medicine, Indianapolis, USA; 3) University of Connecticut Health Center, Farmington, USA.

The age at onset of alcohol dependence (AD) is a critical moderator of genetic associations for alcohol dependence. The present study evaluated whether genetic variants can predict the occurrence of AD in large high-risk families from the Collaborative Study on the Genetics of Alcoholism (COGA). The primary analysis was performed in 2312 subjects from 118 of the largest and most densely affected families. Analyses were limited to the subjects who were regular drinkers. We tested for association between age at onset of AD and variants on the Illumina Omniexpress array using a Cox proportional hazards regression model after adjusting for the effects of birth cohort, gender and family structure. Non-dependent subjects were right censored at the age at last interview. A Schoenfeld residual test was also performed to test the proportionality for the strongest signal. This family-based analysis identified a SNP, rs9847462 near micro RNA on chromosome 3, that is strongly associated (P = 1.03 × 10-11) with the age at onset of AD among regular drinkers. rs9847462 and other tagged SNPs were located under the region of linkage peak identified in our previous analysis. The results indicated that with each year of increase in age, carriers of the minor allele of rs9847462 were 1.5 times more likely to become AD than those homozygous for the wild-type genotype. By the age of 20 years, nearly 60% of subjects homozygous for the minor allele of rs9847462 variant were alcohol dependent in comparison to 20% of subjects homozygous for the major allele. This SNP was also associated with AD symptom count (P = 8.53 × 10-7) and DSM4 AD (P = 3.04 × 10–8) in COGA families, but time to onset analysis significantly improved the power to identify these genetic variants. The present study also identified several other novel variants with suggestive evidence of association with age at onset of AD. These results indicate that there is a moderating effect of age on the risk of AD and time to onset analysis can increase the power to identify novel signals for AD.

1272T

Genetic determinants of the natural history of Alzheimer's. *X. WANG[1], O.L. Lopez[2,3], R.A. Sweet[2,4], M.M. Barmada[1], F.Y. Demirci[1], M.I. Kamboh[1,3].* 1) HUMAN GENETICS, University of Pittsburgh, PITTSBURGH, PA; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA,; 3) Alzheimer's Disease Research Center, University of Pittsburgh, Pittsburgh, PA; 4) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA.

Alzheimer's disease (AD) is a devastating neurodegenerative disease and is characterized by a gradual cognitive and functional decline. Alzheimer's disease is a significant public health problem worldwide and the cost associated with AD is staggering. In order to effectively plan for the future care and treatment of AD patients, it is important to understand the factors that influence the natural history of AD. Although multiple factors have been identified that may affect the natural history of AD, including time to admission to nursing home (NH) and time to death, no studies have been conducted on the genetic determinants of the natural history of AD. In the present study, we have performed a genome-wide association study (GWAS) on 1,187 European American AD patients in order to identify genetic factors that may affect the time to NH admission and time to death. All patients were examined at the University of Pittsburgh Alzheimer's Disease Research Center who had at least one follow-up evaluation (mean follow-up time 4.3 + 2.7 years, range 0.8–18 years). Genotyping was performed using the Illumina Human Omni 1-Quad BeadChip and the analysis was conducted using Cox proportional hazards regression under an additive genetic model with adjustment for age, sex, education, baseline MMSE score, medication, psychosis and the first four principal components. The established genetic risk factor, APOE*4, was not strongly associated with NH admission (P= 0.05) and death (P=0.15). However, we found multiple top significant signals for time to admission to NH on chromosome 6 (P<1E–06) in three adjacent genes: ZUFSP (P=1.08E–07), RSPH4A (P=4.15E–07), and KPNA5 (P= 6.45E–07). We also observed multiple suggestive loci for time to death of AD, including the top signal between the KDR and SRD5A3 genes on chromosome 4 (P=1.82E-07), ARTNL1 on chromosome 10 (P=1.38E–06), SYNPO on chromosome 5 (P=2.36E–06), PPP1R36 on chromosome 14 (P=3.76E-06) and NEUROD4 on chromosome 12 (P=4.288E-06).Our data s suggest that natural history of AD has genetic basis and this novel finding needs to be confirmed in independent studies.

1273F

A genome-wide association study of the response to cognitive behavioral therapy in children with anxiety disorders. *J.R.I. Coleman[1], K. Lester[1], C. Curtis[1], J.L. Hudson[2], C. Creswell[3], G. Breen[1], T.C. Eley[1].* 1) MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK; 2) Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia; 3) Winnicott Research Unit, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK.

Anxiety disorders are the most common psychiatric disorders, with a prevalence in adults of approximately 30%. Psychosocial treatments, including cognitive behavioral therapy (CBT), are the primary treatment modality for anxiety disorders in the United Kingdom. Response (at 6-12 month follow-up) for CBT is estimated at 65%, demonstrating heterogeneity. Evidence suggests individual differences in treatment response have a genetic basis. Preliminary evidence exists for beneficial effects of the 5HTTLPR SS genotype, and for the T allele of rs6330 (NGF gene), on CBT response at 3-12 month follow-up in a cohort of children with a variety of anxiety disorders (Eley et al., 2012, Lester et al., 2012). Such therapygenetic candidate gene studies motivated our current genome-wide association study (GWAS).

Genetic samples were gathered from 1272 children (aged 7–13) diagnosed with anxiety disorders, undergoing CBT. The sample originated from ten different sites in the UK, USA, Australia and Western Europe. Clinical Severity Ratings were established at baseline, upon completion of treatment, and at follow-up. DNA was extracted using a standard protocol, concentrated by filtration, and genotyped using the recently released Illumina HumanCore-Exome array, which assays roughly 250000 SNPs and 250000 exomic variants. The initial outcome variable was primary disorder presence at follow-up. The SNPs most strongly associated with this outcome are presented. Additional measures include genome-wide complex trait analysis (GCTA) and quantitative trait heritability analyses. Further analyses will explore the association between SNPs and changes in Clinical Severity Ratings, and investigate response immediately post-treatment. This is, to our knowledge, the first GWAS of response to psychosocial treatment, and the first treatment response GWAS to be performed in anxiety disorder. Past analyses in related phenotypes such as antidepressant response suggest that the effect sizes of variants underlying treatment response are likely to be small. As the sample size of this study is relatively small, it is probably underpowered to detect such variants. However, GCTA could yield valuable insight; one study that found no SNPs at genome-wide significant levels for antidepressant response gave a SNP-heritability estimate of 42% from GCTA.

1274W

Genome-Wide Association Study of Shared Components of Reading Disability and Language Impairment. *J.D. Eicher[1], N.R. Powers[1], L.L. Miller[2], S.M. Ring[2], J.R. Gruen[1,3], Pediatric Imaging, Neurocognition, and Genetics Study.* 1) Genetics, Yale University, New Haven, CT; 2) School of Social and Community Medicine, University of Bristol, Bristol, UK; 3) Pediatrics and Investigative Medicine, Yale University School of Medicine, New Haven, CT.

Written and verbal language are inherently intertwined neurobehavioral traits vital to the development of communication skills. Unfortunately, disorders involving these traits~specifically reading disability (RD) and language impairment (LI)~are common and prevent affected individuals from developing adequate communications skills, leaving them at risk for adverse academic, socioeconomic, and psychiatric outcomes. Both RD and LI are complex traits that frequently co-occur, leading us to hypothesize that these disorders share genetic etiologies. To test this, we performed a genome wide association study on individuals affected with both RD and LI in the Avon Longitudinal Study of Parents and Children. The strongest associations were seen with markers in ZNF385D (OR=1.81, minimum p=5.45 × 10–7). These associations were strengthened when examining LI cases (OR=1.62, minimum p=6.96 × 10–8). We replicated association of ZNF385D using receptive vocabulary measures in the Pediatric Imaging Neurocognitive Genetics study (minimum p=0.00245). We then used diffusion tensor imaging fiber tract volume data on 16 fiber tracts to examine the implications of replicated markers. ZNF385D was a predictor of overall bilateral fiber tract volumes, specifically in the inferior longitudinal, inferior fronto-occipital, and temporal superior longitudinal fasculi. Here, we present evidence for ZNF385D as a candidate gene for individuals with both RD and LI. The implication of the transcription factor ZNF385D in RD and LI underscores the importance of transcriptional regulation in reading and language. Further study is necessary to discern the target genes of ZNF385D and how it functions within neural development, particularly in fiber tracts vital to fluent language.

1275T

A genome wide association between migraine in bipolar disorder and Neurobeachin. *K.K. Jacobsen[1, 2, 3], S. Johansson[1, 2, 3], C.M. Nievergelt[4], T. Zayats[1, 2, 3], T.A. Greenwood[4], H.S. Akiskal[4, 5], J. Haavik[1, 3, 6], O.B. Fasmer[3, 6, 7], J.R. Kelsoe[4, 5], K.J. Oedegaard[3, 6], BiGS Consortium.* 1) Department of Biomedicine, University of Bergen, Bergen, Norway; 2) Center for Medical Genetics and Molecular Medicine, Haukeland Universtiy Hospital, Bergen, Norway; 3) K. G. Jebsen Center for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway; 4) Department of Psychiatry, University of California San Diego, San Diego, USA; 5) Department of Psychiatry, VA hospital, San Diego, USA; 6) Division of Psychiatry, Haukeland University Hospital, Bergen, Norway; 7) Section for Psychiatry, Department of Clinical Medicine, University of Bergen, Bergen, Norway.

Background: Migraine is a common headache disorder, with a prevalence of approximately 12%. It is characterized by unilateral attacks, and in some patients accompanied by visual aura symptoms (1). Bipolar disorder is a mood disorder ranging from severe depression to mania, with migraine as a common comorbidity seen in 25–45% (2,3). Bipolar disorder patients with migraine are more likely to have a worse outcome than those without migraine(4). Methods: We performed a genome wide association analysis on 460 bipolar patients with self-reported migraine and 914 bipolar patients without migraine. The individuals are from the TGEN sample, a part of the Bipolar Genetics Study (BiGS). Replication was attempted in the GAIN sample, a separate part of BiGS. Results: We found a genome wide significant association between migraine in bipolar disorder and rs1160720, an intronic single nucleotide polymorphism (SNP) in NBEA (P-value 2.97×10–8, OR: 1.82, 95% CI: 1.47-2.25). The SNP resides in a linkage disequilibrium block with several other associated SNPs, spanning several exons. We were not able to replicate our finding in the GAIN sample. Discussion: NBEA encodes the protein neurobeachin, which is involved in the transport of neurotransmitter receptors, among them glutamatergic receptors (5). This receptor system is implicated in both bipolar disorder and migraine (6). Our study provides putatively new clues to the cause of migraine in bipolar disorder, though further studies are needed to verify this association.

1276F

Genome-wide association study of HLA-DQB1*06:02 negative essential hypersomnia. *SS. Khor[1], T. Miyagawa[1], H. Toyoda[1], M. Yamasaki[1], Y. Kawamura[2], H. Tanii[3], Y. Okazaki[4], T. Sasaki[5], L. Lin[6], J. Faraco[6], T. Rico[6], Y. Honda[7], M. Honda[7,8], E. Mignot[6], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan; 2) Yokohama Clinic, Warakukai Medical Corporation, Yokohama 221-0835 Japan; 3) Department of Psychiatry, Mie University School of Medicine, Mie 514-8507 Japan; 4) Metropolitan Matsuzawa Hospital, Tokyo 156-0057, Japan; 5) Graduate School of Education, The University of Tokyo, Tokyo 113-0033 Japan; 6) Stanford Center for Sleep Sciences and Medicine, Stanford University School of Medicine, 1050A Arastradero Rd., Palo Alto, CA 94304, USA; 7) Department of Somnology, Tokyo Medical University, Tokyo 160-0023, Japan; 8) Sleep Research Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan.

Essential hypersomnia (EHS), a sleep disorder characterized by excessive daytime sleepiness, can be divided into two broad classes based on the presence or absence of the HLA-DQB1*06:02 allele. HLA-DQB1*06:02-positive EHS and narcolepsy with cataplexy are associated with the same susceptibility genes. In contrast, there are fewer studies of HLA-DQB1*06:02 negative EHS which, we hypothesized, involves a different pathophysiological pathway than does narcolepsy with cataplexy. In order to identify susceptibility genes associated with HLA-DQB1*06:02 negative EHS, we conducted a genome-wide association study (GWAS) of 125 unrelated Japanese EHS patients lacking the HLA-DQB1*06:02 allele and 562 Japanese healthy controls. A comparative study was also performed on 268 HLA-DQB1*06:02 negative Caucasian hypersomnia patients and 1761 HLA-DQB1*06:02 negative Caucasian healthy controls. We identified three SNPs that each represented a unique locus~ rs16826005 (P = 1.02E-07; NCKAP5), rs11854769 (P = 6.69E-07; SPRED1), and rs10988217 (P = 3.43E-06; CRAT) that were associated with an increased risk of EHS in this Japanese population. Interestingly, rs10988217 showed a similar tendency in its association with both HLA-DQB1*06:02 negative EHS and narcolepsy with cataplexy in both Japanese and Caucasian populations. This is the first GWAS of HLA-DQB1*06:02 negative EHS, and the identification of these three new susceptibility loci should provide additional insights to the pathophysiological pathway of this condition.

1277W

Genome-wide Association Study and Admixture Mapping of Age-related Cognitive Decline in African Americans. *T. Raj[1,2,3], L.B. Chibnik[1,2,3], C. McCabe[1], B. Stranger[4], H. Hendrie[5], L. Barnes[6], T. Foroud[7], J.M. Murrell[8], D.A Bennett[6], K.S. Hall[9], D.A. Evans[10], P.L. De Jager[1,2,3].* 1) Department of Neurology, Brigham and Women's Hospital, Boston MA; 2) The Broad Institute, Cambridge MA; 3) Harvard Medical School, Boston MA; 4) Section of Genetic Medicine, University of Chicago, Chicago IL; 5) Department of Medicine, Indiana University, Indianapolis IN; 6) Rush Alzheimer Disease Center, Rush University Medical Center, Chicago, IL; 7) Department of Medical and Molecular Genetics, Indiana University, Indianapolis IN; 8) Department of Pathology and Laboratory Medicine and Indiana Alzheimer Disease Center, Indianapolis IN; 9) Department of Psychiatry, School of Medicine, Indiana University, Indianapolis, IN; 10) Department of Internal Medicine, Rush Institute for Healthy Aging, Rush University Medical Center, Chicago, IL.

Objective: To leverage genome-wide data from several cohorts so as to identify genomic determinants of age-related cognitive decline among older African Americans.

Data and Methods: We examined genome-wide genotyping data for 3,964 unrelated older African American (AA) subjects who had at least two repeated measures of cognition from five (CHAP, IIDP, ROS, MAP, and MARS) prospective community-based studies. We compared them to 2,703 unrelated non-Hispanic European Americans (EAs) from the ROS, MAP and CHAP studies. Within each cohort, individual cognitive tests were combined to form an aggregate measure of global cognition. We used linear mixed effects models to characterize individual paths of change in cognition, controlling for age, sex and education as fixed effects. After quality control, imputation was performed using the 1000 Genomes Project Phase I combined reference panels of EA and AA ancestry. Common variants were meta-analyzed across cohorts using weighted fixed-effect models. HAPMIX and MIXSCORE programs were used to estimate ancestry at each locus, and to test for association of cognitive decline with locus-specific proportion of ancestry.

Results and Conclusion: *APOE* ε4 haplotype was strongly associated with rate of cognitive decline in AAs ($P= 1.92 \times 10^{-14}$) but the magnitude of the effect was significantly weaker in AAs ($\beta= -0.01$) compared to EAs ($\beta= -0.05$). In AAs we replicated previously known loci in EAs including *ABCA7* ($P= 4.13 \times 10^{-4}$), *PICALM* (7.35 × 10–4), and *EPHA1* ($P=8.06 \times 10^{-4}$). We discovered one genome-wide significant association at the *TRPS1* locus (chr 8q23), and replicated that association in EAs ($P(DISC) = 2.52 \times 10^{-7}$; $P(REP) = 9.79 \times 10^{-3}$; $P(JOINT) = 1.37 \times 10^{-8}$). The top variant in the region is significantly associated with cis gene expression of *TRPS1* in primary monocytes. We also found another variant at the *TEK* locus (chr 9p21) with suggestive evidence of association ($P(DISC) = 4.10 \times 10^{-7}$). Using genome-wide admixture mapping, we did not detect any significant admixture peak, suggesting that genetic factors alone may not explain the differences in age-related cognitive decline between AAs and EAs.

1278T

Genome-wide association study of growth rate and energy status in lymphoblastoid cell lines. *A.R. Sanders[1,2], W. Moy[1], H.H.H. Göring[3], J. Freda[1], D. He[1], S.L. Fuentes[1], J. Duan[1,2], P.V. Gejman[1,2], Molecular Genetics of Schizophrenia (MGS) Collaboration.* 1) Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry and Behavioral Sciences, University of Chicago, Evanston, IL; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

The determinants of growth rate and energy status in LCLs, whether genetic, environmental, or both, remain largely unknown, though growth rate and energy status vary across LCLs and strongly influence many cellular phenotypes. We performed a GWAS using 2,060 samples from the European ancestry (EA) portion of the MGS case-control collection with transformation site, age, EBV load, sex, caseness, and ancestry PCs as covariates. For the top 500 most associated SNPs for each trait, we extracted their genes (or if intergenic, the genes closest on each side), resulting in 429 genes (growth rate) and 383 genes (energy status). The genomic inflation factors for the growth rate (1.000) and the energy status (1.014) GWAS indicated little cryptic population substructure. We found some associations for growth rate with age (p=0.018), sex (p=1.7×10^{-6}), and EBV load (p=8.3×10^{-5}), and for energy status with sex (p=1.8×10^{-4}), caseness (p=0.0076), transformation site (p=0.024), and EBV load (p=0.0090), suggesting the utility of including such covariates in the analysis. We found no genome-wide significant association. For growth rate, the strongest association with additional support was with rs7750067 (p=5.9×10^{-7}), with supportive findings (six SNPs with $5\times10^{-6}<p<10^{-4}$) over this intergenic region, which contains a predicted microRNA (miRNA), ENSG00000266073. The nearest RefSeq genes are *SLC25A51P1*, *EYS*, and *BAI3*. *EYS* contains multiple epidermal growth factor (EGF)-like and LamG domains and is implicated in retinitis pigmentosa. *BAI3* encodes a brain-specific angiogenesis inhibitor, which is an adhesion-G protein-coupled receptor and has been shown to control dendritic arborization growth and branching in cultured neurons. Both *EYS* and *BAI3* are expressed in various tissues, including brain and blood. Pathway analyses revealed the most enriched GO-term for growth rate to be neuron projection (FDR=0.08), and for energy status two were significant: regulation of cell proliferation (FDR=0.0067) and neuron differentiation (FDR=0.042). We describe the effect of epidemiological variables and of specific genomic loci on growth rate and energy status in LCLs. Our results have potential to lead to a better understanding of these important cellular traits. We will present data on the expanded sample of LCLs at the meeting, projecting to reach ~2,500 EA subjects, along with ~1,000 African American subjects.

1279F

Distinct genetic contributions to Tourette Syndrome and Obsessive Compulsive Disorder as revealed by cross-disorder genome-wide studies. *D. Yu[1, 2] on behalf of TS GWAS Consortium, TSAICG, and IOCDFGC.* 1) CHGR, Massachusetts General Hospital, Boston, MA; 2) Stanley Center, Broad Institute, Boston, MA.

Investigation of the genetic relationship of phenotypically-related heritable diseases is important for identifying both shared and distinct genetic risk factors and ultimately for providing molecular insights into diagnosis and treatment. Polygenic score analysis of genome-wide association study (GWAS) data is a powerful tool to examine aggregate cross-disorder genetic effects. Likewise, enrichment analysis of top GWAS signals using expression quantitative loci (eQTLs) from specific tissues is emerging as a complementary approach to assess shared, putatively functional variants between disorders. Here we applied both methods to investigate the genetic relationship between Tourette syndrome (TS) and obsessive-compulsive disorder (OCD), two developmental neuropsychiatric disorders with overlapping clinical symptoms and high rates of comorbidity. In a combined GWAS of TS and OCD in 2,723 cases, 5,667 ancestry-matched controls, and 290 OCD parent-child trios, no markers achieved genome-wide significance (p= 5×10^{-8}) after imputation to the data from the 1000 Genomes Projects. Among 9,099 single nucleotide polymorphisms (SNPs) with p<1×10^{-3}, we found significant enrichment for eQTLs in several brain regions (empirical p<0.001 for frontal cortex and cerebellum, p=0.005 for parietal cortex). While top loci from the individual TS and OCD GWAS studies both demonstrated brain eQTL enrichment, the specific brain regions differed between the two disorders, with TS enriched strongly with cerebellar eQTLs (p<0.001) and OCD with eQTLs from frontal cortex (p<0.001). Within-disorder polygenic score analyses identified a significant polygenic component for OCD (p= 2×10^{-4}) and a smaller polygenic component for TS (p=0.06), explaining 3% and 0.6% of the phenotypic variance, respectively. Reducing the sample size of OCD to that of TS resulted in a decreased, but still significant OCD polygenic signal (p=0.01), explaining 1.4% of OCD phenotypic variance. Intriguingly, no shared polygenic signal was detected in cross-disorder polygenic analysis; furthermore, the OCD polygenic signal was attenuated when TS cases with co-morbid OCD were included. Together, these data suggest that TS and OCD both have a polygenic component, but may have distinct genetic risk loci. In addition, OCD in the presence of TS/CT may have different underlying genetic susceptibility compared to OCD alone.

1280W

Coding variations and the risk of adult ADHD. An exome-chip association study in adult ADHD from the IMpACT. *T. Zayats, S. Johansson, J. Haavik, IMPACT Consortium.* Biomedicine, University of Bergen, Bergen, Bergen, Norway.

Introduction: Attention deficit hyperactivity disorder (ADHD) is a highly heritable childhood onset neuropsychiatric condition that often persists into adulthood. Still, genetics of ADHD, and particularly ADHD in adults is largely unknown. The main purpose of this study was to perform a genome-wide scan of adult ADHD using the newly available Illumina HumanExome12v1 chip, and to evaluate the performance of this technology for genotyping common and rare variants. Materials and Methods: The analyses were carried out using DNA samples collected by the International Multicenter persistent ADHD Consortium (IMpACT). The evaluation of HumanExome12v1 chip was implemented on 2215 individuals, 25 of which were also whole exome sequenced using Roche-NimbleGen Sequence Capture EZ Exome v2 kit and paired-end 100nt sequencing on the Illumina HiSeq. All participants were genotyped on Illumina HumanExome12v1 chip. Genotypes were called in Illumina GenomeStudio V2011.1 software, with additional genotype assignments in zCall software. Performance of HumanExome12v1 chip as well as GenomeStudio V2011.1 software was assessed by direct comparison of data obtained from different DNA sources (blood and saliva), duplicate samples and sequenced genotypes. Mendelian consistency testing was also utilized. Sanger sequencing was used to validate rare variants with minor allele frequency <1%. Association testing was carried out as single marker logistic regression correcting for population substructure and gender in PLINK software. Results: We developed a quality control analysis pipeline for the variants of HumanExome12v1 chip using series of steps implemented in PLINK, GenomeStudio V2011.1 and zCall softwares. Overall, the performance of HumanExome12v1 chip was comparable to that of next generation sequencing. However, it is worth mentioning that mismatch rates were notably higher for rare variants with minor allele frequency < 1% compared to the rest of the chip; the exact specifics will be discussed in details. Association testing in a total of 2600 adult ADHD cases and 7000 controls are currently being performed and will be presented. Conclusion: This study provides insight into the performance of the newly available Illumina HumanExome12v1 chip, serving practical guidance to its quality control with special emphasis on rare variants. Being the largest systematic adult ADHD study to date, this study expected to shed new light on our understanding of adult ADHD genetics.

1281T

Genome-Wide Association Study of Rate of Cognitive Decline in Six Alzheimer's Disease Cohorts. *R. Sherva[1], AL. Gross[2], S. Mukherjee[3], S. Newhouse[4,12], J. Kauwe[5], L. Munsie[6], AJ. Saykin[7], DA. Bennett[8], LB. Chibnik[9,13,14], P. Kraft[10], PK. Crane[3], RC. Green[11,15], Alzheimer's Disease Neuroimaging Initiative and GENAROAD Consortium.* 1) Department of Biomedical Genetics, Boston University School of Medicine, Boston, MA; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Department of Medicine, University of Washington, Seattle, WA; 4) King's College London, Institute of Psychiatry, London, UK; 5) Departments of Biology and Neuroscience, Brigham Young University, Provo, UT; 6) Tailored Therapeutics, Eli Lilly and Company, Indianapolis, IN 46285; 7) Indiana University School of Medicine, Indianapolis, IN; 8) Rush Alzheimer's Disease Center, Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA; 9) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology & Psychiatry, Brigham and Women's Hospital Boston, MA, USA; 10) Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA; 11) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 12) NIHR Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation; 13) Department of Neurology, Harvard Medical School, Cambridge, MA, USA; 14) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 15) Partners Healthcare Center for Personalized Genetic Medicine, Boston, MA.

Background: Variability in cognitive decline in Alzheimer's disease (AD) cannot be entirely explained by demographic, health, or clinical features and genetics are likely to play a role. Determining genetic factors contributing to the variability in the rate of cognitive decline in AD might help reveal previously unidentified biological pathways responsible for disease progression. **Methods:** We developed composite measures of global cognition in seven studies of persons with AD: Alzheimer's Disease Neuroimaging Initiative, National Alzheimer's Coordinating Center, Religious Orders Study, Rush Memory and Aging Project, clinical trial of the drug Tarenflurbil, the Cache County Study on Memory, Health, and Aging, and the AddNeuroMed study (N=3,128). We performed a pre-statistical harmonization to identify tests in common across studies, then used factor analysis consistent with item response theory to estimate cognitive ability scores for each observation at all time points in all studies. Heritability (h^2) of rate of decline was estimated using the Genome-wide Complex Trait Analysis (GCTA; J Yang *et al*. 2010) package to estimate the proportion of variance of slopes calculated within each individual based on their cognitive ability scores at each time point that can be explained by the genome-wide SNP data (SNP-based h^2). Association tests were performed on the cognitive ability scores using generalized estimating equations regressions with an interaction between months with AD and SNP genotype. Results were combined in a meta-analysis. **Results:** The median follow-up time was 1.5 years (range 0–19 years). The composite had interval-level properties, was internally consistent (Cronbach's α=0.96), and demonstrated reliable measurement precision over a broad range of cognitive ability. Rate of decline in our pooled sample was heritable (h^2=0.38, P=0.06). We identified several variants strongly associated with rate of decline with evidence from multiple samples, including SNPs in intergenic regions on chromosomes 18 (P_{meta}=1.23E-08) and 19 (P_{meta}=2.33E-08). APOE ε-4 allele counts were not associated with rate of decline in any cohort where it was available (ADNI, NACC, and the Tarenflurbil trial, P>0.05). **Conclusions:** Genetics influence rate of decline, although there is little evidence that AD risk genes have a strong influence.

1282F

Rare Variant Discovery in Late-Onset Alzheimer's Disease using Exome Arrays. *L.-S. Wang[1], A.C. Naj[2], C. Cruchaga[3], S. Mukherjee[4], R. Graham[5], B. Kunkle[5], A. Partch[1], L. Cantwell[1], P. Crane[6], A.M. Goate[3], J.L. Haines[7], R. Mayeux[8], L. Farrer[9], M.A. Pericak-Vance[5], T. Behrens[10], G.D. Schellenberg[1], Alzheimer's Disease Genetics Consortium.* 1) Dept Pathology and Laboratory Medicine, Univ Pennsylvania, Philadelphia, PA; 2) Dept Biostatistics and Epidemiology, Univ Pennsylvania, Philadelphia, PA; 3) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO; 4) Div General Internal Medicine, University of Washington, Seattle, WA; 5) John P. Hussman Institute for Human Genomics, Univ Miami, Miami, FL; 6) Dept Medicine, University of Washington, Seattle, WA; 7) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 8) Sergievsky Center, Columbia University Medical Center, New York, NY; 9) Biomedical Genetics Program, Boston University School of Medicine, Boston, MA; 10) Genentech, Inc., San Francisco, CA.

Two genome-wide association (GWA) studies in 2012 increased the number of susceptibility genes for LOAD to 9. Last year the rare APP A673T variant as well as several rare variants in the TREM2 genes were also reported by three studies. Exome chips use proven chip-based genotyping technology to genotype nearly 250,000 rare functional variants identified in whole-exome sequencing studies of more than 12,000 subjects, and represent a powerful tool for identifying rare functional variants involved in complex diseases like LOAD. The Alzheimer's Disease Genetics Consortium (ADGC) has generated or collected genotype data of 16,525 samples from four cohorts using Illumina HumanExome BeadChips (version 1), which consisted of 247,870 exonic SNPs, indels, known GWAS hits, and other relevant markers selected from a consortium of more than 12,000 whole exome sequencing samples. Genotype calls were done using clustering profiles generated from GenenTech (>35,000 samples). 14,355 samples (8,221 AD cases, 6,134 controls) passed QC and were divided into discovery and replication cohorts. Our preliminary analysis led to the following discoveries. We applied several quality control steps including (a) removing mismatched samples using GWA SNP array data genotype concordance, (b), removing subjects and variants with low call-rate (>5% missingness), and (c) excluding monomorphic variants, after which we retained 5,801 subjects and 145,681 markers. We found that the APP A673T variant is extremely rare in the American Caucasian population (2 heterozygotes in 14,355 samples). We performed analyses with SKAT (Optimized Sequence Kernel Association Test) on rare SNPs (MAF<2%), adjusting for age-at-onset (cases)/age-at-exam (controls), number of APOE ε4 alleles, and population principal components. No genome-wide significance was found. We found that the rare signal from TREM2 is replicated in the SKAT analysis (P=6.97e–5). SKAT analysis on known AD genes detected rare variants SORL1 (P= 0.0236) and PICALM (P=8.45e–5), showing that sophisticated association tests are needed to tease out rare variant signal. These findings suggest that other rare exonic variants may contribute to LOAD risk. PCR validation of genotype calls is being performed. We will report on the replication of these preliminary findings in this presentation.

1283W

A Noise-Reduction GWAS in Multiple Sclerosis. *J.P. Hussman[1,2], A.H. Beecham[1], M. Schmidt[1], J.L. McCauley[1], J.L. Haines[3], M.A. Pericak-Vance[1], International Multiple Sclerosis Genetics Consortium.* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Hussman Foundation, Ellicott City, MD, USA; 3) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, USA.

Genome-wide association studies (GWAS) have proven useful in identifying disease susceptibility genes in multiple sclerosis. However, GWAS have focused on the testing of each single nucleotide polymorphism (SNP) independently. Given it is likely that the disease variants have not been directly genotyped, tests that account for multiple flanking SNPs in linkage disequilibrium (LD) with the disease variants may increase the power to detect association. We have used the GWAS noise reduction (GWAS-NR) approach to identify LD blocks which are significantly associated with multiple sclerosis. As GWAS-NR uses a linear filter to identify genomic regions demonstrating correlation among association signals in multiple data sets, we have divided the genotype data from a previously published GWAS in a total sample of 9722 cases and 17376 controls into seven population specific strata from Australia, Central Europe, the Mediterranean, Finland, Scandinavia, United Kingdom, and the United States. Within each stratum, we adjusted for the first five principal components from Eigenstrat. Of the 87838 LD blocks in the GWAS data, defined by the confidence intervals algorithm, we have identified 1434 blocks with p-value < 5.0E−02. A total of 226 of the 1434 blocks show p-value < 5.0E-08 using a modified version of the Truncated Product Method, 138 being within the Major Histocompatibility Complex (MHC) on chromosome 6 (28-34 MB). Of the remaining 88, 24 contain SNPs previously identified in GWAS at a genome-wide level and 47 lie within 2 MB of previously identified SNPs. The remaining 17 are in regions novel to previous GWAS efforts. In particular, four of the 17 blocks are on chromosome 1 in the region of FCRL2, FCRL3, and FCRL4. Previous studies have found that the Fc receptor-like (FCRL) molecule may contribute to the autoimmune disease process. Other noteworthy findings include blocks within JAK1 and COL11A2. Like TYK2, which has been shown to be associated with multiple sclerosis, JAK1 is a non-receptor tyrosine kinase. Three blocks in COL11A2, located within the MHC at ~33 MB, show significance (p-values < 5.0E-08). This is of interest as our recent analysis of sequence data in multiplex families also pointed to this gene, with one of the identified variants being located within one of the LD blocks. These data support the use of this approach to identify additional genetic factors contributing to multiple sclerosis risk.

1284T

Genome-Wide Association Study of Alcohol Consumption in the Kaiser Permanente/UCSF Genetic Epidemiology Research on Adult Health and Aging Cohort. *E. Jorgenson[1], L. Shen[1], T. Hoffmann[2,3], M. Kvale[3], Y. Banda[2,3], D. Ranatunga[1], N. Risch[2,3], C. Schaefer[1], C. Weisner[1,4], J. Mertens[1].* 1) Kaiser Permanente Division of Research, Oakland, CA; 2) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 4) Department of Psychiatry, University of California San Francisco, San Francisco, CA.

Alcohol consumption is a common trait with a complex etiology. There is strong evidence that both genetic and environmental factors play a role in determining an individual's alcohol use, including patterns of alcohol consumption and the risk of alcohol dependence. A number of genome-wide association studies of alcohol dependence and alcohol consumption, utilizing samples of several thousand subjects, have identified association signals that require further confirmation in independent samples. Here, we report the results of a genome-wide association study of alcohol consumption in the Kaiser Permanente/UCSF Genetic Epidemiology Research on Adult Health and Aging Cohort (n=110,266), the largest study to date. We identified genome-wide significant associations between SNPs in the ethanol metabolism pathway and measures of alcohol consumption. Among non-Hispanic white subjects who report drinking alcohol, we observed an association between rs1229984, which encodes an arginine to histidine change in the amino acid sequence of alcohol dehydrogenase 1B (ADH1B), the major metabolizing enzyme of ethanol, and a reduction of 1.29 drinks per week per copy ($p=1.2*10^{-26}$). We also observed a similar effect, a reduction of 1.47 drinks per week ($p=6.9*10^{-5}$), in Latino subjects, and a more modest reduction of 0.38 drinks per week (p=0.05) in East Asian subjects. The smaller effect of rs1229984 on the number of drinks consumed per week in East Asian subjects is due to the presence in East Asian populations of rs671, a polymorphism that encodes an amino acid substitution from glutamine to lysine in aldehyde dehydrogenase 2 (ALDH2), the major metabolizing enzyme of acetaldehyde, the intermediate product in the ethanol metabolism pathway. In East Asian subjects, rs77768175, which is in strong linkage disequilibrium with rs671, was associated with a reduction of 0.80 drinks per week (p=0.008). More strikingly, subjects who carry the rs77768175 G allele are considerably less likely to report any alcohol consumption (OR = 0.32, $p=3.6*10^{-58}$).

1285F

Transethnic genome-wide meta-analysis of Alzheimer disease in Alzheimer's Disease Genetics Consortium. *G. Jun[1,2,3], J. Chung[1], J. Kozubek[1], C. Reitz[6], B.N. Vardarajan[6], K.L. Lunetta[2], J.L. Haines[7], M.A. Pericak-Vance[8], R. Kuwano[9], R. Mayeux[6], G.D. Schellenberg[10], L.A. Farrer[1,2,3,4,5], Alzheimer's Disease Genetics Consortium.* 1) Medicine (Biomedical Genetics), Boston University, Boston, MA; 2) Biostatistics, Boston University, Boston, MA; 3) Ophthalmology, Boston University, Boston, MA; 4) Neurology, Boston University, Boston, MA; 5) Epidemiology, Boston University, Boston, MA; 6) Neurology, Columbia University, New York, NY; 7) Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 8) The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 9) Molecular Genetics, Niigata University, Niigata, Japan; 10) Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA.

The Alzheimer Disease Genetics Consortium (ADGC) has performed genome-wide association (GWA) studies for late-onset Alzheimer's disease (LOAD) in European American (EA), African American (AA), and Japanese (JPN) populations. GWA analyses were conducted separately in samples containing 22,289 EA (11,641 cases, 11,341 controls), 5,028 AA (1,489 cases, 3,539 controls), and 1,845 JPN (951 cases, 894 controls) subjects using imputed genotypes for 7,911,420 SNPs in a logistic regression model adjusting for age and sex. Results from the three groups for the total sample (ALL) and subgroups of APOE ε4-positive(58% of cases and 27% of controls) and ε4-negative (38% cases and 69% controls) subjects were combined using meta-analysis. Top-ranked SNPs in transethnic meta-analysis were further examined for ethnic-specific or transethnic signals by evaluating results within each ethnic group. Significant associations (transethnic meta P-value $<10^{-6}$) were identified with SNPs in novel LOAD loci in the ALL (*MTOR*, *HBEGF*, *CASS4*, and near *ODZ2*) and in ε4-positive (*METTL19*, *GRIK2*, *GJC1*, *TRMT44*, *SH2D4B*, and near *EOMES*) groups. We observed a novel genome-wide significant (GWS) association ($P<5\times10^{-8}$) with a *GJC1* SNP in the ε4-positive group, but this finding was evident in Caucasians only (rs11871429: $P_{EA}=9.1\times10^{-9}$, $P_{AA}=0.83$, $P_{JPN}=0.17$, and $P_{EA\text{-}AA\text{-}JPN}=6.1\times10^{-8}$). Associations for other novel loci were most significant in the transethnic meta-analysis (top result: *SH2D4B* SNP rs12772279 in ε4-positive subjects; $P_{EA}=4.0\times10^{-5}$, $P_{AA}=0.015$, $P_{JPN}=0.16$, $P_{EA\text{-}AA\text{-}JPN}=2.8\times10^{-7}$). Top-ranked transethnic signals showed the same direction of the effect but the minor allele frequency varied substantially across the ethnic groups. In addition to genes in the *APOE* region, GWS results for previously established LOAD loci were observed with SNPs in *BIN1*, *PICALM*, *CR1*, *MS4A6E*, and *PTK2B* in ALL subjects (best SNP: rs586274 in *PICALM*; $P_{EA}=5.8\times10^{-11}$, $P_{AA}=0.633$, $P_{JPN}=3.0\times10^{-4}$, $P_{EA\text{-}AA\text{-}JPN}=1.1\times10^{-12}$). Among other established LOAD loci in EA, different SNPs showed the smallest P values in the AA or JPN samples, whereas top-ranked SNPs were not significant (P>0.1). An attempt to replicate and extend the top findings in independent Caucasian (N~9,000) and Hispanic (N~5000) samples is underway. This study suggests that the transethnic approach can detect novel LOAD susceptibility genes whose effects are either independent of or influenced by APOE genotype.

1286W

Pathway-based polygenic scores and variation explained by functional SNPs in schizophrenia. *K.K. Nicodemus[1], G. Donohoe[1], D. Tropea[1], D. Morris[1], K. Kendler[2], F. O'Neill[3], D. Walsh[4], B. Riley[2], M. Gill[1], A. Corvin[1], Wellcome Trust Case Control Consortium 2.* 1) Department of Psychiatry, Trinity College Dublin, Dublin, Ireland; 2) Virginia Commonwealth University, Virginia, USA; 3) Queen's University Belfast, Belfast, UK; 4) Health Research Board, Dublin, Ireland.

Schizophrenia (SZ) is a complex disorder with aetiology due to multiple genetic effects, which are likely to be enriched in specific pathways. Using experimentally-derived pathways from SZ-associated genes in the cellular adhesion molecule (CAM), DISC1, miR-137, NRXN1, TCF4 and ZNF804A pathways, we assessed the polygenic score contribution to variation explained in case status in the WTCCC2 schizophrenia case (N = 1378)-control (N = 1086) GWAS. Further, we assessed the contribution of functional SNPs (fSNPs) versus intronic SNPs not in LD with fSNPs in amount of variation explained.

We created polygenic scores based on p-value thresholds and weighted by the log10(odds ratio) from the PGC1 SZ GWAS. The number of genes/SNPs included were: CAM (125/6332), DISC1 (130/5384), mir-137 (474/21368), NRXN1 (131/5931), TCF4 (566/9424) and ZNF804A (138/3080). Restricting to fSNPs produced (% total): CAM 4.8, DISC1 5.8, miR-137 5.0, NRXN1 5.3, TCF4 8.9 and ZNF804A 8.5. Nagelkerke's R2 was estimated using logistic regression. P-values were obtained using a LRT. We compared the R2 from the set of functional SNPs to that obtained using 1000 same-size sets of randomly-selected intronic SNPs not in LD ($r2 < 0.2$) with the fSNPs.

The amount of variation explained (R2, LRT p) was: CAM (0.25%, 0.031), DISC1 (0.95%, 2.9e-05), TCF4 (0.76%, 0.00018), ZNF804A (0.91%, 4.3e-05), NRXN1 (0.27%, 0.026) and miR-137 (2.7%, 1.1e–12). The use of fSNPs explained a significant proportion of variation estimated (% R2 attributable to functional SNPs): CAM 84, DISC1 54, TCF4 34, ZNF804A 32, miR-137 5.2 and NRXN1 1.2. Further, the % of LE intronic SNP sets exceeding the R2 for the fSNPs was: CAM 1.9%, DISC1 8.2%, TCF4 29%, ZNF804A 0% miR-137 56.0% and NRXN1 92%, indicating fSNPs may explain more variation.

The polygenic score based on fSNPs explained a significant proportion of the total variation explained and was larger than that able to be explained using intronic SNPs in LE with fSNPs. Assuming equal variation explained by each SNP, the fSNP polygenic score should explain 4-9% of the total; in all but 2 pathways the percent variation explained by fSNPs was > 30%. The use of fSNPs may improve interpretability of the polygenic score by restricting to SNPs with known function and by reducing the number of SNPs to examine in follow-up studies.

1287T

Meta analysis identifies TSNARE1 as novel Schizophrenia / Bipolar susceptibility locus. *P. Sleiman[1], D. Wang[2], J. Glessner[1], D. Hadley[1], R.E. Gur[3], N. Cohen[2], Q. Li[2], H. Hakonarson[1], Janssen-CHOP Neuropsychiatric Genomics Working Group.* 1) Center Applied Genomics, CHOP, Philadelphia, PA; 2) 4Janssen Research & Development, LLC, Raritan, NJ; 3) Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104, USA.

We carried out a meta-analysis of combined schizophrenia, schizoaffective, and bipolar cohorts that resulted in the identification of six genome-wide significant loci, including one novel locus at chr8q24.3, encompassing TSNARE1 ($P=1.28\times10^{-9}$). The study included 13,394 schizophrenia and bipolar cases and 34,676 controls. Of these: 3,182 schizophrenia (of which 377 were classified as schizoaffective) cases and 1,032 bipolar I cases were collected from 28 clinical trials conducted by Janssen Research & Development, LLC. These samples were matched to 15,277 and 8,000 controls, respectively, from the biorepository at the Center for Applied Genomics (CAG) of the Children's Hospital of Philadelphia (CHOP). In addition, 1,157 cases meeting DSM-IV-TR criteria for schizophrenia or schizoaffective disorder from the Center for Applied Genomics (CAG) at The Children's Hospital of Philadelphia and the Department of Psychiatry at the University of Pennsylvania, School of Medicine and 2,107 controls from the biorepository at CAG were also included in the analysis. The remaining 8,023 schizophrenia cases and 9,292 controls were part of the Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC), as previously described, and were obtained from the NIMH as schizophrenia distribution 9 (https://www.nimhgenetics.org/). The function of the TSNARE1 gene remains unknown. A recent publication suggests it may have evolved within the vertebrate lineage from the harbinger transposon superfamily. Bioinformatic predictions based on phylogenetic ancestry indicate it may bind SNARE (soluble N-ethylmaleimide-sensitive factor attached protein receptor) proteins and have SNAP receptor activity. TSNARE1 may therefore have a vertebrate-specific function in intracellular protein transport and synaptic vesicle exocytosis.

1288F

Psychiatric Genomics Consortium quadruples schizophrenia GWAS sample-size to 35,000 cases and 47,000 controls. *S. Ripke, Schizophrenia Group of the Psychiatric Genomics Consortium.* ATGU, Mass Gen Hosp, Boston, MA.

The PGC (Psychiatric Genomics Consortium) is an international group of researchers whose major aim is to maximize the utility of extant psychiatric GWAS through mega-analysis. In a previous study, our first wave of genome-wide schizophrenia association analysis identified multiple loci involved in this genetically complex and clinically heterogeneous disorder (Nature Genetics, 2011). While around 20,000 individuals were necessary to achieve this result, detailed analysis of the data suggested that there are many more genes to discover, and that this should be possible by further increase of sample size.

Here we present an update of this international endeavor, which now comprises 35,476 schizophrenia cases and 46,839 controls coming from 52 substudies. The presented data is imputed into 1000 Genomes (Aug, 2012) and analyzed using standard logistic regression with ancestry components as covariates. All index SNPs with a p-value smaller than 1×10^{-6} were used for replication lookup in an independent GWAS analysis with 1,500 cases and 66,000 controls.

The number of independent genome-wide significant regions in this newest round of meta-analysis increased to 108 ($P < 5\times10^{-8}$). These results increase the number of loci strongly implicated in schizophrenia by more than 80. The loci implicated include prior targets (MIR137, CACNA1C, ZNF804A) along with a host of new targets many of which are now implicated by multiple lines of genomic evidence (DRD2, KCTD13).

These results are in line with prior predictions and developments in other complex disease GWAS with sufficiently large samples like Crohn's disease. They provide new insights into the biology of schizophrenia.

1289W

Polymorphisms in the CACNA1C gene region and amygdala activation during encoding of negative emotional stimuli. *L. Gschwind[1], C. Vogler[3], D. Coynel[2], A. Milnik[1], V. Freytag[1], K. Spalek[2], D. de Quervain[2,3], A. Papassotiropoulos[1,3,4].* 1) Molecular Neurosciences, University of Basel, Basel, Basel, Switzerland; 2) Cognitive Neuroscience, University of Basel, Basel, Basel City, Switzerland; 3) Psychiatric University Clinic, University of Basel, Basel, Basel City, Switzerland; 4) Department Biozentrum, Life Sciences Training Facility, University of Basel, Basel, Switzerland.

Background: Imaging genetics links brain imaging data (e.g. functional Magnetic Resonance Imaging (fMRI)) to genetic variation. Narrowing down the neuronal correlates of psychopathology, i.e. identifying patterns of alternated brain activation could serve as valuable endophenotype in genetic research. Differences in amygdala activation, a brain region that is involved in processing of emotional content, have frequently been reported comparing psychiatric patients to healthy controls. Recent studies reported genotype dependent activation differences of the amygdala for the intragenic SNP rs1006737 in CACNA1C (Wessa et al., 2010 and Tisli et al, 2013), a gene implicated in psychiatric disorders. We hypothesized that intragenic variation of CACNA1C might affect amygdala activation also in healthy individuals. Therefore we tested variants in the CACNA1C gene for association with mean amygdala activation in response to emotional stimulus material during an fMRI session in a large sample of healthy individuals. Methods: A sample of N=917 healthy young Swiss individuals encoded a set of 24 negative emotional pictures taken from the International Affective Picture System in a 3 Tesla fMRI scanner. Genotyping of the CANA1C gene was done using the Affymetrix Human SNP Array 6.0. A total of 183 SNPs fulfilled genotype quality criteria (MAF > 5 %, Deviation from HWE p>0.01). Automated segmentation procedures of structural Magnetic Resonance images were used to create individual masks for the accurate extraction of individual brain activation. Subsequently, amygdala activation was calculated contrasting negative emotional pictures vs. scrambled pictures. Genetic associations were tested under the assumption of an additive model using the Plink software package. Results: The lowest p-value out of the 183 SNPs assessed was found for SNP rs7139329 (p=0.00052), which did not survive conservative multiple testing (Bonferroni corr. p = 0.09). Further studies are warranted to elucidate the role of genetic variation of CACNA1C for amygdala activation in healthy individuals.

1290T

Genetic variation underlying amygdala-volume is highly enriched with schizophrenia susceptibility variants in healthy young individuals.
P. Lee[1,2,3], A. Holmes[1,4], P. Gallagher[1], L. Germine[1], M. Hollinshead[1], J. Roffman[1], R. Buckner[1,4], J. Smoller[1,2,3]. 1) Massachusetts General Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 4) Harvard University, Cambridge, MA.

Background: Variations in amygdala volume have been implicated in the pathophysiology of schizophrenia, bipolar disorder, and autism, but the genetic basis of these volumetric differences remains undefined. Here we report genome-wide association analysis of amygdala volume in a cohort of 1,426 healthy individuals and the genetic relationship between this brain structure and major neuropsychiatric disorders. **Methods:** Our study is based on the Harvard/MGH Brain Genomics Superstruct Project (GSP), a neuroimaging and genetics study of brain and behavioral phenotypes comprising more than 3,500 healthy subjects. Structural MRI images of amygdala volume were assessed for all GSP subjects using a standardized imaging protocol (3T, 12-channel coil, T1). Genome-wide genotyping was performed on 1870 subjects of European ancestry, yielding the discovery GWAS data of 1,140,419 SNPs. All study subjects were young adults with no history of psychiatric illnesses or major health problems ($18 \leq age \leq 35$). Imputation using the 1000 Genomes produced the allele dosages of 8,839,342 SNPs ($R2>0.3$; $MAF>0.01$), for which single-variant association was assessed using linear-regression. Covariates included age, gender, handedness, intracranial volume, scanner, console, and MDS factors to control for potential population sub-stratification. Genetic relationships of amygdala-volume-associated SNPs to major neuropsychiatric disorders were examined using multivariate enrichment analyses of genomic inflation factor, λ, as previously described by Stein et al. **Results:** Single-SNP-based linear-regression identified four genomic regions that show significant association with amygdala volume at $P < 1e\text{-}06$. While the top association locus on chromosome 3q24 (highest $P=4.8e\text{-}08$; $\beta=-81.33$mm3; SE= 14.90mm3) resides on a non-genic region, nearby brain-expressed genes of potential interest include PLOD2 and PLSCR4. Furthermore, enrichment analyses revealed inflated association of schizophrenia susceptibility variants [9] with amygdala volume differences ($\lambda=3.931$; $P<5e\text{-}05$). **Conclusions:** Using a neuroimaging GWAS analysis of healthy young adults, we identified genetic variants influencing individual differences in amygdala volume. We also found a potential role for schizophrenia susceptibility variants in modulating normal variation of amygdala volume, suggesting an etiologic link between amygdala structural changes and emotional/cognitive abnormalities present in this serious brain disorder.

1291F

Brain N-acetylaspartate is Increased in mouse models of Pelizaeus-Merzbacher disease: implication for human imaging findings.
K. Inoue[1], Y. Itoh[1], S. Saito[2], I. Aoki[2], J. Barkovich[3], J. Takanashi[2,4]. 1) Dept MR & BD Res, Natl Inst Neurosci, NCNP, Kodaira, Japan; 2) Molecular Imaging Center, National Institute of Radiological Sciences,Chiba, Japan; 3) Dept. Radiology & Biomedical Imaging, Univ. California San Francisco, CA , USA; 4) Dept. Pediatrics, Kameda Medical Center, Kamogawa, Japan.

In order to evaluate a hypothesis that hypomyelinating process may affect N-acetylaspartate (NAA) and N-acetylaspartylglutamate (NAAG) biochemical pathways, we examined brains of mouse models of Pelizaeus-Merzbacher disease. Two distinct lines, myelin synthesis-deficient (msd) and Plp1-transgenic (Tg) mice, each modeling point mutation and overexpression respectively, were subjected to study. We performed magnetic resonance imaging and proton magnetic resonance spectroscopy (1H-MRS) of the thalamus for msd and wild-type mice with a 7.0 tesla magnet. NAA and NAAG were independently measured by high performance liquid chromatography (HPLC). Immunohistochemical analysis using anti-Mbp, Gfap, Ng2, and NeuN antibodies were also performed. We also performed HPLC on Tg mice.In msd mice, 1H-MRS revealed increased tNAA, creatine, glutamine, and glutamate, and decreased choline (Cho). HPLC analysis revealed increases of both NAA and NAAG in the msd brains. Histologically, the msd brains revealed hypomyelination and astrogliosis. Oligodendrocyte progenitor cells and neurons were normal in number in the thalamus wherein 1H-MRS was obtained. In Tg mice, HPLC analysis also revealed increases of both NAA and NAAG. Despite of distinct molecular and cellular pathology between point mutation and overexpression, we observed similarly increased NAA and NAAG in both mouse lines, suggesting a presence of common downstream biochemical pathway. We hypothesized that a primary increase of NAA resulting from the lack of mature oligodendrocytes may lead to a secondary increase of NAAG. Such neurochemical derangement observed in the PMD mice may provide critical information in understanding MRS finding in human PMD patients with unique increase in total NAA. In conclusion, increased total NAA with decreased Cho detectable on 1H-MRS may be an important marker for PMD, which can be distinguished from more common neurological disorders that have decreased tNAA.

1292W

SZT2 mutations in infantile encephalopathy with epilepsy and callosal dysgenesis.
L. Basel-Vanagaite[1, 2, 3, 4], T. Hershkovitz[5], E. Heyman[6], M. Raspall-Chaure[7], N. Kakar[8], P. Smirin-Yosef[2], M. Vila-Pueyo[7], L. Kornreich[2], H. Thiele[10], H. Bode[9], I. Lagovsky[2], D. Dahary[11], A. Haviv[11], M. Pasmanik-Chor[12], P. Nürnberg[10], C. Kubisch[8], M. Shohat[1,2,3], A. Macaya[7], G. Borck[8]. 1) Rabin Medical Center, Petah Tikva, Israel; 2) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Felsenstein Medical Research Center, Petah Tikva, Israel; 4) Schneider Children's Medical Center of Israel, Petah Tikva, Israel; 5) Rambam Health Care Campus, Haifa, Israel; 6) Assaf Harofeh Medical Center, Zerifin, Israel; 7) Vall d'Hebron Research Institute (VHIR), Autonomous University of Barcelona, Barcelona, Spain; 8) Institute of Human Genetics, University of Ulm, Ulm, Germany; 9) Children's Hospital, University of Ulm, Ulm, Germany; 10) University of Cologne, Cologne, Germany; 11) Toldot Genetics Ltd., Hod Hasharon, Israel; 12) Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.

Epileptic encephalopathies are severe developmental disorders characterized by early onset epilepsy and poor neurological outcome. The genetic origins and underlying molecular processes of early onset epileptic encephalopathies have been described in only a few well defined syndromes. The objective of this study was to identify the disease-related gene in two unrelated patients with a distinctive phenotype characterized by refractory epilepsy, absent developmental milestones, thick corpus callosum and cavum septum pellucidum on brain MRI. In order to identify the disease-associated gene, we performed whole exome sequencing of the DNA of two probands, a 10 year-old girl of Iraqi Jewish origin and a 9 year-old Spanish boy who both presented with an infantile encephalopathy with refractory epilepsy. In both patients, we identified biallelic mutations of the Seizure Threshold 2 (SZT2) gene. The causative mutations include a homozygous nonsense mutation in the girl patient and compound heterozygosity for a nonsense mutation and an exonic splice site mutation in the boy. The latter mutation leads to exon skipping and premature termination of translation. Thus, all three mutations are predicted to result in nonsense-mediated mRNA decay and/or an N-terminal protein truncation and thereby loss of SZT2 function. SZT2 contains a putative superoxide dismutase motif and is presumably localized to the peroxisome where it has been suggested that it protects from oxidative stress and glutamate toxicity. Thus, glutamate-mediated excitotoxicity is a plausible mechanism underlying epileptogenesis caused by SZT2 mutations. We conclude that mutations in SZT2 are a novel cause of a severe type of autosomal recessive infantile encephalopathy with intractable epilepsy, thick corpus callosum and cavum septum pellucidum. While the cellular function of SZT2 is unknown and no diseases caused by SZT2 mutations have been reported in humans, Szt2 is known to influence seizure threshold and epileptogenesis in mice. Truncating mutations of Szt2 can confer low seizure thresholds and embryonic lethality in mice, phenotypes that are well consistent with the severe seizures and developmental arrest in our patients.

1293T

Association between Alzheimer's genetic risk scores and hippocampal and amygdala volume.
M.K. Lupton[1], W. Wen[2], K. Mather[2], H. Brodaty[2], D. Ames[2], J.N. Trollor[2], N.J. Armstrong[3], A. Simmons[4], S. Lovestone[4], J.F. Powell[4], P.S. Sachdev[2], M.J. Wright[1], GERAD1 Consortium, Addneuromed Consortium, Alzheimer's Disease Neuroimaging Initiative. 1) Neuroimaging Genetics, Queensland Institute of Medical Research, Herston, QLD, Australia; 2) Brain and Ageing Research Program, School of Psychiatry, University of New South Wales, Sydney, Australia; 3) Garvan Institute of Medical Research, Darlington, NSW 2010, Australia; 4) Institute of Psychiatry, King's College London, London. SE5 8AF, UK.

Alzheimer's disease (AD) is the most common form of dementia with an estimated 35.6 million sufferers worldwide . Recent GWAS findings have identified twenty one common genetic risk alleles, in addition to APOE ε4, which effect risk for late onset AD. However, there is limited understanding of how these genetic risk factors affect the magnetic resonance imaging phenotypes of hippocampal and amygdala volumes, which are the first to be affected in AD pathogenesis. Evidence of an early influence would increase our understanding of how these genetic variants work to affect risk. In addition, the identification of robust biomarkers would allow both screening and targeting of interventions to subpopulations of those most at risk. Here we test the hypothesis that there is an association between AD genetic risk variants and hippocampal and amygdala volumes in a large sample (N=2729). We included two AD case control cohorts from the Alzheimer's disease Neuroimaging Initiative (ADNI, N=743) and Addneuromed Consortium (N=356), two elderly (Memory and Ageing Study (MAS, N=542), Older Australian Twins Study (OATS, N=199) and one young (Queensland Twin Imaging (QTIM N=632) population cohorts. We tested for an effect in each cohort, and using Meta analysis, in AD and mild cognitive impairment (MCI) groups, as well as a healthy older group (with sample sizes of 271, 614 and 711 respectively). We found a significant association of the AD risk score with reduced hippocampal and amygdala volumes in both AD case control cohorts as well as reduced hippocampal volumes in the AD and MCI groups. The addition of the AD risk variants outside the APOE genotype did significantly strengthen association within these cohorts and disease groups. Importantly, we found no association between the AD risk score and the subcortical volumes in either the healthy older (N=704) or young (N=465) adult groups, and no association with the APOE ε4 genotype, which is in contrast to some previous underpowered studies. Although there is an effect of AD genetic risk scores on hippocampal and amygdala volume overall in case control cohorts, and in separate MCI and AD groups, there is no association in healthy older people. The lack of association, especially with APOE genotype calls into question the usefulness of these MRI measures in the prediction of future decline to MCI and AD. Alternatively the genetic and MRI measures presented here may be valid but independent predictors.

1294F

Genome-wide linkage analysis suggests oligogenic regulation of the human parieto-occipital 10-Hz rhythmic activity.
E. Salmela[1], H. Renvall[2], M. Vihla[2], J. Kujala[2], R. Salmelin[2], J. Kere[1, 3]. 1) Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland; 2) Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University, School of Science, Espoo, Finland; 3) Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden.

The human cerebral cortex shows several intrinsic oscillations (rhythms) that can be characterized with non-invasive neuroimaging methods such as magnetoencephalography (MEG). The most prominent of them is the 10-Hz rhythm recorded over the parieto-occipital cortices. This rhythm is strongly attenuated for example by opening the eyes, and its reactivity is widely used to probe cortical functions in both healthy and clinical populations. However, little is known about its underlying molecular mechanisms. To study the possible genetic determinants of the parieto-occipital 10-Hz rhythm in a normal population, we measured spontaneous brain activity with MEG in 98 pairs of healthy siblings while the subjects had their eyes closed and open (3 minutes each). The cortical activity was recorded with a 306-channel Elekta Neuromag neuromagnetometer, and amplitude spectra at each channel were calculated using Fast Fourier Transformation (FFT). The subjects were genotyped on Affymetrix 250K StyI SNP arrays, yielding genotypes for more than 28,000 single-nucleotide polymorphisms (SNPs) after filtering for genotyping quality and linkage disequilibrium.

The brain activity was quantified from the difference spectra between "eyes closed" and "eyes open" conditions. Width of the main spectral peak at ~10Hz, its peak frequency and peak strength were measured at the maximum channels in three regions over the parieto-occipital cortices. The peak strengths of the rhythm were highly heritable ($h^2 > 0.75$), and a variance component-based analysis revealed linkage for both the strength and the width of the spectral peak. The highest linkage peak was seen for the width of the spectral peak over the left parieto-occipital cortex on chromosome 10q23.2 (LOD = 2.814, $p < 0.03$ based on 1001 permutations with Merlin). This genomic region contains several functionally plausible genes, including *GRID1* which encodes a subunit of glutamate receptor channels that mediate the fast excitatory synaptic transmission in the central nervous system. Altogether, 12 suggestive or significant linkage peaks were seen across the 9 phenotypes tested; in 1001 permutations, the probability of observing 12 equally high peaks by chance was 0.046. Overall, the results demonstrate the potential of genetic analysis in linking macroscopic cortical phenotypes with molecular-level processes controlled by specific genes.

1295W

Genetic resilience to neurodegeneration in the presence of tau pathology.
T. Hohman, M. Koran, T. Thornton-Wells. Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

The disease cascade in Alzheimer's Disease (AD) involves tauopathies, β-amyloid plaques, progressive neurodegeneration, and cognitive impairment. Certain individuals present with pathologic protein biomarkers but no clinical manifestation of AD. The current project sought to identify genetic variants that modify the relationship between biomarkers of pathology and ventricular volume (neurodegeneration). We also investigated whether the observed protein-gene interactions had an effect on levels of neuroinflammatory markers in plasma. We used ventricular volume from Magnetic Resonance Imaging (MRI) images quantified in Freesurfer as our outcome measure. Our statistical model controlled for age, education, diagnosis, and gender. Our term of interest was a gene x Phosphorylated Tau (Ptau) interaction. Ptau was quantified from protein levels in cerebrospinal fluid (CSF). Genes were identified using a Genome-wide Association Study (GWAS) approach to look for significant single nucleotide polymorphism (SNP) x Ptau interactions. We corrected for multiple comparisons using a False Discovery Rate threshold of $p < 0.05$. In addition, we tested for consistency across two subsets of the ADNI study. We identified one SNP x Ptau interaction that was consistent across both the ADNI-1 and ADNI-2/GO data sources: POT1 (rs4728029; t = 5.269, p =1.82x 10-7). The POT1 x Ptau interaction explained two percent of the variance in ventricular volume even when including APOE and the other covariates previously mentioned in the model. POT1 expression has been related to telomere length in previous research (Kondo et al., 2004), and telomere length has been associated with a neuroinflammatory response in Alzheimer's disease (Panossian et al., 2003). Therefore, in posthoc analyses we tested the Ptau x rs4728029 interaction in relation to neuroinflammatory markers measured in plasma. Indeed, this interaction was related to levels of interleukin 6 receptor (IL6R), providing additional support for a possible neuroinflammatory mechanism that explains the observed genetic interaction. These results suggest that genetic variation may modulate the neuroinflammatory response to Alzheimer's tau pathology and ultimately modify risk for the neurodegenerative cascade. Future work will use a more focused pathway approach to further investigate genetic resilience related to a beneficial neuroinflammatory response.

1296T

Genetic interactions found between calcium channel genes modulate amyloid load measured by positron emission tomography. *M. Koran, T. Hohman, T. Thornton-Wells, Alzheimer's Disease Neuroimaging Initiative.* Vanderbilt University, Nashville, TN.

Late-onset Alzheimer's disease (LOAD) is known to have a complex, oligogenic etiology, with considerable genetic heterogeneity. We investigated the influence of genetic interactions between genes in the Alzheimer's disease (AD) pathway on amyloid-beta (Aβ) deposition as measured by AV-45 ligand positron emission tomography (PET) to aid in understanding this disease's genetic etiology. The Alzheimer's Disease Neuroimaging Initiative (ADNI) cohorts were used for discovery and replication. We discovered and validated a significant genetic interaction between the genes *RYR3* and *CACNA1C*, both of which encode calcium channels expressed in the brain. The results shown here support previous animal studies implicating interactions between these calcium channels in amyloidigenesis and suggest that the pathological cascade of this disease may be modified by interactions in the amyloid-calcium axis. Future work focusing on the mechanisms of such relationships may inform targets for clinical intervention. .

1297F

Genetic interactions within inositol-related pathways are associated with longitudinal changes in ventricle size. *T. Thornton-Wells[1], T. Hohman[1], S. Meda[2], M. Koran[1].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Olin Neuropsychiatry Research Center, Harford Hospital, Hartford, CT.

The genetic etiology of late onset Alzheimer disease (LOAD) has proven complex, involving clinical and genetic heterogeneity and gene-gene interactions. Recent genome wide association studies (GWAS) in LOAD have led to the discovery of novel genetic risk factors; however, the investigation of gene-gene interactions has been limited. Conventional genetic studies often use binary disease status as the primary phenotype, but for complex brain-based diseases, neuroimaging data can serve as quantitative endophenotypes that correlate with disease status and more closely reflect pathological changes. In the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort, we tested for association of genetic interactions with longitudinal MRI measurements of the inferior lateral ventricles (ILVs), which have repeatedly shown a relationship to LOAD status and progression. We performed linear regression to evaluate the ability of pathway-derived SNP-SNP pairs to predict the slope of change in volume of the ILVs. After Bonferoni correction, we identified four significant SNP-SNP interactions in the right ILV (RILV) corresponding to gene-gene pairs *SYNJ2-PI4KA, PARD3-MYH2, PDE3A-ABHD12B* and *OR2L13-PRKG1* and one significant interaction in the left ILV (LILV) corresponding to *SYNJ2-PI4KA*. The SNP-SNP interaction corresponding to *SYNJ2-PI4KA* was the same in both the RILV and LILV and was the most significant interaction in each (RILV: $p=9.13\times10^{-12}$; LILV: $p=8.17\times10^{-13}$). Both genes belong to the inositol phosphate signaling pathway which has been previously associated with neurodegeneration in AD.

1298W

Associations of the rs4818 polymorphism in the COMT gene with demyelination of the inferior longitudinal fasciculus in the white matter of the brain and cognition in schizophrenia patients. *Z. Kikinis[1], K. Green[1], M. Giwerc[1], S. Bouix[1], N. Makris[1], N. Schreiber[2], G. Corfas[1], R. Kucherlapati[1], R. Kikinis[1], M. Kubicki[1, 2], M.E. Shenton[1, 2].* 1) Harvard Medical School, Boston, MA; 2) Harvard Medical School, Brockton, MA.

A single nucleotide polymorphism (SNP), rs4818, located at the 4th intron of the Catechol-O-methyltransferase (COMT) gene, has been observed to cause changes in expression of the COMT protein. COMT is a schizophrenia susceptibility gene and variants of rs4818 have been reported to be associated with variations in cognitive performance in schizophrenia. However, the impact of rs4818 polymorphism on the white matter of the brain has yet to be investigated. Here, we compared CC, CG and GG carriers of the rs4818 SNP for changes in the inferior longitudinal fasciculus (ILF). The ILF is a long association fiber tract connecting occipital and temporal lobes in the white matter of the brain. We used MR-Diffusion Tensor Imaging (MR-DTI), to analyze white matter microstructure of this tract and we applied neuropsychological tests to assess cognitive performance. Methods: MR-DTI images were acquired on 30 patients with chronic schizophrenia and 26 matched healthy controls and tractography of ILF was performed. Changes in MR-DTI output measures, such as Fractional Anisotropy (FA), trace, Axial Diffusivity (AD) and Radial Diffusivity (RD), were analyzed as they might reveal microstructural changes of axons, such as changes in myelination. DNA was extracted from saliva. The effect of the rs4818 polymorphism on cognition was tested using scores of Wisconsin Card Sorting test (WCST). Results: Schizophrenia patients with the CC genotype (N=10) at rs4818 exhibited higher values in RD, higher values in trace and lower values in FA in the ILF in the right hemisphere, which were all statistically significant (p=0.008, p=0.02, p=0.05). Increases in RD, while no changes in AD and decreases in FA have been reported previously in animal studies as consequence of demyelination in brain white matter. Schizophrenia patients with CG/GG and the control groups with either genotype did not differ in the DTI values. Further, RD significantly and positively correlated to the scores of several subtests of the WCST (p<0.02) in the group of schizophrenia patients with the CC genotype at rs4818. This suggests that demyelination of the axon is associated with poorer performance on the test. Conclusion: Our results suggest that the CC genotype at rs4818 of the COMT gene is associated with demyelination of the white matter tract ILF in schizophrenia patients. These changes in myelination might affect working memory and cognition in schizophrenia patients with this particular genetic make up.

1299T

Identification of robust biomarkers of neuronal and glial metabolic changes in spinocerebellar ataxia type 1,2,3 and 7. *F. Mochel[1], I. Mawusi[1], T.M. Nguyen[1], D. Rinaldi[1], C. Jauffret[1], R. Valabregue[2], P.G. Henry[3], A. Brice[1], A. Durr[1].* 1) Department of Genetics, Hôpital de La Salpêtrière, Paris, France; 2) CENIR, Institut du Cerveau et de la Moelle, Paris, France; 3) University of Minnesota, Center for Magnetic Resonance Research, Minneapolis, Minnesota, USA.

Background: Spinocerebellar ataxias (SCAs) belong to the group of polyglutamine repeat disorders and lead primarily to neurodegeneration in the cerebellum and the pons. We recently demonstrated that even the most sensitive clinical scores would require large number of patients to assess any therapeutic benefit. Therefore, the identification of robust biomarkers is critical to assess disease progression for therapeutic development. Methods: 1H-NMR spectroscopy was performed at 3T to determine the neurochemical profile of 24 metabolite concentrations in the vermis and pons of a unique cohort of 66 SCA patients - SCA1 (N=18), SCA2 (N=13), SCA3 (N=22) and SCA7 (N=13) - as well as in healthy controls with similar median age (N=34). Results/Interpretation: Compared to controls, SCAs patients displayed a significant decrease of neuronal metabolites, N-acetylaspartate and glutamate, but increased glia-related metabolites, glutamine and myoinositol. The neuronal loss in both affected regions was associated with a significant increase in creatine and phosphocreatine suggesting compensatory energetic mechanisms. Of note, there was a strong negative correlation between ataxia rating score (SARA) and total N-acetylaspartate in the pons of SCA2 (r = -0.7942, p = 0.0186), SCA3 (r = –0.7868, p= 0.0008) and SCA7 (r = -0.7614, p = 0.0105). In the vermis, SARA score correlated with total creatine in SCA1 (r = 0.7573, p = 0.0007), SCA2 (r = 0.6310, p = 0.0278) and SCA3 (0.5679, p = 0.0072). A correlation with myoinositol was also found in SCA2 (r = 0.6142, p = 0.0336) and SCA3 (r = 0.5831, p =0.0055). The PCA confirm that neuronal metabolites (N-acetylaspartate and glutamate) in the vermis and the pons vary inversely to glia-related metabolites (myoinositol) and energy-related metabolites (creatine and phosphocreatine), suggesting compensatory mechanisms to the neuronal loss.

1300F

Analysis of Structural Variants in the DISC1 gene: Association Test with P300 ERP and Brain Volumes. *V. De Luca[1], G. Spalletta[2].* 1) Dept Neurogenetics, Univ Toronto, Toronto, ON, Canada; 2) Neuropsychiatry, IRCCS Santa Lucia Foundation, Italy.

In a large Scottish family with a (1;11)(q42;q14.3) translocation, the carriers of this balanced translocation were significantly associated to reduction in the amplitude of the P300 event-related potential (ERP). In healthy controls, we tested the association between two non-synonymous SNPs in DISC1 gene with P300 amplitude and the volumes of the brain structures that are the putative P300 generators. We have investigated the measurement of the visual P300 event-related potential (ERP) during a modified Stroop task. For this study, 32 healthy subjects (drug free) were recruited for the ERP arm and 100 subjects were recruited for the structural MRI arm. The mean age in the ERP was 34.53+/−8.5. There were 18 males and 14 females. The analysis of the SNP rs 821616 (Cys704Ser) was performed using the ABI7000. For the neutral word condition, there was no difference between the subjects who were carrying the variant 704Ser and those who did not carry this variant [F (1/ 26) = 1.732, p=0.200]. In the structural MRI arm, we did not find association with the SNP rs821616. Although our results did not show an association between Cys704Ser in DISC1 and P300 amplitude, this pilot study suggests a new methodology to combine multiple samples with related endophenotypes to test the pleiotropic effect of specific functional variants.

1301W

Clinical and Molecular Investigation of Spinocerebellar Ataxia with Hypogonadism: Not So Far, Not So Close. *C. Lourenco[1], C. Sobreira[1], P. Frassinete de Medeiros[2], M. Gonzalez[3], F. Speziani[3], S. Zuchner[3], W. Marques Jr[1].* 1) Neurology, Univ Sao Paulo, Ribeirao Preto, SAO PAULO, Brazil; 2) Genetics Service , Federal University of Campina Grande, Campina Grande, Brazil; 3) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

BACKGROUND: The association between cerebellar ataxia and hypogonadism was first described in four sibs by Holmes in the beginning of the XX century, and has since become known as Holmes type ataxia. Several syndromes with hypo/hypergonadotrophic hypogonadism and ataxia have been published, however there is a remarkable clinical heterogeneity among them. Here, we present the clinical data and molecular/biochemical studies of nineteen Brazilian patients with cerebellar ataxia and hypogonadism. MATERIAL AND METHODS: All patients were evaluated in the neurogenetics clinics by geneticists, neurologists and endocrinologists. Brain MRI, ophtalmological exam, EMG/NCV, hormone and biochemical tests, screening for CDG disorders (IEF and MS analysis), karyotype, muscle biopsy with chain respiratory enzyme assays and measuremente of coenzyme Q10, molecular tests for Friedreich ataxia and for SCAs (types 1, 2, 3, 6 and 7) were performed in the course of the investigation.Exome sequencing as performed in selected patients with a consistent phenotype. RESULTS: All patients had cerebellar ataxia, but the age of the onset was variable; it was worthy to note that ten patients had early onset ataxia (in the first decade of life). Consanguinity of parents was noted in three families; seven patients had hypergonadotropic hypogononadism. Mental retardation was seen in five patients with hypergonadotropic hypogonadism. None of the patients had chromosomal anomalies. Molecular tests for Friedreich and SCAs 1, 2 ,3 , 6 and 7 were all negative. Optic atrophy and retinochoroidal degeneration were found in five patients; axonal neuropathy was present in four patients. Cerebellar atrophy with pons or prominent vermis involvement was a constant feature. In two patients with ataxia and hypergonadotropic hypogonadism, coenzyme Q10 deficiency was confirmed in muscle biopsy.Two unrelated adult patients with hypergonadotropic hypogonadism had biochemical features of CDG Ia. One family - with four affected sibs - have features consistent with a rare neurological disorder , Boucher-Neuhauser syndrome. CONCLUSIONS: The association between cerebellar ataxia and hypogonadism comprise hetereogenous entities whose clinical investigation can elighten the pathological basis of these fascinating neuroendocrinological syndromes. Screening for CDG and Coq10 deficiency should be done in such patients as a part of the work-up investigation.

1302T

Novel ATM mutation in late-onset autosomal recessive cerebellar ataxia with neuropathy. *H. Shimazaki[1], R. Sugaya[1], J. Honda[1], A. Meguro[2], I. Nakano[3].* 1) Neurology, Jichi Medical University, Shimotsuke, Tochigi, Japan; 2) Hematology, Jichi Medical University, Shimotsuke, Tochigi, Japan; 3) Neurology, Tokyo Metropolitan Neurological Hospital, Fuchu, Tokyo, Japan.

Backgrounds: Autosomal recessive spinocerebellar ataxias (ARSCA) comprise many types of diseases. Most frequent ARSCA is Friedreich ataxia, but other types are relatively rare. We encountered a consanguineous family with two patients of late-onset cerebellar ataxia with neuropathy. We attempt to identify the causative gene mutation of this family with ARSCA. Methods: We investigated the proband with detailed neurological examination, blood examination, brain MRI, electrophysiological study, nerve biopsy and gene analyses. Results: The neurological examination revealed cerebellar ataxia, hand tremor, distal muscle wasting and diminished tendon reflexes. Patients had no conjunctival telangiectasias and showed immunodeficiency. Blood examination showed slightly elevated AFP. Brain MRI demonstrated marked cerebellar atrophy. Electrophysiologic study and nerve biopsy showed axonal neuropathy. We could not detect GAA repeat expansions of FXN gene, and sequencing analyses could not reveal pathologic substitutions in the APTX, SETX, and TDP1 genes. Whole-exome sequencing (WES) could identify the novel homozygous missense mutation in the ATM gene. This homozygous mutation was found in another patient, co-segregated within the family members and not found in 200 Japanese control DNAs. Subsequently, the patients suffered and died from bile duct cancer and chronic lymphocytic leukemia. Conclusion: We could identify ATM mutations in adult, late-onset ARSCA family. Ataxia-Telangiectasia (AT) is usually early-onset and show immunodeficiency and telangiectasia. We should consider AT even in late-onset ARSCA without telangiectasia and immunodeficiency. WES is one of the useful methods for identify the causative mutation in atypical adult-onset autosomal recessive cerebellar ataxias.

1303F

Modelling collybistin deletion neuropathology in induced pluripotent stem cells (iPSC)-derived neurons. *A. Sertie[1,2], K.G. Oliveira[1], C.O.F. Machado[1,2], C. Rosemberg[1], F. Kok[1,3], M.R. Passos-Bueno[1].* 1) Genetics and Evolutionary Biology Department, Institute of Biosciences, University of São Paulo, Brazil; 2) Experimental Research Center, Albert Einstein Jewish Hospital, Sao Paulo, Brazil; 3) Clinics Hospital, School of Medicine, University of São Paulo, Brazil.

Collybistin (CB) is a neuron-specific guanine nucleotide exchange factor implicated in inhibitory synapse development and plasticity that cluster and localize gephyrin and inhibitory neurotransmitter receptors to the postsynaptic membrane. We have identified a Brazilian patient with a deletion of the entire collybistin gene (ARHGEF9, Xq11.2) who shows severe mental retardation, epilepsy, and autistic behavior. In order to gain further insight into the underlying cellular neuropathology, we generated iPSC-derived neural progenitor cells (NPCs) and neurons from this patient and two unaffected individuals. iPSC were generated from skin fibroblasts using standard retroviral reprogramming technology. iPSC differentiation into NPCs and neurons was performed under conditions that favor the generation of either excitatory or inhibitory neurons. Expression of pluripotency-, NPC- and neuron-specific markers was determined by immunocytochemistry and/or western blotting. mTOR signaling pathway analysis was performed in NPCs treated nor not with rapamycin by western blotting. We observed that all control and patient iPSC clones express pluripotent markers such as Nanog, Oct4, Sox2 and Lin28. NPCs were positive for early neural precursor markers, such as Nestin and Musashi1. Mature neurons were positive for neural markers β-III Tubulin and Map2; in addition, these cells also express PSD-95 and GABARγ2 proteins, markers for excitatory and inhibitory synapses respectively. Interestingly, collybistin expression, which is not detected in normal fibroblasts, was reactivated in control IPSC, NPCs and mature neurons. We did not detect any significant differences between patient and control cells with respect to the reprogramming and differentiation capacities. However, iPSC and NPCs derived from the patient showed increased proliferation rate. Because mTOR signaling has been linked to cell proliferation and the CB partner gephyrin interacts with mTOR, we investigated the mTOR signaling pathway in patient-derived NPCs. Compared to control NPCs, our preliminary results suggested altered mTOR signaling in patient NPCs. Together, our results suggest that iPSC-derived NPCs and neurons from the patient carrying a deletion of the CB gene provide very promising model systems to explore the roles of CB in human neural development and physiology, as well as the mechanisms underlying the cognitive impairment in CB-deficient patients. Support: FAPESP, CNPq, Autismo & Realidade.

1304W

Genetic basis of Math genetic difficulty. *M.R.S. Carvalho[1, 3], M.de Miranda[1, 3], G. Pena[6], A. Julio-Costa[4], P. Pinheiro-Chagas[4, 7], L. Salvador[2], M. Andreata[1], V.G. Haase[2, 4, 5].* 1) Departamento de Biol Geral, Univ Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 2) Departamento de Psicologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 3) Pós-Graduação em Genética, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 4) Pós-Graduação em Neurociências, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 5) Pós-Graduação em Saude da Criança e do Adolescente, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 6) Pós-Graduação em Enfermagem, Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; 7) Inserm-CEA Cognitive Neuroimaging unit NeuroSpin Center.

Math learning difficulty (MD) is a common cognitive condition affecting from 3 to 6% of the school-age population. Impacts of MD are low scholar achievement, low self-esteem, low employability, and low income. Social and economical factors and low educational standards may contribute to MD. However, a primary, genetic predisposition was also proposed. Familial aggregation has been reported for MD, but there is only two reports in the literature analyzing family history. Heritability and relative risks were never estimated. This project was approved by the Ethics in Research Committee of the Universidade Federal de Minas Gerais. It is part of a population based case-control study, in which a random sample of the school age population of Belo Horizonte city, Brazil, was collected. In order to ascertain individuals with MD, a two phase strategy was adopted. In the first phase, a Math achievement test (TDE) was responded by 1564 children. Students with results below the percentile 25 in TDE were further evaluated with a battery of tests covering intelligence, attention, mental state, memory, working memory, fine motor coordination, language, somato-sensorial function. Students with IQ<85 were excluded. A family history was collected with an specifically developed, structured questionary composed by simple questions such as 'can he/she add?', 'can he/she subtract?', 'can he/she multiply?', 'can he/she divide?', 'can he/she exchange money', 'can he/she discriminate left and right', 'can he/she tell the time on the clock?', 'can he/she go around by himself/herself', 'can he/she write?', 'can he/she read?'. Besides, data were collected on health conditions, consanguinity, scholarship, profession, and familial income. Fifty-eight family histories were collected (34 cases and 24 controls). Information was obtained on 1778 relatives, 981 biological ones. MD frequency among the 1564 students that took part in the first phase was estimated in 7%. No significative differences in sexual proportion were observed among MD index cases. Familial recurrence was reported in 32/34 families of the affected children. MD was reported in 27% of the mothers, 16% of the fathers, and 40% of the brothers and sisters of the affected children. These results are similar to those found in the two previous reports. Heritability was estimated in 70% for MD. Both autosomal dominant and X-linked recessive inheritance were detected among the MD families. Support: CNPq, FAPEMIG.

1305T

Whole Transcriptome Analysis of Fetal and Adult Ts1Cje Brains Provides New Perspectives for Prenatal and Postnatal Treatment of Neurocognition in Down Syndrome. *F. Guedj[1], J.LA Pennings[2], L.C Graham[1], E.L Newman[3], K.A Miczek[3], D.W Bianchi[1].* 1) Mother Inftant Research Institute, Tufts Medical Center, Boston, MA, USA; 2) Laboratory for Health Protection Research (GBO), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; 3) Department of Psychology, Tufts University, Bacon Hall, 530 Boston Ave, Medford, MA, USA.

Background: Learning deficits in Down syndrome (DS) emerge in early childhood and are associated with microcephaly and reduced neurogenesis. Although these anomalies originate during fetal life, pre-clinical trials are mainly focused on adults. We have proposed using non-invasive prenatal testing (NIPT) as an opportunity for prenatal treatment to improve cognition in DS (PrenatDiagn 2013;33:614). We hypothesize that prenatal treatment of brain and cognitive anomalies can be achieved using a systems biology approach. As baseline information, we analyzed whole transcriptome changes in embryonic and adult brains of the Ts1Cje mouse model of DS, as well as neonatal behavior via the ultrasonic vocalization (USV). **Methods:** For gene expression studies, total RNA was prepared from adult cortex and hippocampus (Ts1Cje=6/WT=5) and E15 whole brains (Ts1Cje=5/WT=5) for hybridization to Affymetrix mouse gene 1.0ST arrays. Data were analyzed to identify genes that are differentially regulated. Functional analyses were performed using GSEA and DAVID to identify altered signaling pathways in Ts1Cje brains. For behavioral studies, P7 pups were separated from their dams and the number of USVs was recorded. Motor activity was assessed by estimating the number of grid crossings and rollings on the testing platform. **Results:** More genes were significantly differentially regulated in the Ts1Cje vs. WT E15 brains (n=71) compared to the hippocampus (n=30) and cortex (n=7) at FDR-BH < 20%. Interestingly, similar regulation patterns are observed at both stages. Functional analyses highlighted imbalances of several cellular pathways (inflammation, oxidative stress, apoptosis and G-protein signaling) at both stages. Importantly, cell cycle genes (28 genes) and amino acid transporters (7 genes) are exclusively affected in fetal brain. Ts1Cje neonates produced more ultrasonic vocalizations (sum of ranks= 274.50) than WT pups (sum of ranks=221.50) (p<0.01,Mann-Whitney test). They also display abnormal motor activity indicated by reduced number of grid crossings (Sum of ranks=141.50) versus WT (Sum of ranks=141.50) (p<0.01). **Conclusions:** Data acquired from analysis of the brain transcriptome and early neonatal behavioural screening in Ts1Cje mice suggest novel pathway differences and behavioral abnormalities that are already present during fetal and neonatal life in DS. These differences provide baseline information on which to evaluate the effect of in utero treatment strategies.

1306F

Monoamine Oxidase Deficiency: The clinical relevance of personal genomics in a new developmental brain dysfunction disorder. *D. Moreno-De-Luca*[1], *E.R. Riggs*[2], *D.H. Ledbetter*[2], *C.L. Martin*[2], *J.F. Cubells*[3]. 1) Department of Psychiatry Yale University 300 George Street, Suite 901 New Haven, CT 06511; 2) Geisinger Health System 100 North Academy Avenue Danville, PA 17822; 3) Department of Human Genetics Emory Autism Center Department of Psychiatry and Behavioral Sciences Emory University School of Medicine Woodruff Memorial Building, Suite 7213 1638 Pierce Drive Atlanta, GA 30322.

Now considered the standard of care for developmental disabilities, chromosomal microarray testing has opened the door for detailed phenotypic delineation of new neurodevelopmental syndromes, defined by specific molecular criteria. Here we describe the phenotypic, familial, and genetic characterization of a male patient with a deletion involving the X-linked MAOA and MAOB genes, essential for amine-neurotransmitter homeostasis. The male proband presented at the age of 27 for psychiatric care. Facial dysmorphology, autism and ID, plus a family history remarkable for four maternal uncles with ID and premature deaths due to stroke, prompted evaluation by array comparative genomic hybridization (aCGH). We used the International Collaboration for Clinical Genomics (ICCG, formerly ISCA) consensus microarray design, containing 180K oligonucleotide probes across the whole genome. To fine-map initial results, we used a custom-designed high-density exon targeted microarray based on the same platform. aCGH revealed a 897 kb deletion on the X chromosome (chrX:42,625,499-43,523,142) that entirely removes MAOA and disrupts the coding region of MAOB. The patient's phenotype including ID, complete lack of productive language and very limited receptive language, complex wringing limb and body movements, constant motor restlessness, severe constipation with radiographic evidence of a flaccid large bowel, self-injurious behavior (hand-biting), and frequent brief syncopal episodes, is remarkably similar to other individuals with deletions involving MAOA and MAOB, but strikingly different from Brunner syndrome, arising from a point mutation in MAOA alone. The family was advised to refrain from giving the patient foods containing tryamine, and they were provided a list of sympathomimetic medications to avoid. The clear genotype-phenotype correlation in this individual, and the dramatic family history of premature death, is similar to the presentations of other males with deletions disrupting both MAO-encoding genes. Our observations and prior reports establish MAO deficiency as a new developmental brain dysfunction disorder with explicit clinical implications, namely the need for a low tyramine diet and clear contraindication of sympathomimetic agents. Knowledge of the genetic etiology of the phenotype in these patients allows for a tailored clinical management targeting directly the deficient pathway and serves as an example of personalized medicine.

1307W

VRK1 mutations associated with complex motor and sensory axonal neuropathy plus microcephaly. *W. Wiszniewski*[1], *C. Gonzaga-Jauregui*[1], *T. Lotze*[2], *L. Jamal*[3], *S. Penney*[1], *I. Campbell*[1], *D. Pehlivan*[1], *J. Hunter*[2], *S. Woodbury*[2], *G. Raymond*[4], *A. Adesina*[2], *S. Jhangiani*[5], *J. Reid*[5], *D. Muzny*[5], *E. Boerwinkle*[6], *R. Gibbs*[5], *J. Lupski*[1], *Centers for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Texas Children's Hospital, Houston, TX; 3) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD; 4) Department of Neurology, University of Minnesota, Minneapolis, MN; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) Human Genetics Center and Institute of Molecular Medicine, University of Texas-Houston Health Science Center, Houston, TX.

Here, we report three children from two unrelated families with a previously uncharacterized complex axonal motor and sensory neuropathy accompanied by severe non-progressive microcephaly and cerebral dysgenesis. We performed whole-genome and targeted whole-exome sequencing in these three affected subjects. Using genome-wide sequence analysis, we identified compound heterozygous mutations in two affected siblings from one family and a homozygous nonsense mutation in the third unrelated patient in the vaccinia-related kinase 1 (VRK1) gene. VRK1 encodes a serine/threonine kinase that is crucial for cell cycle progression and cell division and is proposed to be involved in nervous system development and maintenance. It is an early response gene that directly phosphorylates and regulates p53 and has crucial roles throughout the cell cycle. It has been proposed that VRK1 acts downstream of CLP1, and the disruption of the latter was shown recently to result in abnormal neuronal apoptosis leading to a similar neurodegenerative phenotype defined by microcephaly, brain dysgenesis and axonal neuropathy in mice. We hypothesize that a similar mechanism of abnormal neural apoptosis may explain the neurologic phenotype observed in patients with VRK1 mutations. Patients with rare diseases and complex clinical presentations represent a challenge for clinical diagnostics. Genomic approaches are allowing the identification of novel variants in genes for very rare disorders enabling a molecular diagnosis. Genomics is also revealing a phenotypic expansion whereby the full spectrum of clinical expression conveyed by mutant alleles at a locus can be better appreciated.

1308T

Evidence for phenotypically-linked bounds on the role of *de novo* loss of function variation in autism spectrum disorders. *E.B. Robinson*[1, 2, 3, 4], *K.E. Samocha*[1, 2, 3, 4], *J. Kosmicki*[1, 3, 5], *L.M. McGrath*[6], *B.M. Neale*[1, 2, 3, 4], *M.J. Daly*[1, 2, 3, 4]. 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Medical and Population Genetics Program, Broad Institute of Harvard and MIT, Cambridge, MA; 4) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA; 5) Center for Biomedical Informatics, Harvard Medical School, Boston, MA; 6) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA.

Background The distribution of cognitive ability in individuals with autism spectrum disorders (ASDs) ranges from severe impairment to high intelligence. While ASD with intellectual disability has been associated with a greater frequency of deleterious genetic events than ASD alone (e.g. Girirajan et al. 2012), the extent to which cognitive variation can be used to delineate bounds on the influence of *de novo* variation in ASDs has not been well characterized. The goal of this study was to determine whether *de novo* loss of function (LOF) mutations were associated with ASDs across the range of cognitive ability in 955 sequenced trios from the Autism Consortium and Simons Simplex Collection. Method Participants were included if they attempted any one of the 8 IQ tests used across the consortia that offered a comparable estimate of full scale IQ (e.g. the Differential Ability Scales, the WISC; n=801). To maintain adequately powered subsets, we divided the sample into three groups: 1) measured IQ at or above average (≥100; n=229), 2) measured IQ below average IQ (<100; n=457), and 3) IQ test attempted but not successfully completed (n=115). We compared the observed rate of *de novo* LOF mutations per exome in each group against the expected rate in the general population as estimated in the Exome Sequencing Project (0.09 per exome). Results The ASD sample as a whole featured an excess of *de novo* LOF mutations (observed rate= 0.13, p=2.05e−7). However, there was no *de novo* LOF excess in the group with IQ≥100 (observed rate=0.08, p=0.59). The association between ASDs and *de novo* LOF variants became stronger when evaluated separately within the groups with measured IQ<100 (observed rate=0.16, p=2.37e-7) and IQ attempted by not successfully completed (observed rate=0.21, p= 9.72e-5). The difference in *de novo* LOF rate between the IQ≥100 and other groups was also statistically significant (p=0.001). Discussion These findings suggest that *de novo* LOF variation does not play a role in ASDs when intelligence exceeds the general population average. Further, they suggest that ASDs include etiologic subtypes that can be distinguished through phenotypic variation. We discuss these findings in terms of their relevance to study design in ASD genetics.

1309F

Phenotypic categorization of putative pathogenic Copy Number Variants (CNVs) in a population of Autism Spectrum Disorder (ASD) patients. *I.C. Conceição[1,2,3], B. Oliveira[1,2,3], C. Correia[1,2,3], M. Rama[1], J. Coelho[1], C. Café[4], J. Almeida[4], S. Mouga[4,5], F. Duque[4,5], G. Oliveira[4,5,6], A.M. Vicente[1,2,3].* 1) Instituto Nacional de Saúde Dr Ricardo Jorge, Lisbon, Portugal; 2) Center for Biodiversity, Functional & Integrative Genomics, Lisboa, Portugal; 3) Instituto Gulbenkian de Ciência, Oeiras, Portugal;; 4) Unidade Neurodesenvolvimento e Autismo, Centro de Desenvolvimento, Hospital Pediátrico (HP) , Centro Hospitalar e Universitário de Coimbra (CHUC), Coimbra, Portugal; 5) Instituto Biomédico de Investigação em Luz e Imagem, Faculdade de Medicina da Universidade de Coimbra, Portugal; 6) Faculdade de Medicina da Universidade de Coimbra, Portugal ;.

The in depth characterization of CNVs in patients with ASD is fundamental to improve the distinction between benign and disease-causing structural variants, defining the clinical and genomic attributes with better predictive power for the development of CNV categorization methods for diagnostic purposes. In this study we characterize and categorize CNVs of clinical significance present in a population of 342 ASD patients, genotyped by the Autism Genome Project using SNP arrays. We selected all high confidence genic CNVs that did not overlap more than 20% in sequence with benign CNVs identified in public databases control populations (N=5459), and included CNVs containing syndromic genes/gene regions associated with ASD. A total of 180 genic CNVs were identified, ranging from 5 kb to 3 Mb, 62% of which were deletions. Genic CNVs included between one (73% of all genic CNVs) to 25 genes in a single CNV. Most CNVs (86%) were present in a single individual, but 23 common CNVs with a frequency of 1% or higher, distributed in 6 genomic regions, were identified in 26 individuals. Each of these common CNVs were present in 3 to 6 individuals, and encompassed candidate gene/gene regions for ASD, such as 16p13.11, *DPYD*, *PARK2* and *VPS13B*. Network analysis of the 314 genes encompassed by these CNVs yielded a network including 85 genes, enriched in central nervous system development and regulation of cellular catabolic processes. 23.5% of genic CNVs were *de novo*. CNV inheritance was further explored using parental information on autistic traits assessed using the Social Responsiveness Scale (SRS) and the Broad Autism Phenotype Questionnaire (BAPQ). A significant excess of BAPQs positive for autistic traits was observed in CNV-transmitting fathers, mainly in the 'aloof' personality. Familial correlations assessed for all pairs using the SRS for parents and probands showed a significant correlation between parents, suggestive of assortative mating in ASD. CNVs were further categorized according to various phenotypic attributes. We observe a trend towards individuals with intelectual disability, dysmorphisms and family history of neuropsychiatric disorder having a higher CNV burden, larger average CNVs sizes and a higher number of genes *per* CNV and *per* individual. Data mining algorithms are being employed for additional detailed categorization of specific CNVs according to clinical data, including ASD severity, social communication, language, regression and epilepsy.

1310W

Fragile X, intermediate, and premutation alleles in the Autism Genetic Resource Exchange (AGRE). *W. Brown, A. Glicksman, X. Ding, N. Ersalesi, C. Dobkin, S. Nolin.* Dept Human Gen, NYS Inst Basic Res, Staten Island, NY.

AGRE is an autism family registry and resource that includes primarily multiplex families, having two or more children affected by autism spectrum disorders. Family pedigree, phenotypic and genotypic data, and genetic material is available. They represent a relatively unbiased sample. ~35% of the first 480 families were noted to have had some genetic testing prior to entering the registry. The individuals receive ADI-R and ADOS. We have screened the FMR1 locus in one proband from each of 1742 families. We found 8 families that had the fragile X mutation present. Among these, 6 were found among the first set of 480 families. The prevalence of fragile X among the ~312 AGRE families that had had no prior genetic screening was ~ 1.9%. An estimate of the IQ score of the autistic subjects was 98±23 by Raven and 90±21 byStanford-Binet testing with 10-14% <70. Thus, the AGRE sample has a higher IQ distribution than typical for fragile X subjects (mean ~40±25). Previous prevalence studies of fragile X in autistic samples range from 0 to 16%; with a mean of ~4%; (Feinstein 98). Our initial 1.9% is similar to a report of 1.6% among 123 unrelated autistic individuals (Bailey 93), but lower than the 13% we found on an earlier multicenter study of 183 individuals (Brown 86). A growing awareness of fragile X syndrome has decreased the probability of fragile X in these multiplex autism families due to screening and exclusion from AGRE, and due to higher shifts in the IQ distribution for more recent autism cohorts. The overall observed frequency was was 8/1742 or ~ 0.5%. This finding still confirms an association of fragile X and autism exists. We tested to see if there is an association of autism with premutations or intermediate alleles. Among the 1535 male probands tested, there were 2 with premutation (59 & 64 CGGs) and 12 with intermediate (45-54 CGGs) alleles for an intermediate prevalence of 0.78%. Among the 206 female probands tested there were 2 with premutations (55, 59) and 7 with intermediate alleles. Since females have two alleles, dividing by 2 gives an intermediate allele prevalence of 1.7% in female alleles or an overall intermediate allele prevalence of 0.98%. Our control value was 1.15%. Thus, there was no excess of intermediate or premutation alleles among the AGRE registry autistic probands. This finding indicates autism is NOT associated with intermediate (45-54) or premutation (55-200) alleles in the AGRE probands.

1311T

Genome-wide association study identifies variants for major depression through age at onset stratification. *R.A. Power[1], K. Tansey[1], H. Buttenschøn[2], H. Lee[3], S. Cohen-Woods[4], S. Ripke[5], N.R. Wray[3], P.G.C. MDD Working Group[6], C.M. Lewis[1].* 1) Social Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, London, United Kingdom; 2) Institut for Klinisk Medicin, Aarhus Universitet, Denmark; 3) Queensland Brain Institute, University of Queensland, Australia; 4) University of Adelaide, Australia; 5) Massachusetts General Hospital Simches Research Center, USA; 6) Consortium.

Major depressive disorder (MDD) is a highly prevalent and disabling disease. Despite moderate heritability and the pooling of large GWAS samples, no replicable genetic variants have been established for MDD. Here we used age at onset (AAO) to dissect the considerable heterogeneity within MDD, and better understand the underlying genetic architecture of this disorder. In the Psychiatric Genomics Consortium's sample of 9 studies containing 8,920 MDD cases and 9,521 controls, we split cases into 8 subsets based on AAO percentile within their study (i.e. first 8th of earliest onset cases, second 8th of earliest onset, etc.). These octiles were analysed systematically against controls to look at both early and late-onset specific risk factors. A replication threshold of p=5E−7 was used, with SNPs replicated in a sample of 6 studies containing 4,922 cases and 15,141 controls. GCTA and polygenic profile scoring were used to examine the heritability of early vs. late onset MDD, and how it overlapped with other psychiatric disorders. Ten independent SNPs met our replication threshold with AAO specific MDD sub-groups, of which two were genome-wide significant in the meta-analysis with the replication sample. These were rs7647854 on chromosome 3 (p=1.0E−11, OR=0.81) and rs7950328 chromosome 11 (p= 1.0E-08, OR=1.35). Further analysis showed that the effect of these SNPs was specific to the 50% latest onset cases, and increased in effect size as AAO increased. With a median AAO at age 27 across samples, 'late-onset' here reflects an adult-onset form of MDD rather than one developing in old age. We found no difference in the heritability of early and late-onset MDD, though did find a significantly greater burden of schizophrenia and bipolar risk variants in early-onset cases. In this study we have identified the first genome-wide significant replications for major depression and helped redefine its clinical picture. Our results suggest that adult-onset MDD is a more homogenous disorder with risk variants of larger effect, while adolescent-onset MDD is genetically more similar to both schizophrenia and bipolar disorder. We also highlight how the approach of more homogenous phenotyping can increase power beyond simply increasing sample size.

1312F

Divergent and convergent quantitative dysmorphic phenotypes among neuropsychiatric patients with rare de novo Copy Number Variants.
C.K. Deutsch[1], F. Momen-Heravi[2], R. Francis[1], A.T. Hunt[1], J.M. Stoler[3], L.B. Holmes[4], J. Sebat[5]. 1) Psychobiology Program, Eunice Kennedy Shriver Ctr, Waltham, MA; 2) Harvard Catalyst, Harvard Medical School, Boston, MA; 3) Clinical Genetics, Boston Children's Hospital, Boston, MA; 4) Medical Genetics Unit, MassGeneral Hospital for Children, Boston, MA; 5) Beyster Center for Molecular Genomics of Neuropsychiatric Disease, University of California San Diego, La Jolla, CA.

Over the last 25 years, quantitative dysmorphology methods have found increasing application in clinical and medical genetics. This has been facilitated by the publication of Farkas' atlas Anthropometry of the Head and Face. This craniofacial surface measurement system has become a popular standard reference, providing the most extensive normative database extant for direct craniofacial measurement in North America (N=1312). These anthropometric measurements have been used to render the diagnosis of anomalies both objective and reliable. Until recently geneticists have had to rely on cumbersome hard-copy look-up tables for these norms. Through the support of NIH, we have made the use of these norms more user-friendly by creating a computer-based interactive craniofacial measurement database for them. This software, FaceValue, provides tables of normative measurements conditioned on patient demographics. With these tables, the clinical/ medical geneticist configures operational definitions of anomalies, computes their z-scores as continuously distributed variables, and generates reports based on output measures. Here, we have applied FaceValue to the assessment of dysmorphic features in children with rare de novo Copy Number Variants (CNVs), using an anthropometric protocol to document for the first time clusters of specific, quantitatively-defined anomalies for molecular cytogenetic duplications and deletions. We provide worked examples of craniofacial anomalies among specific CNVs, including cases of 1q21.1 duplication, 15q11.2 deletion, and 22q11.2 deletion and duplication. These cases have also been phenotyped with respect to a spectrum of neuropsychiatric diagnosis, including but not confined to autistic disorders. Further, we present worked examples of dysmorphology observed in cases of teratogenesis in the form of prenatal exposure to anticonvulsants. In addition to distinctive combinations of dysmorphic features among these patients, convergent phenotypes emerged that are shared among multiple cases with autistic disorders. This convergence may signal common underlying forms of maldevelopment contributed by heterogeneous etiologies. Grant support: Simons Foundation SFARI (REACH Study; Sebat J, PI), Interactive Craniofacial Normative Database (R42DE016442; Deutsch CK, PI), Simons Foundation SFARI (Dysmorphology in Autism; Deutsch CK, PI).

1313W

Late-onset Alzheimer disease neuropathology genomic screen identifies novel loci for neuritic plaque and other AD neuropathology features.
K.L. Hamilton-Nelson[1], G.W. Beecham[1], A.C. Naj[3], L.-S. Wang[3], E.R. Martin[1], R. Mayeux[4], J.L. Haines[5], L.A. Farrer[6], G.D. Schellenberg[3], M.A. Pericak-Vance[1], T.J. Montine[2], Alzheimer's Disease Genetics Consortium. 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Fl; 2) Department of Pathology, University of Washington, Seattle, WA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Taub Institute of Research on Alzheimer Disease, Columbia University, New York, NY; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 6) School of Medicine, Boston University, Boston, MA.

Background: Late-onset Alzheimer disease (LOAD) is a highly heritable neurological disease with several known genetic risk loci (APOE, CR1, etc). Most of these loci have small effects on risk (e.g., odds ratios ≈1.1-1.2) and, aside from statistical association, no clear connection to LOAD etiology. To further investigate underlying genetic mechanisms of LOAD, we have performed a genome-wide association study (GWAS) of AD neuropathology and related phenotypes, including a neuropathology-confirmed case-control analysis, and analyses of neuropathology features, including neuritic plaques (NP), lewy bodies (LB), amyloid angiopathy (AA), medial temporal sclerosis (MTS), AD Braak stage, and vascular brain injury (VBI). We used this expanded neuropathology approach to limit the effects of phenotypic heterogeneity, and provide additional insights into AD subphenotypes. Methods: We examined 4,914 samples from 11 datasets in the Alzheimer Disease Genetics Consortium which were genotyped with high-density chips and imputed to a 1,000 Genomes Project reference panel. Primary pathology was assessed on ordinal scales (e.g., none, sparse, moderate, or frequent NP); statistical analyses were performed on either the ordinal or binary (i.e., presence or absence) scales as appropriate. Association was performed using logistic regression for binary traits and polytomous logistic regression for ordinal traits, followed by meta-analysis. Subjects examined included 3,887 neuropathologically-confirmed LOAD cases and 1,027 neuropathologically-confirmed cognitive controls. Results: Associations of APOE and BIN1 with LOAD were confirmed. Additionally, several novel LOAD associations were found, including PHF21B ($P=2.0\times10^{-8}$), and SMOX ($P=9.0\times10^{-7}$). Multiple loci were associated with the presence of neuritic plaques, including APOE ($P=1.8\times10^{-30}$), GALNT7 ($P=6.0\times10^{-9}$), ABCG1 ($P=8.0\times10^{-9}$), and a region near LMX1B ($P=4.3\times10^{-8}$). Additional loci were found to be associated with several ordinal neuropathology traits including LB, AA, MTS, VBI, and AD Braak staging (data to be presented). Conclusion: These results confirm several known AD risk loci and implicate novel loci in the etiology of LOAD neuropathology features, neuritic plaque in particular. Additional analyses including gene-based and pathway-based tests are being performed and will be reported, including thorough investigations of the known LOAD candidate genes.

1314T

Canine Multiple System Degeneration is Associated with Distinct SERAC1 Mutations in Two Different Dog Breeds. *J. Guo[1], R. Zeng[1], G.S. Johnson[1], T. Mhlanga-Mutangadura[1], E. Morava[2], T. Kozicz[2], J.F. Taylor[3], D.P. O'Brien[4], R.D. Schnabel[3].* 1) Department of Veterinary Pathobiology, University of Missouri College of Veterinary Medicine, Columbia, MO; 2) Hayward Genetics Center, Tulane University, LA; 3) Division of Animal Sciences, University of Missouri College of Agriculture, Food and Natural Resources, Columbia, MO, USA; 4) Department of Veterinary Medicine and Surgery, University of Missouri College of Veterinary Medicine, Columbia, MO, USA.

Canine Multiple System Degeneration (CMSD) is an autosomal recessive neurodegenerative disease. The initial cerebellar ataxia appears at 3 to 6 months and progresses to severe parkinsonism requiring euthanasia during the second year of life. Marked degeneration of the cerebellum, caudate nucleus, putamen, and substantia nigra are apparent at necropsy. CMSD has been recognized in two breeds: Kerry Blue Terrier (KBT) and Chinese Crested (CC). Previously, we mated CMSD obligate carriers from the two breeds and produced two normal and two CMSD-affected siblings, indicating that in both breeds CMSD is caused by the same mutation or by allelic mutations. Linkage mapping with samples from both breeds restricted the CMSD locus to a 5.2 Mb segment of CFA1 which contained 83 genes. We used Illumina sequencing technology with paired-end libraries to generate a whole genome sequence (WGS) with a 24-fold average coverage for a CMSD-affected KBT and a WGS with a 21-fold average coverage for a CMSD-affected CC. The reads from each WGS were separately aligned to the canine reference genome sequence (build 3.1) with NextGENe software (SoftGenetics). The detected sequence variants were prioritized according to predicted functionality and were filtered to remove variants also present in 54 other WGSs generated for dogs not affected with CMSD. After filtering, two homozygous sequence variants from the KBT alignment and four homozygous sequence variants from the CC alignment were in genes located within the previously mapped CFA1 region. Among these were a SERAC1 nonsense mutation in exon 15 in the KBT and a 4 bp deletion in a SERAC1 splice donor site sequence in the CC. RT-PCR amplification of RNA from a CMSD-affected CC showed that the SERAC1 splice donor site sequence deletion causes exon 4 skipping and a frame shift. All 6 CMSD-affected KBTs and all 20 CMSD-affected CCs were homozygous for their respective mutations, whereas, none of the 287 genotyped CMSD-free dogs were homozygous for these alleles. Mutations in human SERAC1 have been associated with a rare progressive neurodegenerative disease known as MEGDEL syndrome. MEGDEL syndrome patients exhibit dystonia, deafness, Leigh-like syndrome, impaired oxidative phosphorylation, and 3-methylglutaconic aciduria. The pathologic pathways underlying the neurodegeneration in this disease are poorly understood. Dogs with nullifying mutations in canine SERAC1 may prove useful as models to investigate these pathways.

1315F

Does gene function predict cognition and behaviour in X-linked intellectual disability? *K. Baker[1], G. Scerif[2], D.E. Astle[3], P.C. Fletcher[4], F.L. Raymond[1].* 1) Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom; 2) Department of Experimental Psychology, University of Oxford, United Kingdom; 3) MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom; 4) Behavioural and Clinical Neurosciences Institute, University of Cambridge, United Kingdom.

Genetic diagnosis is rapidly becoming a reality for many more individuals and families affected by Intellectual Disability (ID). Post-diagnostic challenges include establishment of a prognostic evidence base for ultra-rare disorders and elucidation of neurodevelopmental mechanisms. To this end, we have collected standardised medical, behavioural and cognitive data on participants with ID in the GOLD (Genetics of Learning Disability) cohort, who carry pathogenic variants in X-linked genes implicated in diverse neuronal functions including receptor dynamics, vesicle trafficking, ubiquitination and palmitoylation. Our aim is to determine which aspects of phenotype may be predicted by gene function. Severity of global impairment (assessed via Vineland Adaptive Behaviour Scales) varies widely within groups. However, beyond global impairment, function-associated characteristics start to emerge. For example, mutations in membrane-associated guanylate kinase genes (e.g. *DLG3*, *PAK3*) are associated with poor cognitive task performance (assessed via WASI-II), and high rates of behavioural problems particularly affecting attention and anxiety (assessed via the Developmental Behavior Checklist). Mutations in pre-synaptic and post-synaptic genes are associated with deficits in accuracy and speed-of-responding on computerised assessments of visual and auditory attention. Lastly, we have identified a previously-unreported association between *ZDHHC9* and childhood-onset nocturnal seizures, highlighting the importance of post-diagnostic medical phenotyping. The specificity of these observations will be tested via investigation of larger post-diagnostic samples and additional functional groups.

1316W

Serotonin Transporter Methylation and Response to Cognitive Behaviour Therapy in Child Anxiety Disorders. *S. Roberts[1], K.J. Lester[1], C.C.Y Wong[1], J.L. Hudson[2], C. Creswell[3], J. Mill[1,4], T.C. Eley[1].* 1) King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, UK; 2) Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, NSW, Australia; 3) Winnicott Research Unit, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK; 4) University of Exeter Medical School, Exeter University, St Luke's Campus, Exeter, UK.

Anxiety disorders are the most common psychiatric disorders in childhood and are associated with a range of social and educational impairments. Cognitive Behaviour Therapy (CBT) is effective for the majority of childhood anxiety cases, although up to 40% of children retain significant impairments after CBT. Recent research in the newly emerging field of 'therapygenetics' indicates that individual differences in treatment response may have a genetic basis. Of particular interest is the Serotonin Transporter gene (SERT), which has been associated with disorders such as anxiety and depression. Furthermore, the short (S) allele of the serotonin promoter polymorphism (5HTTLPR) has been associated with poorer outcomes in high-stress environments but better outcomes in positive environments. A study from our group demonstrated that anxious children with the SS genotype showed an increased response to CBT, suggesting an interaction between genes and environmental influences. Epigenetic mechanisms that influence transcriptional regulation (such as DNA methylation) have been shown to be susceptible to the environment, making them plausible candidates for the biological embedding of experience. We tested whether DNA methylation of CpG sites upstream of the SERT promoter were associated with response to CBT in a subsample of 168 clinically anxious children. Buccal swabs DNA was collected at pre- and post-treatment, and SERT methylation determined using the Sequenom EpiTyper. Treatment response was defined as the presence or absence of all anxiety disorder diagnoses at follow-up. We detected a significant 3-way interaction between clinical response, percentage methylation at pre- and post-treatment and CpG site ($F(5,77)= 2.39$, $p = 0.046$). A significant difference between treatment responders and non-responders was found in percentage methylation change (averaged across all sites) across treatment time ($t(115)= 3.20$, $p = 0.002$), with this effect being particularly strong for CpG site 4 in particular ($t(92)= 3.43$, $p = 0.001$). At this site, non responders displayed a decrease in methylation from pre to post treatment time points at CpG site 4 (mean change = -6.3%), while treatment responders showed an increase in methylation (mean change = 3.5%). These findings imply that responders and non-responders show differences in DNA methylation change across treatment time, most prominently at a CpG site upstream of the serotonin transporter gene.

1317T

Using transcriptomic data to understand the processes involved in antidepressant treatment response. *K. Hodgson[1], K.E. Tansey[1], G. Coppola[2], G. Breen[1], R. Uher[1,3], P. McGuffin[1].* 1) SGDP, Institute of Psychiatry, Kings College London, London, United Kingdom; 2) Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, USA; 3) Department of Psychiatry, Dalhousie University, Halifax, NS, Canada.

Research indicates that antidepressant response is a complex trait, involving interactions between both genetic and environmental factors (Tansey et al 2012). However the mechanisms by which antidepressants have their effects is unclear, and clinicians remain unable to identify who will respond best to which drug. Gene expression levels may capture the molecular changes that occur with antidepressant treatment, and can be influenced by both genetic and environmental factors. By examining the transcriptome and its alteration during treatment, this study aims to understand the biological mechanisms of antidepressant action, and identify if any of these processes are linked to individual variability in treatment response. Blood samples were taken from patients recruited into the Genome-Based Therapeutic Drugs for Depression (GENDEP) study both prior to treatment (week 0) and after receiving an antidepressant for eight weeks. RNA was extracted from whole blood, and processed on the Illumina HumanHT-12 v4 Expression BeadChip. After quality control measures, 123 patients had transcriptomic data available from both time points. Gene probes which varied significantly in expression levels between week 0 and week 8 were identified, and gene probes with expression levels significantly associated with antidepressant response (as measured by percentage change in the Montgomery-Asberg Depression Rating Scale) were identified. Network-level analyses of gene expression changes were undertaken using weighted-gene coexpression network analysis (WGCNA) to examine networks of gene coexpression, for analysis of alterations in gene-connectivity during treatment and links to treatment response. No significant associations were observed when performing the gene probe level analysis, using a threshold of FDR $q<0.1$. Network level analysis undertaken using WGCNA identified modules of coexpressed genes shared between week 0 and week 8 samples. However, the changes in the gene probe levels within each module (module eigengenes) from week 0 and week 8 did not differ significantly over the period of drug treatment. Whilst previous research looking into gene expression changes associated with antidepressant treatment has focused on candidate genes. However, our results indicate that any gene expression changes occurring as a result of antidepressant treatment are not of a large enough magnitude to be detected at transcriptome-wide significance at either the probe or network level.

1318F

MicroRNA-137 is Associated with Epigenetic Variation at the HCG9 Gene in the Amygdala. *A.P.S. Ori[1], K.R. van Eijk[2], M.F. Stokman[3], E. Strengman[2], The. Netherlands Brain Bank[4], M.P.M. Boks[5], S. de Jong[1], R.A. Ophoff[1, 6].* 1) Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, CA; 2) Department of Medical Genetics, University Medical Centre, Utrecht, The Netherlands; 3) Department of Clinical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 4) The Netherlands Institute of Neuroscience, Amsterdam, The Netherlands; 5) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, Utrecht, The Netherlands; 6) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

Noncoding RNAs are widely abundant in cells and fulfill critical roles as transcriptional and post-transcriptional regulators. Noncoding RNAs have been implicated in the regulation of the epigenetic mechanisms. MicroRNA-137 (miR-137) is a small non-coding RNA important for neurodevelopment. A large genome-wide association study recently suggested miR-137 mediated dysregulation as an etiologic mechanism in schizophrenia. This study aims to investigate the role of miR-137 function in epigenetic variation across regions of the human brain.

We investigated 249 postmortem brain samples originating from 44 non-demented control, 6 schizophrenia and 11 bipolar disorder individuals. Data was collected on whole genome DNA methylation and microRNA-137 expression. Linear models were used to investigate the relationship between microRNA expression and DNA methylation levels.

MicroRNA-137 expression showed to be regional specific with highest expression in cortical regions, limbic system and basal ganglia and lowest expression in the cerebellum. In addition, miR-137 expression showed a strong correlation with methylation levels at a CpG site of the HLA complex group 9 gene (HCG9) in the amygdala. This correlation was significant ($b = 6.98$, $t = 9.52$, $p=0.0066$) and showed to be independent of disease status.

This study demonstrates a relationship between the schizophrenia associated microRNA miR-137 and DNA methylation at HCG9 in the amygdala. This is a brain region involved in processing emotions and memories and reproducibly implicated in schizophrenia pathology. Interestingly, the HCG9 gene is located within the major histocompatibility complex (MHC) class 1 region, which is also associated to schizophrenia. Future work involves investigating the effect of the miR-137 schizophrenia risk variant and miR-137 target genes on miR-137 expression levels and epigenetic variation at the HCG9 gene. Taken together, this study reveals microRNA-137 expression across the human brain and provides key insights on cross talk between this small RNA molecule and DNA methylation. These findings contribute to understanding etiologic mechanisms in schizophrenia pathology.

1319W

The genetic architecture of white matter integrity: Insights from quantitative genetics and genome-wide associations. *E. Sprooten[1], E.E.M. Knowles[1], D.R. McKay[1], M.A. Carless[2], M.A.A. de Almeida[2], A. Winkler[3], T. Dyer[2], J.E. Curran[2], H. Göring[2], R. Olvera[2], P. Kochunov[4], P. Fox[5], L. Almasy[2], R. Duggirala[2], J. Kent[2], J. Blangero[2], D.C. Glahn[1].* 1) Department of Psychiatry, Yale University, New Haven, CT; 2) Department of Genetics, Texas Biomedical Research Institute, University of Texas Health Science; 3) FMRIB, Oxford University; 4) Maryland Psychiatric Research Center, University of Maryland; 5) Research Imaging Institute, University of Texas Health Science Center San Antonio.

Background: White matter integrity as measured by diffusion tensor imaging (DTI) is heritable and implicated in several complex phenotypes including major depressive disorder, bipolar disorder, schizophrenia and dementia. We investigated the genetic architecture of white matter and its homogeneity across regions using quantitative genetics and genome-wide association (GWAS) of global and regional fractional anisotropy (FA).

Methods: Participants included 776 Mexican-American individuals from extended pedigrees. DTI data were acquired on a Siemens 3T Trio scanner (b=0 and b=700 s/mm2) along 55 non-collinear directions and a spatial resolution of 1.7x1.7x3mm. TBSS was applied to create white matter skeletons representing the centers of white matter within each subject. Heritability estimates of mean FA within this skeleton and within 48 regions of interest, and genetic correlations between them, were calculated using SOLAR. Genome-wide association was conducted utilizing Illumina microarrays under an additive genetic model. Covariates included age, sex, their interactions, and the first 4 principal components to account for population stratification.

Results: FA was significantly heritable (h2=0.52, p=1.09*10−10). Genetic correlations between global FA and each of the tracts were between .28 and .94. A number of genome-wide significant SNPs were found for global FA. The strongest association was on 17q24.1 with the intergenic SNP rs10853057 (p=3.18*10−10) close to the gene GNA13, and the second strongest SNP rs12249377 (p=3.26*10-10) was in the serotonin receptor HTR7 gene on 10q23.31. An additional three significant hits were on chromosomes 12 (p=7.97*10−10), 16 (p=1.94*10−9) and 1 (p=2.00*10−8). The betas of the GWAS of the global FA measure explained on average 44% (15%-70%) of the rankings of the betas of sub-regions of interest.

Conclusions: FA is a heritable trait and a powerful quantitative phenotype to examine in genome-wide association studies. Genetic correlations and comparisons of GWAS results suggest that a substantial portion of genetic influence on white matter structure is uniform across the brain. Our results encourage further research into the biological roles of these genome-wide significant SNPs in health and disease.

1320T

Association of KIBRA With Episodic and Working Memory: A Meta-Analysis. *A. Milnik[1,2], A. Heck[1,3,4], C. Vogler[1,3,4], H.-J. Heinze[2,5,6], D.J.-F. de Quervain[4,7], A. Papassotiropoulos[1,3,4].* 1) Division of Molecular Neuroscience, Department of Psychology, University of Basel, Basel, Switzerland; 2) Department of Neurology, University of Magdeburg, Magdeburg, Germany; 3) Life Sciences Training Facility, Department Biozentrum, University of Basel, Basel, Switzerland; 4) University Psychiatric Clinics, University of Basel, Basel, Switzerland; 5) Leibniz Institute for Neurobiology, Magdeburg, Germany; 6) Helmholtz Center for Neurodegenerative Diseases, Magdeburg, Germany; 7) Division of Cognitive Neuroscience, Department of Psychology, University of Basel, Basel, Switzerland.

WWC1 was first implicated in human cognition through a genome wide association study in 2006 that reported an association of the intronic single nucleotide polymorphism (SNP) rs17070145 with episodic memory performance. WWC1 encodes the protein KIBRA, which is almost ubiquitously expressed. Together with its binding partners, KIBRA is assumed to play a role in synaptic plasticity. T-allele carriers of SNP rs17070145 have been reported to outperform individuals that are homozygous for the C-allele in episodic memory tasks. Here we report two random effects meta-analyses testing the association of rs17070145 with episodic and working memory. All currently available population-based association studies that investigated effects of rs17070145 on episodic or working memory were included in the analyses. Where performance measures for multiple domain-specific tasks were available for a given study population, averaged effect size estimates were calculated. The performed meta-analyses relied on 17 samples that were tested for episodic memory performance (N = 8,909) and 9 samples that had performed working memory tasks (N = 4,696). We report a significant association of rs17070145 with both episodic (r = 0.068, P = 0.001) and working memory (r = 0.035, P = 0.018). In summary, our findings indicate that SNP rs17070145 located within KIBRA explains 0.5% of the variance for episodic memory tasks and 0.1% of the variance for working memory tasks in samples of primarily Caucasian background.

1321F

A Meta-analysis including 9,038 individuals confirms an interaction effect between depression and *FTO* genotype on BMI. *M. Rivera[1,2], T. Corre[3,4], C. Wolf[5], D. Czamara[5], S. Kloiber[5], B. Muller[5], M. Preisig[6], G. Breen[1], I. Craig[1], A. Farmer[1], C. Lewis[1], P. McGuffin[1].* 1) MRC SGDP Centre, Institute of Psychiatry, King's College London, London , London, United Kingdom; 2) CIBERSAM, University of Granada, Spain; 3) Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 5) Max-Planck-Instutite of Psychiatry, Munich, Germany; 6) Department of Psychiatry, CHUV, Lausanne, Switzerland.

Background Depression and obesity are leading causes of disease burden and disability and major public health concerns worldwide. Both conditions are highly prevalent and are major risk factors for chronic (physical) diseases. The association between depression and obesity has repeatedly been reported in many studies. The role of the *FTO* gene in BMI and obesity has been confirmed in many independent studies. Recently, we have reported for the first time that depression amplifies the effect of *FTO* gene variation on BMI. The aim of the present study is to replicate these findings by investigating the *FTO* rs9939609 polymorphism in a meta-analysis including four independent studies consisting of depression cases and controls. **Methods** The sample consists of 5,134 depression cases and 3,904 controls from different studies: Radiant (3,251), PsyCoLaus (2,994), GSK (1,677) and MARS (1,116). As common inclusion criteria were available information on depression phenotypes, BMI and genotype data for rs9939609 *FTO* polymorphism. The distribution of BMI was positively skewed in all studies. We therefore transformed the data to $Log_{10(}$BMI) to achieve a closer approximation to normal distribution. In each individual study, linear regression models for quantitative traits assuming an additive genetic model were performed to test for the interaction between rs9939609 polymorphism and depression for an effect on Log_{10}BMI. All individuals were of white European ancestry. A classical approach meta-analysis with effect size estimates and standard errors was performed using the statistical package METAL. **Results** The results supported a significant interaction between *FTO* rs9939609 genotype and depression in relationship to Log_{10}-BMI (β=0.092, SE=0.033, p=0.0058). There was significant heterogeneity among studies. The sources of heterogeneity are being investigated. The meta-analysis interaction results show that in cases with depression there is an increased of 0.092 units of BMI for each *FTO* rs9939609 risk allele. **Discussion** This is to date the first meta-analysis investigating the relationship between *FTO*, BMI and depression. The results confirm that a history of depression increases the effect of *FTO* gene on BMI. This finding could have implications for predicting which patients with depression are at risk of high-BMI related disorders and potentially highlights how to improve prevention, management and treatment programs.

1322W

The genetic basis of solitude: lower heterozygosity in people with strict definition autism spectrum disorders and their parents. *S. Huang, X. Lu, J. Liang, D. Yuan.* State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China.

Autism Spectrum Disorders (ASDs) are >4 times more common in males and parents of people with ASDs often show mild forms of autistic-like characteristics or broad phenotypes. Strict (STR) and spectrum (SPC) definition ASDs differ mainly in social deficits with STR only slightly lower in IQ. Our recent work reveals a link between nucleotide diversity and complex traits/diseases consistent with stabilizing selection on SNP amounts and a priori truth on entropy and order. Heterozygosity (Het), a measure of random diversity, may be under stabilizing selection. From a priori reasoning, solitude should be linked with more homozygous (Hom) genomes since social deficits and tendency to marry people like oneself favor less admixture. Parents of ASDs are more populated in science and engineering fields demanding both IQ and creativity. While ASDs may be under negative selection, solitude is critical to creativity. An aesthetic hallmark of great creative works is unity in harmony of opposites in a pure and high contrast fashion. A genome ranked high in aesthetics with more Hom parts should make more creative people with all around balanced traits. Remarkably, previous work indeed shows less social but more novelty seeking behavior in inbred versus hybrid mice. We analyzed SNP dataset from dbGAP of 1154 ASD trios of European ancestry and correlated each individual's number of Het SNPs with STR vs SPC status. We found a moderate but significant correlation of low Het to STR cases (Spearman $r = 0.086$, $P < 0.01$) and their fathers ($r = 0.08$, $P < 0.01$). The average Het of STR cases is lower than SPC ($P < 0.05$, t test). STR sons have lower Hom minor allele content or amount (MAC) than their fathers ($P = 0.02$, t test), and the same holds for SPC sons and fathers ($P = 0.006$). (MAF was determined using parents.) Similar results were found for 148 trios of non-European ancestry. Finally, we analyzed relevant control cohorts from dbGAP including the Framingham study. Low Het associates significantly with living alone, never married, lower birth order, and higher education level and income. These results suggest that function and solitude in STR fathers may be at precarious near imbalance under more Hom DNAs. A moderate decrease in their sons in Hom MAC coupled with a slight increase in Het may tilt the balance towards extreme functions at either end without much improvement on solitude. The same may also apply to SPC involving more drop in Hom MAC.

1323T

Shared genetic vulnerability for Attention Deficit Hyperactivity Disorder, Substance Use and Gambling in Australian Adolescents. *P.A. Lind[1], D.A. Hay[2], T.J. Cicero[3], N.G. Martin[4], S.E. Medland[1].* 1) Quantitative Genetics, Queensland Institute of Medical Research, Brisbane, Australia; 2) School of Psychology, Curtin University, Perth, Australia; 3) School of Medicine, Department of Psychiatry, Washington University, USA; 4) Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia.

Previous studies have shown that the prevalence of substance use (including legal and illicit drugs) and gambling in children with Attention Deficit Hyperactivity Disorder (ADHD) is higher than in the general population. The aim of this study was to explore whether cigarette smoking, alcohol drinking and gambling in adolescents with ADHD was due to shared genetic risk factors. We analyzed self-report substance use and gambling data in twin participants from the Genetics of Inattention in Australia study (N = 500 individuals) who completed a structured psycho-social interview. Data included age of first use, number of DSM-IV symptoms for substance dependence or pathological gambling, the maximum number of cigarettes smoked in 24 hr (MAXCIG), the maximum number of alcoholic drinks consumed in 24 hr (MAXDR), and the number of different types of gambling used (PGVAR). Using quantitative genetic modelling, we found that genetic risk factors for ADHD also increased DSM-IV dependence symptoms (genetic correlations, r_g, .03-.19), the number of cigarettes consumed ($r_g = .45$) and types of gambling tried ($r_g = .07$), and reduced the age of onset for gambling, smoking and using alcohol (r_g ranges from −.07 to −.48). However, a negative genetic correlation was observed with alcohol consumption ($r_g = -.29$). These results suggest that the variants influencing ADHD behaviours influence substance use and gambling in later life and provide support for the development of an ADHD-substance use cross disorder consortium.

1324F

Genome-wide association study for Cognitive Decline. *LB. Chibnik[1, 2], L. Yu[3], T. Raj[1, 2], J. Xu[1], N. Patsopoulos[1, 2], BT. Keenan[1], R. Sherva[4], SE. Leurgans[3], D. Blacker[5], RS. Wilson[3], EM. Reiman[6], M. Huentelman[7, 8], RC. Green[9, 10], LA. Farrer[4], P. Crane[11], R. Mayeux[12, 13], R. Lipton[14], GD. Schellenberg[15], DA. Evans[16], PL. De Jager[1, 2], DA. Bennett[3], Alzheimer's Disease Genetics Consortium.* 1) Department of Neurology, Harvard Med Sch/Brigham & Women's Hosp, Boston, MA; 2) Program for Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL; 4) Boston University, School of Medicine, Boston, MA; 5) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 6) Neurogenomics Division, Translational Genetics Research Institute and Arizona Alzheimer's Consortium, Phoenix, AZ; 7) Banner Alzheimer's Institute and Department of Psychiatry, University of Arizona, Phoenix, AZ; 8) Department of Psychiatry, University of Arizona, Tucson, AZ; 9) Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA; 10) Partners Healthcare Center for Personalized Genetic Medicine, Boston, MA; 11) Department of Medicine, University of Washington, Seattle, WA; 12) Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, NY; 13) Gertrude H. Sergievsky Center, Columbia University, New York, NY; 14) Departments of Neurology, Epidemiology and Population Health and the Montefiore Headache Center, Albert Einstein College of Medicine, Bronx, NY; 15) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 16) Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, IL.

Background: Cognitive decline, especially decline in episodic memory, is the clinical hallmark of Alzheimer's disease (AD) and known to start long before the onset of clinically diagnosed AD dementia. Thus, it may serve as a useful quantitative phenotype for GWAS and other 'omics' analyses. Methods: We included non-Hispanic White participants from four prospective community-based cohort studies (ROS, MAP, CHAP, EAS) who were non-demented at study entry and have at least two repeated measures of cognition. Within each cohort, individual cognitive tests were combined to form aggregate measures of global cognition and episodic memory. Genotype data from each cohort was quality controlled and imputed using 1000Genomes reference panel. We used linear mixed effects models to characterize individual paths of change in cognition, modeling age, sex and education as fixed effects and intercept and slope as random effects. We extracted individual residual cognitive slope estimates from the models and fit a linear regression model for each single nucleotide polymorphism (SNP), adjusting for population stratification. Results were meta-analyzed across cohorts. Results: A total of 3534 subjects were analyzed. The average age at enrollment in each study ranged from 72 to 81, follow-up ranged from 2–18 years and incidence of clinical diagnosis of AD ranged from 19% and 34%. The most significant loci for decline in both global cognition and episodic memory was APOE ($p=3.1\times10^{-34}$ and $p=2.1\times10^{-24}$, respectively). Although no other non-APOE locus reached genome-wide significance, we found suggestive results at two other regions. First, near the EDAR gene, previously found suggestive in a large AD case/control GWAS, with the top locus in the intergenic region 23kb upstream of EDAR reaching a $p=8.9\times10^{-8}$ for global cognition and $p=1.0\times10^{-4}$ for episodic memory and second, on the RAB3GAP2 gene with the top locus (non-coding intron) reaching a $p=1.3\times10^{-7}$ for global cognition and $p=1.3\times10^{-6}$ for episodic memory. We are currently analyzing data from two other cohort studies for validation. Conclusion: One previously un-described SNP and one possible AD associated SNP are observed to be associated with cognitive decline in a meta-analysis of community-based elders followed longitudinally. Our findings to date suggest that cognitive decline may be a useful phenotype for identifying genomic variation associated with AD or other factors influencing cognitive decline.

1325W

Combining RNA-sequencing and genotyping data to identify genes and pathways associated with major depression. *S. Mostafavi[1], A. Battle[1], X. Zhu[1], J. Potash[2], M. Weissman[3], J. Shin[4], K. Beckman[5], C. Haudenschild[6], C. McCormick[7], R. Mei[8], M. Gameroff[2], H. Glindes[3], P. Adams[2], F. Goes[9], F. Mondimore[9], D. MacKinnon[9], L. Notes[10], B. Schweizer[9], D. Furman[1], S. Montgomery[1], A. Urban[1], D. Koller[1], D. Levinson[1].* 1) Stanford University, Stanford; 2) University of Iowa, Iowa City; 3) Columbia University, NY; 4) National Cancer Institute, Bethesda; 5) University of Minnesota, Minneapolis; 6) Personalis, Menlo Park; 7) Ilumina, La Jolla; 8) Centrillion Biosciences, Palo Alto; 9) Johns Hopkins University, Baltimore; 10) American University, Washington DC.

Altered gene expression levels in disease can reflect the effect of sequence variation, environmental factors and their interaction with genetics, as well as the effects of the disease processes itself. To gain insights into biological mechanisms that are relevant to major depressive disorder (MDD) by identifying genes and pathways with altered expression levels, we compiled genome-wide gene expression, genotype, and physiological data in a case/control cohort of MDD from a large population-based sample. We analyze 463 recurrent MDD cases and 459 controls, that were recruited from a nationally-representative survey research panel. We sequenced RNA from whole-blood, genotyped these individuals for common SNPs, and derived information on psychiatric and medical history. Using these data, we assessed the association of MDD status with each gene and with sets of genes (pathways), while accounting for the influence on expression levels of potentially confounding variables such as environmental factors, and computational estimates of cell type proportions. Our analysis consists of three components: (1) association analysis of single gene expression levels with MDD; (2) analysis of association of MDD with canonical gene pathways, and hypergeometric tests of the enrichment of pathways in subsets of the most strongly-associated genes; and (3) joint analysis of association of gene expression and eQTLs (SNPs associated with expression, or expressed quantitative trait loci) with MDD. In the analysis of association between MDD and expression levels of autosomal genes, a significant excess of low p-values was observed, but no single-gene association was significant after genome-wide correction. Pathway-based analyses of expression data detected a significant association (FDR<0.05) between MDD and the interferon-I signaling pathway, with increased expression levels in MDD cases, where secondary analysis did not identify this result subject to additional confounding factors. Finally, joint evaluation of gene expression and eQTL genotypes identified a significant association with CINP, a gene involved in cell cycle arrest. The results support the hypothesis that altered immune signaling plays a role in the pathogenesis and/or the persistence and progression of MDD, and in particular implicates type I interferon signaling in pathology of MDD.

1326T

The PD Brain Map Project: Mapping the Transcriptional Architecture of Dopamine Neurons in Human Brain. *C.R. Scherzer[1,2,6], C. Vanderburg[2], T.G. Beach[3], C.H. Adler[4], J.J. Locascio[1,2], N. Pochet[5,6], A. Regev[5], B. Zheng[1], Z. Liao[1].* 1) The Neurogenomics Laboratory, Harvard Medical School and Brigham & Women's Hospital, Cambridge, MA; 2) Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Sun City, AZ; 4) Department of Neurology, Arizona Mayo Clinic, Scottsdale, AZ; 5) The Broad Institute of MIT and Harvard, Cambridge, MA; 6) Department of Neurology, Brigham & Women's Hospital, Boston, MA.

While the number of human genes has shrunk to an estimated ~22,000, a hidden universe of an ever-increasing number of tens of thousands of non-coding RNAs and splice variants is revolutionizing our thinking about the complexity of the human brain. We hypothesize that in complex genetic neurodegenerative diseases such as Parkinson's (PD) combinatorial effects of genetic and environmental risks disrupt the ordered flow of genetic information into vulnerable brain cells through cell-type-specific modulation of transcript abundance, transcription start site use, and sequence. The PD Brain Map Project's goal is to chart the flow of information from the entire human genome into a prototype brain cell-type - the dopamine neuron. Genetic variation between more than 100 individuals is examined for correlation with differences in transcribed elements — both protein-coding and non-coding — to identify regions of the genome that influence whether, how, and how much a transcript is expressed in this specific cell type *in situ* in human brains. Transcriptomes of control and diseased brains are probed using laser-capture microdissection, massively parallel sequencing of near ultra-low amounts of RNA, and expression Quantitative Trait Locus analysis. The PD Brain Map will provide a high-resolution encyclopedia of transcribed elements in a prototype cell-type in human brains, help to understand inherited susceptibility to PD, and highlight targets for precision therapies. Pilot phase results indicate that tens of thousands of isoforms of protein-coding genes and of non-coding RNAs are abundantly expressed in dopamine neurons and suggest a more diverse and complex transcriptional architecture than previously imagined. Support: U01 NS082157; W81XWH-BAA-12-1; Michael J. Fox Foundation; U24 NS072026; P30 AG19610.

1327F

Adult onset painful sensory polyneuropathy caused by a dominant NAGLU mutation. *M. Tetreault[1,2], MJ. Dicaire[1], P. Allard[3], K. Gehring[4], D. Leblanc[3], N. Leclerc[5], J. Mathieu[5], B. Brais[1].* 1) Neurogenetics of motion laboratory, Neurology and Neurosurgery, Montreal Neurological Institute, Montreal, PQ, Canada; 2) Centre of Excellence in Neuroscience of Université de Montreal, CHUM Research Center, Montréal, Québec, Canada; 3) Laboratoire de génétique médicale, CHU-Ste-Justine, Montréal, Québec, Canada; 4) Department of Biochemistry, McGill University, Montréal, Québec, Canada; 5) Cliniques de maladies neuromusculaires, Jonquière, Québec, Canada.

Late-onset painful sensory neuropathies are usually considered to be acquired and most commonly associated with common diseases such as diabetes. Adult presentations of known hereditary forms such as Amyloidosis and Fabry's are accompanied by other organ involvement. We recruited a large French-Canadian family (45 individuals; 21 affected cases) with a dominantly inherited late onset painful sensory neuropathy to uncover the underlying genetic mutation. The main clinical features of this dominant painful sensory polyneuropathy are the appearance of excessive cramps as early as 20 years old, constant painful paresthesias in the feet and later the hands appearing on average around age 55 (30–65) that interfere with sleep and progresses into a mild sensory ataxia. Electrophysiological studies are normal until late in the course of the disease where it documents a sensory polyneuropathy. Four affected individuals were sent for exome sequencing. Analysis of rare variants shared by all affected cases led to a list of four candidate variants. Segregation analysis in all recruited individuals has shown that only the p.Ile403Thr variant in the α-N-acetyl-glucosaminidase (NAGLU) gene is segregating with the disease. Recessive NAGLU mutations cause the severe childhood lysosomal disease Mucopolysaccharidosis IIIB (MPS-IIIB). Family members carrying the mutation showed a significant decrease of the enzymatic function. The late onset and variable severity of the symptoms may have precluded the description of such symptoms in parents MPS IIIB cases since they usually are lost to follow-up. The identification of a dominant phenotype associated with a NAGLU mutation supports that some carriers of lysosomal enzyme mutations may develop later in life much milder phenotypes.

1328W

Exome Sequencing of Extended Families Identify Putative Novel Candidate Genes for Autism. *A. Patowary, I. Stanaway, R. Nesbitt, W. Raskind, D. Nickerson, Z. Brkanac.* University of Washington, Seattle, WA.

Autism is a complex neurodevelopmental disorder behaviorally characterized by impairments in social interactions, communication and repetitive and restricted patterns of behavior. Autism has a strong genetic component. Currently more than 100 genetic causes of autism are recognized accounting for approximately 20% of cases and indicating high levels of genetic heterogeneity. Recently additional genes with de-novo mutations with large effects were identified in simplex families, with each gene accounting for a very small fraction of cases. In similar manner to genetics of simplex autism it is likely that for a subset of familial cases rare highly penetrant genetic variants are causal as well. Identification of genes responsible for familial autism is important as genetic architecture of familial autism might have important epidemiological implications. Under assumption that causal mutations are private or very rare, exome sequencing is a powerful tool for identification of highly penetrant variants in families with multiple affected cases. We performed whole exome sequencing in 27 NIMH and University of Washington autism families. We have selected families that include affected cousins and frequently additional affected siblings to minimize variant sharing between affected family members. The exome sequencing using NimbleGen SeqCap v2.0 and Illumina HiSeq2000 was performed at University of Washington Northwest Genomic Center. Sequences were aligned to the reference genome with BWA and variants were called with GATK. Functionally annotations of the variants were performed using ANNOVAR. As candidate variants we consider variants that are private, (not present in dbSNP, 1000 genomes or ESP6500) and functional (coding, missense, stop-gain, stop-loss, splice). The annotation and frequency filtering resulted in identification of 22±12 shared private functional variants in each family. In aggregate our sample has generated 198 private candidate variants. We have compared our candidate gene list with list of 424 genes with single de-novo mutations from recent exome sequencing studies. We have found 7 genes that are present in de-novo cases and in our families as well. As a next step, to confirm association with autism for the genes with mutations in simplex and familial cases, we are using molecular inversion probes and next generation sequencing to perform a large gene based case-control study of 1500 familial autism cases and 500 unscreened controls.

1329T

Identification of Rare Variants in Childhood Onset Schizophrenia using Exome Sequencing. *A. Ambalavanan[1], S.L. Girard[2], J. Gauthier[2], L. Xiong[2], A.D. Laporte[2], D. Spiegelman[2], E. Henrione[2], O. Diallo[2], S. Dobrzeniecka[2], P.A Dion[2], R. Joober[1,3], J.L. Rapoport[4], G.A. Rouleau[1,5].* 1) Department of Human Genetics, McGill University, Canada; 2) Centre de Recherche du Centre Hospitalier de Université de Montréal (CRCHUM), and Department of Medicine, Université de Montréal, Montreal, Quebec, Canada; 3) Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; 4) Child Psychiatry Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, USA; 5) Montreal Neurological Institute, McGill University, Canada.

Background: Childhood Onset Schizophrenia (COS) is a rare severe form of schizophrenia for which definitive genetic causes remain elusive even though the disease itself is heritable. The linkage and association studies of COS have been ineffective in attempting to identify explicit risk variants. We now know that de novo mutations play a significant role in the genetic mechanism of Schizophrenia, thus, we hypothesized that rare de novo mutation in different genes account for a portion of COS individuals. **Methods:** Our approach is to identify causative genes by targeted capture of the exome of COS proband and their respective parents followed by massively parallel sequencing of each individual in the family. We prioritized coding variants that are found only in the affected children within the family as well as variants found to be de novo. Each candidate variant is validated by Sanger sequencing in the proband and both parents. **Results:** We found interesting variants, including many de novo mutations and these potential variants are being validated with Sanger sequencing and further sequencing is being conducted to explore the genes and their major biological pathways **Conclusion**: We believe that insights from the identification of COS genes may help to device diagnostic tools which in turn may help to intervene earlier before the onset of symptoms which ultimately lead to the identification of therapeutic avenues for better treatment.

1330F

Family Genomics Reveals a Network of Calcium Signaling Genes Underlying Bipolar Disorder. *S.A. Ament[1], G. Glusman[1], H.C. Cox[1], D.E. Mauldin[1], S.Z. Montsaroff[1], N.D. Price[1], S. Szelinger[2], D.W. Craig[2], H.J. Edenberg[3], L. Hood[1], F.J. McMahon[4], J.R. Kelsoe[5], J.C. Roach[1], The Bipolar Genome Study.* 1) Institute for Systems Biology, Seattle, WA; 2) Translational Genomics Institute, Phoenix, AZ; 3) Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN; 4) National Institute of Mental Health, Bethesda, MD; 5) Department of Psychiatry, University of California, San Diego, CA.

We sequenced 200 personal genomes from 43 multiplex families with bipolar disorder to elucidate its genetic causes. Analysis of single pedigrees suggested oligogenic causation, with multiple uncommon, segregating coding and putative regulatory variants in each pedigree but little evidence for highly penetrant monogenic causes of disease. We developed new computational tools to aggregate candidate genomic variants across multiple pedigrees, and gene- and network-level tests for their statistical association with bipolar disorder. Comparison of the 43 sequenced bipolar disorder pedigrees to whole genome sequences from 158 individuals in 34 non-bipolar pedigrees revealed single genes and functionally-related gene networks that harbored segregating candidate variants frequently in bipolar pedigrees but rarely in control pedigrees. Voltage-gated calcium channels, calmodulin-dependent protein kinases, and GABA receptors were among the pathways most strongly associated with bipolar disorder in this analysis. We chose 30 top candidate genes for re-sequencing in 6000 additional bipolar disorder cases and controls. Our results suggest that a network of rare genomic variants in multiple genes with synaptic functions increase risk for bipolar disorder. In addition, our results suggest that -- despite multigenicity - family genomics is a powerful approach to discover genetic causes of common, complex diseases.

1331W

Partial deletion of the *RBFOX1* gene in a 16-year-old male patient with autism spectrum disorder, attention deficit hyperactivity disorder and intellectual disability. *O. Bartsch, D. Galetzka, E. Weis, S. Schweiger, J. Winter.* Institute of Human Genetics, Johannes Gutenberg University Mainz, Mainz, Germany.

The RBFOX1 protein is a muscle, heart and brain specific RNA-binding protein that regulates the alternative splicing of various genes in neuronal development and maintenance. Copy number variations within the *RBFOX1* gene (alias *FOX1*, *A2BP1*, OMIM *605104) are increasingly realized as causes of human neurodevelopmental disorders including intellectual disability, autism spectrum disorder (ASD), epilepsy, attention deficit hyperactivity disorder (ADHD), and schizophrenia. Here, we report on a 16-year-old male with ASD, mild intellectual disability (IQ 59, tested at age 13 years) and a heterozygous deletion spanning 100.3 kb in the 5' region of the *RBFOX1* gene. From age 2 years onwards, he showed speech delay and behavioural abnormalities including autoaggression, head banging and teeth grinding. He responded towards touch with beating, frequently cried himself to sleep and had fits of anxiety, e.g. towards noise. He also had frequent fits of laughter and tantrums without obvious cause. Neurological examinations at 5 and 6 years of age indicated mild muscular hypotonia, reduced gross and fine motor skills, clumsiness and ADHD (agitation, inability to concentrate, distractibility, impulse control disorder, and emotional lability). At age 16 years he read fluently with proper intonation and understanding (school grade 3 level) but could not write, perform simple mathematical calculations or hold eye contact. The deletion on chromosome 16p13.2 (chr16:6,751,824-6,852,175, UCSC hg19) included the promoter region and exon 1 of transcript variant 6 of *RBFOX1*. A previous report describes a patient with remarkably similar molecular and clinical findings (Mikhail et al., 2011, AJMG-A 155A:2386-2396). Our study further supports the role of *RBFOX1* gene mutations and in particular, of heterozygous loss-of-function mutations of *RBFOX1* transcript variant 6, in the etiology of ASD, ADHD and intellectual disability.

1332T

Genetics and Cellular Neurobiology of Severe Mental Illness. *C.G. Bouwkamp[1], F.M.S. De Vrij[1], N. Gunhanlar[1], S.T. Munshi[1], M.P.H. Coesmans[1], M. Quadri[2], G. Breedveld[2], J.A. Maat-Kievit[2], B. Oostra[2], V. Bonifati[2], S.A. Kushner[1].* 1) Dept of Neurobiological Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands; 2) Dept of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands.

The neurobiological mechanisms of severe psychiatric disorders such as schizophrenia, bipolar disorder, and autism remain elusive. The aim of our project is to uncover the molecular and cellular neurobiology of psychiatric disorders by identifying rare highly-penetrant causative Mendelian genetic variants in severely affected families. A profound problem in studying the neurobiology of psychiatric disorders has always been the wide phenomenological heterogeneity. In fact, there is emerging consensus that the diagnosis of psychiatric disorders reflect a multitude of different syndromes at the biological level. Our study attempts to circumvent the difficulties inherent in studying genetically unrelated patients by using exome and whole-genome sequencing in a family-based design, which has yielded some notable recent successes in the literature. Families are recruited throughout the Netherlands from psychiatric clinics, national patient organizations, and the Erasmus MC department of Clinical Genetics. In particular, we have included families in which there are four or more affected family members who share a similar form of disorder. Affected and unaffected family members undergo structural clinical assessments by clinically-trained staff of our research group and whole-blood DNA is obtained. Exome sequencing or whole-genome sequencing is performed, in combination with linkage and copy number variant analysis using Illumina OmniExpress BeadChips. Called variants are filtered based on linkage regions, inheritance model, coding region, Minor Allele Frequency (MAF) <2 percent; in public databases, and being nonsense, missense or splice-affecting. Remaining variants are Sanger-validated in the family DNA samples and candidate-variants are validated by large case-control cohort screening. Furthermore, we obtain skin biopsies from the family members that are enrolled in our study. Using human induced pluripotent stem cell technology (hiPSC), it is possible to reprogram fibroblasts to pluripotency and subsequently derive neurons under controlled conditions. We are developing standardized methodology for comparing electrophysiological, morphological and molecular biological signatures of affected versus unaffected family members to elucidate the underlying neurobiological mechanisms.

1333F

Mapping the schizophrenia brain: Genes harboring damaging de novo mutations in schizophrenia map to a highly interconnected network of transcriptional co-expression and protein interaction in fetal prefrontal cortex. *S. Gulsuner[1], T. Walsh[1], A.C. Watts[1], M.K. Lee[1], A.M. Thornton[1], S. Casadei[1], C. Rippey[1], H. Shahin[1], M.-C. King[1], J.M. McClellan[2], Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group,.* 1) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 2) Department of Psychiatry, University of Washington, Seattle WA.

Schizophrenia is a chronic neuropsychiatric disorder with a prevalence of 1% and requires lifelong treatment. Despite strong evidence of a familial component, genetic causes of the disease remain elusive. Many patients have no family history of mental illness. Genes responsible for schizophrenia can be revealed by de novo mutations in such patients. Participants in our project were 105 quads or trios comprising a proband with schizophrenia, an unaffected sibling if available and their unaffected parents. Families were selected for absence of history of serious mental illness. Exome sequencing of all 399 persons revealed 57 damaging de novo mutations in 47 of 105 schizophrenia probands and 35 such mutations in 25 of 84 unaffected siblings (X2 = 4.45; P = 0.035). The genes harboring de novo damaging mutations yielded networks based on protein-protein interactions (PPI) and transcriptional co-expression in different brain regions at different developmental stages, as revealed by RNASeq analysis from the BrainSpan Atlas of Developing Brain. The network generated by genes harboring de novo damaging mutations in schizophrenia had significantly more physical interactions (P=0.0005) and significantly greater levels of transcriptional co-expression in dorsolateral (P=0.0001) and ventrolateral (P=0.0004) regions of the prefrontal cortex during fetal development, compared to networks derived from genes mutant in unaffected siblings. Overall, 40 genes with de novo damaging mutations in schizophrenia mapped to this highly interconnected network. These genes function in the brain in neuronal migration, synaptic transmission, signaling, transcriptional regulation, and transport. These results support prefrontal cortex as a critical region in the pathogenesis of schizophrenia. The approach also supports the applicability of proteomic and transcriptome analyses in order to map critical genes to shared biological networks for conditions characterized by extreme genetic heterogeneity.

1334W

Targeted re-sequencing of HDC and SLITRK1 in Tourette syndrome using next-generation sequencing. *A. Inai[1], H. Kuwabara[1], Y. Eriguchi[1], T. Shimada[2], M. Furukawa[1], T. Sasaki[3], M. Tochigi[2], C. Kakiuchi[2], K. Kasai[2], Y. Kano[1].* 1) Depaerment of Child Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Japan; 2) Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Japan; 3) Department of Health Education, Graduate School of Education, University of Tokyo, Japan.

Introduction; Tourette syndrome (TS) is neurodevelopmental disorder characterized by both motor and vocal tics. Family studies and twin studies provide strong evidence for genetic nature of TS. Although heritability of TS is high, genetic factors remain largely unknown. Association studies revealed that common genetic variants explain a modest fraction of heritable risk for TS, suggesting the important role of rare variants for the unexplained heritability. Histidine decarboxylase gene (HDC) and Slit and Trk-like 1 gene (SLITRK1) were proposed as causative factors for TS, however, subsequent studies showed difficulty in confirming some of these results. The aim of the present study is to examine whether TS is related to the HDC and SLTRK1 in Japanese population. Methods; A total of 95 individuals with TS participated in this study. All the participants provided written informed consent after they were given a complete explanation of the study as required by the ethics committee of the University of Tokyo Hospital. The exon region of HDC and SLITRK1 were amplified on TruSeq Custom Amplicon kit following the manufacturer's protocols. Amplified libraries were sequenced on the Illumina Miseq as pair-end 151-bp reads. Sequence reads were mapped to the reference genome (hg19). Variant calls were filtered to coordinate with at least 10×coverage. Results; There were total of 370 single nucleotide variants (SNVs) called in the 95 TS participants. After filtering for synonymous substitutions and database SNPs (db131), four novel and non-synonymous variants remain. Three participants had distinct SNVs on HDC, and other three participants had identical SNV on SLITRK1. PolyPhen algorithm predicted that all three SNVs on HDC disrupted the conformation of the protein. On the other hand, the SNV on SLITRK1 was predicted to be benign mutation. Conclusions; Present study suggested that HDC gene could play a role in the pathogenesis of TS, whereas SLITRK1 might play few or no role in TS susceptibility in Japanese population.

1335T

Exome sequences of multiplex, multigenerational families reveal schizophrenia risk loci involved in fatty acid oxidation. *M.Z. Kos[1], J. Peralta[1], M.A. Carless[1], M. Almeida[1], R.C. Gur[2], M.F. Pogue-Geile[3], D. Roalf[2], V. Nimgaonkar[4], R.E. Gur[2], L. Almasy[1].* 1) Dept Gen, Texas Biomedical Research Institute, San Antonio, TX; 2) Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA; 3) Department of Psychology, University of Pittsburgh, Pittsburgh, PA; 4) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Schizophrenia is a serious mental illness defined by severe psychosis and disruptions in thought and behavior, with a worldwide prevalence of about 1 percent. Although schizophrenia is highly heritable, with estimates hovering around 0.80, much of the genetic liability is yet to be explained. In this paper, we search for susceptibility loci in multiplex, multigenerational families affected by schizophrenia, by targeting coding regions that are more likely to harbor protein-altering variants with potential risk effects. Exome sequencing was performed on 134 samples from eight European-American families, including 25 individuals with DSM-IV schizophrenia or schizoaffective disorder, using Illumina TruSeq technology. The exome variant calls were filtered based on stringent GATK probabilistic quality scores (LODs ≥ 4.0), as well as functional relevance as computed by the SIFT and PolyPhen algorithms. In total, 11,878 nonsynonymous variants with putative deleterious effects, representing 5,867 genes, were tested for their association with the schizophrenia spectrum disorders. Our top association hit is for a common polymorphism (rs1094111; MAF = 0.40) in the gene AMACR, which has been previously implicated in the risk of schizophrenia, a result that remained highly significant after correcting for multiple-testing (corrected P =0.0043). The enzyme coded by this gene, alpha-methylacyl-CoA racemase, is involved in oxidative metabolism of branched-chain fatty acids. Its deficiency can lead to the accumulation of pristanic acid in neural tissue, which may influence brain function and structure. In addition, the SNP rs10378 in the gene TMEM176A is also significant (corrected P = 0.031). Interestingly, permuted pathway analysis (1,000 runs) of the top association hits (P < 0.01; n = 249 genes) revealed significant enrichment of genes involved in 'fatty acid oxidation' (empirical P = 0.027), which includes the gene AMACR. Rare, nonsynonymous variants were found in two other genes involved in this metabolic pathway: a novel one in the gene ACAA1 (P = 0.0016; MAF = 0.008); and rs2894359 in the gene HSD17B4 (P = 0.0065; MAF = 0.025). In conclusion, our association findings from family-based exome sequence data suggest an important role for genes involved in fatty acid metabolism in the risk of schizophrenia, in particular the gene AMACR, providing key insights into possible strategies for preventative and therapeutic treatments.

1336F

Runs-of-homozygosity are associated with intellectual disability and female gender in simplex autism. *E. Morrow[1,2], E. Gamsiz[1], E. Viscidi[1], A. Frederick[1], M. Schmidt[1], E. Triche[1], S. Nagpal[1].* 1) Molecular Cellular Biology, Brown University, Providence, RI; 2) Simons Simplex Genetics Collaborative.

Background: To date there have been few genome-wide studies designed to examine excess homozygosity in autism, particularly in pedigrees without recent shared ancestry, which could indicate the effect of autosomal recessive loci. Methods: We have investigated a possible role for autosomal recessive loci through the study of autosomal runs-of-homozygosity (ROH) using SNP arrays. We have analyzed the Simons Simplex Collection comprised of 2108 North American simplex ASD families. We test the hypothesis that ROH is associated with affected status in those children with autism and co-occurring ID, or in affected females. We also analyze exome sequencing data for rare, homozygous deleterious variants within relevant ROH intervals for sequence mutations in candidate autism genes. Results: We discovered that affected status is associated with an increased burden of ROH in probands with IQ≤70. As minimum block size was increased, we observed an increased burden of ROH in affected probands with IQ≤70 as compared to unaffected matched siblings but no such excess was seen for IQ>70. For example, at a minimum segment size ≥ 2500kb, the total burden of ROH in probands with IQ≤70 was 1.326 times greater than unaffected siblings (p = 0.012). By contrast, the ratio of the ROH burden in affected versus unaffected siblings in the cohort with proband IQ>70 was 1.012 (p = 0.8935). Girls with autism also revealed a statistically significant burden of ROH as compared to their same sex siblings (p < 0.01), whereas boys with autism relative to their male unaffected siblings did not show this pattern. Within the autism cohort, we noticed a strong statistical association between ROH (minimum segment >2500kb) and measures of intellectual (IQ) and adaptive function (Vineland), but not measures of autism symptoms or severity. ROH segments are widely distributed across the entire genome. There are many regions of the genome wherein there is an excess of ROH in affected probands as compared to controls. We did identify several candidate, homozygous deleterious variants that require further study as candidate disease-associated variants. Discussion: Our study argues that autosomal recessive loci may contribute to an important part of the genetic architecture of autism in low IQ probands at least in simplex autism. In addition, we present among the first data that support a model involving a distinct genetic architecture in girls and patients with low IQ autism.

1337W

A Novel Autosomal Dominant Dystonia and Spastic Paraplegia Caused by a Mutation in *ATP5G3*, a gene that encodes for subunit c of mitochondrial ATP Synthase. *D.E. Neilson[1], T. Huang[1], X. Wang[1], N.D. Leslie[1], R.B. Hufnagel[1], D.L. Gilbert[2].* 1) Div. of Human Genetics, Cincinnati Children's Hosp, Cincinnati, OH; 2) Div. of Neurology, Cincinnati Children's Hosp, Cincinnati, OH.

OBJECTIVE To identify the genetic basis for the neurological phenotype in a large, multigenerational family with a novel, autosomal dominant, highly penetrant neurological disease causing childhood onset dystonia and adulthood onset spasticity.

BACKGROUND: We previously characterized the phenotype of this disease which has highly variable onset between the first year and fifth decade. Clinically it presents with progressive, deep brain stimulation-responsive idiopathic generalized dystonia in childhood and/or spastic paraplegia in adulthood. Linkage to chromosome 2q24-2q31 was identified, in a region in which no prior inherited dystonias or spastic paraplegias had been identified.

DESIGN/METHODS: Forty family members were seen and examined of whom 18 were affected. Exome sequencing was performed on two distantly-related family members. The resulting analysis was constrained to the genetic locus on chromosome 2q. Confirmatory Sanger-based sequencing was performed on the remainder of the affected and unaffected family members. Functional analyses of isolated mitochondrial complexes, derived from patient fibroblast culture, were performed.

RESULTS: A single, novel heterozygous mutation (p.106N>K) in ATP5G3 was identified in a highly conserved region of the c subunit of mitochondrial complex V ATP synthase. This mutation segregated with disease in the family. ATP synthase catalyzes ATP synthesis during oxidative phosphorylation. The c subunits of this complex form the trans-membrane proton pore structure. ATP5G1, ATP5G2 and ATP5G3 provide functional redundancy. Mitochondrial assays demonstrate a statistically significant 20% reduction in complex V activity, whereas complexes I through IV displayed normal activity.

CONCLUSIONS: An amino acid substitution in complex c of mitochondrial ATP synthase causes progressive generalized dystonia and spastic paraplegia. Initial studies suggest that the mutant protein interferes with complex V activity, proportional to its relative expression. Tissue specific or stress-induced increases in expression of this mutant ATP5G3 may produce stronger reductions in activity. This may explain the restricted pattern of neuronal dysfunction. *In vivo* and *in vitro* studies currently in progress may reveal a mechanism for neurological dysfunction in this and other progressive neurological diseases.

1338T

Whole exome sequencing of patients with Rett-like features negative for MECP2 mutations:. *H.E. Olson, O. Khwaja, C. LaCoursiere, E. Martin, W.E. Kaufmann, A. Poduri.* Neurology, Boston Children's Hospital, Boston, MA., USA.

Objective: To identify genetic etiologies in cases with Rett syndrome or with Rett-like features when clinical testing for MECP2 mutations or deletions is negative. Methods: A cohort of eleven patients with Rett syndrome-like features, four meeting criteria for the disorder, and negative clinical testing for mutations or deletions in *MECP2* were recruited by a Rett syndrome specialist to the Core for Neurological Diseases at Boston Children's Hospital. We completed a detailed phenotypic analysis and performed whole exome sequencing. Results: Using 2010 diagnostic criteria, three patients had classical Rett syndrome and mutations in *MECP2* (two frameshift deletions and one pathogenic missense mutation). One patient met criteria for atypical Rett syndrome, with neonatal onset epilepsy including focal seizures and epileptic spasms, and had a frameshift deletion in *STXBP1*. The remaining patients had Rett-like features but did not meet criteria for Rett syndrome, most often due to lack of regression. One patient with Rett-like features without epilepsy had a missense mutation in *FOXG1*, and consistent MRI findings. One had a deletion in *MECP2*. For the remaining five, candidate genes were identified including known epilepsy genes. Conclusions: Whole exome sequencing is high yield for patients with Rett syndrome or Rett-like features negative for mutations in *MECP2*, though targeted gene testing may also provide a diagnosis. Genes associated with atypical Rett syndrome, epilepsy, or intellectual disability should be considered in cases not meeting criteria for Rett syndrome or when *MECP2* testing is negative. Clinical criteria supportive of classical Rett syndrome correlated well with *MECP2* mutations.

1339F

Detecting novel mutations of Alzheimer's disease gene using semiconductor sequencing. *Y. Ryoichi[1,2], M. Ryosuke[1,3], M. Hiroyuki[1], I. Yuishin[3], K. Masahito[1], K. Takashi[4], M. Hirofumi[1], M. Noriyoshi[2], K. Hidemi[2], K. Hideshi[1].* 1) Epidemiology, Research Institute for Radiation Biology and Medicine Hiroshima University, Hiroshimashi, Hiroshimaken, Japan; 2) Department of Periodontal Medicine, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan; 3) Department of Neurology, Tokushima University Hospital, Tokushima, Japan; 4) Department of Anatomical Pathology, Hiroshima University Hospital, Hiroshima, Japan.

Alzheimer's disease (AD) is a progressive neurodegenerative disease, characterized by an impaired ability to remember, poor judgment, impaired visuospatial abilities, language functions or changes in personality. Neuronal loss, senile plaque, and neurofibrillary tangle are pathological hallmarks of AD. To date, several genes have been identified as causative of AD, including PSEN1, PSEN2 and APP. These genes are involved in the amyloid-beta pathway, and are included in the new diagnostic guidelines for AD as 'probable AD dementia carriers of a causative AD genetic mutation'. In addition, APOE also points to an association with late-onset AD, in that APOEε4 allele frequency is high in sufferers. Until recently, we screened these genes using Sanger sequencing. However, it was time consuming and costly if used for routine diagnostic purposes. Over the past few years, high-throughput genome technologies have changed the genetic landscape of AD. Here, we used Ion sequencing technology, which uses an integrated semiconductor device, detects the hydrogen ions at non-optical, high speed. To reveal the causative gene of AD, we investigated 45 Japanese AD patients with positive family histories, and 27 sporadic patients with early-onset (<60 years old). In familial cases, we found 5 heterozygous variations in 4 genes. Two variations were detected in the PSEN1 gene. One is a known heterozygous missense mutation in the PSEN1 gene (c.488 A>G, p.H163R, rs63750590). Another novel heterozygous missense variation in the PSEN1 gene (c.1158C>A, p.F386L,). APOEε4 (c.471T>C, p.C130R, rs429358) was detected. The frequency of patients who had the APOEε4 allele was 31% in the familial group. None of the patients with PSEN1,PSEN2 and APP mutations carried the APOEε4 allele. In the early onset group, only a novel heterozygous missense variation (c.1262C>T, p.T421M) was detected in the PSEN2 gene. We detected mutations of the PSEN1 and PSEN2 that are causes of AD, using Ion sequencing technology. This technology can help to diagnose autosomal dominant AD that has been difficult to investigate with conventional methods, and will help to expand both clinical knowledge of the disease and consultations with patients and their families.

1340W

Exome sequencing of Parkinson's disease in order to identify genetic variants with high disease-risk. *W. Satake[1], Y. Suzuki[2], Y. Ando[1], H. Tomiyama[3], M. Yamamoto[4], M. Murata[5], N. Hattori[3], S. Tsuji[6], S. Sugano[2], T. Toda[1].* 1) Div of Neurol/Mol Bra Sci, Kobe Univ Grad Sch Med, Kobe, Japan; 2) Dept of Med Genome Sci, the Univ of Tokyo, Japan; 3) Dept of Neurol, Juntendo Univ Sch Med, Tokyo, Japan; 4) Takamatsu Neurology Clinic, Takamatsu, Japan; 5) Dept of Neurol, Natl Cent Hosp of Neurology and Psychiatry, Kodaira, Japan; 6) Dept of Neurol, the Univ of Tokyo, Tokyo, Japan.

Parkinson's disease (PD) is one of the most common neurodegenerative diseases worldwide, mainly manifesting motor impairment due to degeneration of dopamigenic neurons. Common form of PD appears following a multi-factorial inheritance pattern; that is, the majority of PD-patients show sporadic onset, and 5-10% of the patients show familial aggregation and an elevated relative risk ratio typical of disorders with complex inheritance. Relatives of an affected individual are more likely to have disease-predisposing alleles in common with the affected person than are unrelated individuals. It is reasonable that more genetic variants with high disease-risks of polygenic PD are thought to exist in patients with family history than in sporadic cases, although there are some cases who will be due to single-gene mutation causing parkinsonism of mendelian inheritance that is masked by small family sizes and incomplete penetrance. Therefore, performing exome sequencing of PD-patients with family history will be effective in order to identify genetic variants with high PD-risk, such as rare variants. We performed exome sequencing using 66 patients from 33 families. We obtained genome of 2 affected individuals from each family. We extracted exome using Agilent SureSelect and performed massive parallel sequencing using HiSeq2000 (average depth ×102). After BWA mapping, GATK calling, and dbSNP132 filtering, we detected an average of 565 nonsynonymous SNVs. Because 2 patients from the same family would share alleles with disease risk, we extracted and detected an average of 244 SNVs (per family) which are sheared between 2 patients from the same family. Among these, 793 SNVs were shared between more than 2 families in total, and 35 SNVs showed variant-frequency <0.25% in data of control samples, which may contain candidates of rare-variants with high risk for Parkinson's disease. We will examine association between these variants and PD by association studies using exome sequencing of sporadic cases.

1341T

Clinical exome sequencing identifies two novel IQSEC2 mutations associated with X-linked intellectual disability with seizures: implications for genetic counseling and clinical diagnosis. *S. Tang[1], S.K. Gandomi[1], K.D. Farwell Gonzalez[1], L. Shahmirzadi[1], J. Mancuso[2], P. Pichurin[2], R. Temme[3], S. Dugan[3], W. Zeng[1].* 1) Clinical Genomics, Ambry Genetics, Aliso Viejo, CA., USA; 2) Mayo Clinic, Department of Medical Genetics, Rochester, MN, USA; 3) Children's Hospital & Clinics of Minnesota, St. Paul, MN.

Intellectual disability (ID) is a heterogeneous disorder with a wide phenotypic spectrum. Over 1,700 OMIM genes have been associated with this condition, many of which reside on the X-chromosome. The IQSEC2 gene is located on chromosome Xp11.22 and is known to play a significant role in the maintenance and homeostasis within the neural environment of the brain. Mutations in IQSEC2 have been historically reported as causing non-syndromic X-linked intellectual disability (XLID) characterized by early onset limited intellectual functioning and limited adaptive behavior. Case reports of affected probands show phenotypic overlap with conditions associated with pathogenic MECP2, FOXG1, CDKL5, and MEF2C gene mutations. Affected individuals, however, have also been identified as presenting with additional clinical features including seizures, autistic-behavior, psychiatric problems, and delayed language skills. Although once thought to be a rare cause of XLID, IQSEC2 mutations are becoming more frequently detected through Next Generation Sequencing technologies utilized in clinical diagnosis. To date, a total of four unrelated families and 32 male probands are reported to carry mutations in this gene. Here we report two novel IQSEC2 de novo truncating mutations (c.2582G>C; p.S861T affecting splicing and c.2052_2053delCG; p.C684X) identified through diagnostic exome sequencing (DES) in two unrelated male probands manifesting developmental delay, seizures, hypotonia, plagiocephaly, and abnormal MRI findings. Both patients also presented with other mild features, not typically seen with IQSEC2 mutations, and neither patient presented with behavioral disturbances as seen in previously reported patients. Our two probands expand on the current understanding of genotype-phenotype correlations that exist for this gene. Our data also suggests that patients with truncating mutations in IQSEC2 are more severely affected compared with previously reported cases known to carry missense alterations. Overall, DES established a molecular diagnosis for two patients in whom traditional testing methods were uninformative while contributing to expanding the mutational and phenotypic spectrum of IQSEC2. In addition, our data clearly supports recently published data suggesting that IQSEC2 plays a more significant role in the development of XLID than previously anticipated. It should also be considered as a candidate gene in cases of male probands presenting with ID plus seizures.

1342F

Homozygosity mapping and exome sequencing in consanguineous Pakistani pedigrees with autism spectrum disorder and intellectual disability. *L. Xiong[1,2,4], S. Li[1,2], C.S. Leblond[2], S. Laurent[2], Q. Jiang[1,2], S. Zhou[2], D. Spiegelman[2], A. Ambalavanan[2], M. Christian[2], C. Caron[3,4], B. Forgeot d'Arc[1,2,4], S. Rasheed[5], Z.A. Nanjiani[6], M.Q. Brohi[7], L. Mottron[1,2,4].* 1) Centre de recherche, Institut universitaire en santé mentale de Montréal, Montreal, Canada; 2) Centre de recherche du CHUM, Montreal, Canada; 3) Centre for Excellent in Pervasive and Developmental Disorders, University of Montreal, Montreal, Canada; 4) Department of Psychiatry, University of Montreal, Montreal, Canada; 5) Autism Institute, Karachi, Pakistan; 6) Ma Ayesha Memorial Centre, University of Karachi, Pakistan; 7) Sir Cowasjee Jehangir Institute of Psychiatry, Pakistan.

Background: Pakistan has the highest rate of consanguineous marriage, due to historical, religious, cultural and social reasons. This population characteristic makes Pakistan an excellent location for collecting large families for genetic studies, particularly for autosomal recessive traits and for severe psychiatric disorders plagued by extreme heterogeneity and reduced reproductive fitness in outbred Western populations. **Methods:** In the past few years, with the cooperation from local geneticists and clinicians, we have identified and characterized 6 consanguineous pedigrees with 3-5 affected siblings/first-cousins per pedigree with variable intellectual disability (ID)and autism spectrum disorder (ASD), which fits with a recessive inheritance pattern. We have thoroughly investigated all the symptomatic individuals, including standard clinical examinations, comprehensive psychological evaluations, routine laboratory tests, MRI and EEG investigations; and have collected blood samples from all the affected individuals, their unaffected parents and siblings. We have carried out homozygosity mapping in these 6 ASD/ID pedigrees using high density DNA microarrays. Due to the extreme genetic heterogeneity of ASD/ID, as well as the efficiency and reduced cost of whole exome sequencing, we also performed whole exome sequencing in 2-3 selected affected individuals/pedigree. **Results:** We have identified extensive runs of homozygosity region in each individual genome, as well as shared identical-by-descent regions among the affected individuals in each or branch of these 6 pedigrees independently. Combined with whole exome sequencing, we have so far identified a novel frameshift nonsense mutation in the VPS13B gene in one family with 3 affected brothers with atypical Cohen syndrome and autism phenotype. We have also identified another frameshift nonsense mutation in the CC2D1A gene in one family with 5 affected brothers with intellectual disability and autism behaviors. Genetic validation of potential disease-causing mutations in other 4 pedigrees is underway. **Discussion:** Our results strongly indicate extensive genetic heterogeneity in these consanguineous pedigrees with ASD/ID phenotypes. Direct whole exome sequencing could be the most useful and efficient approach to identify causative disease mutations and to facilitate clinical diagnosis, as well as for much needed early intervention and genetic counseling.

1343W

Mutations in the mitochondrial chaperone *TRAP1* are associated with the triad of chronic fatigue, pain and gut dysmotility: Crazy, criminal, or just caught in the *TRAP*? *R.G. Boles[1,2,3], L.R. Susswein[1], S.A. Wong[1,2], K.J. McKernan[1], A.S Zare[1], T.R. Foss[1], H.A. Hornung[1], K.M Sheldon[1].* 1) Courtagen Life Sciences, Inc., Woburn, MA; 2) Division of Medical Genetics, Children's Hospital Los Angeles, Los Angeles, CA; 3) Medical School, University of Southern California, Los Angeles, CA.

Functional symptoms, such as chronic fatigue, pain, and gastrointestinal dysmotility, are common and can dramatically impair a person's quality of life, including leading to disability. Patients with these subjective symptoms are often discounted by the medical establishment as being complainers, mentally ill, or even causing the condition (Munchausen, including by proxy). While the etiology of functional symptomatology is largely unknown, these conditions are common in patients with mitochondrial disease, suggesting that genes involved in energy metabolism are good candidates in functional disorder pathogenesis. DNA from 270 patients referred for clinical testing was sequenced using Courtagen Life Science's nucSEEK™ Next Generation sequencing platform. The test sequences 1,195 nuclear-encoded genes for proteins that localize to the mitochondria or are associated with phenotypes which mimic mitochondrial disease. Ten patients were found to harbor mutations in the ATPase domain (I253V x 7 patients, E192K x 2, E216* x 1) of TNF receptor-associated protein 1 (*TRAP1*). Nine of 10 patients have a triad of chronic pain, fatigue, and gastrointestinal (GI) dysmotility (the 10th patient does not report dysmotility), versus only 10 of 95 patients without *TRAP1* mutations referred to us for sequencing (P=3×10-7). Only 1 of 17 patients with *TRAP1* mutations outside the ATPase domain have this triad (P=0.0001 v. in domain mutations); rather, these patients present with varying symptoms and the variants are likely unrelated to their disease. *TRAP1* is a mitochondrial antioxidant chaperone involved in preventing apoptosis under oxidative stress. Antioxidant supplementation is often used in the treatment of functional disorders. Therefore, we hypothesize that mutations in the ATPase domain of *TRAP1* prevent appropriate antioxidant chaperoning, which predisposes towards the development of functional disease symptomatology. Anecdotal clinical reports suggest substantial improvement, at least in fatigue and pain, among the 4 of our 10 patients known to be treated on antioxidants to date. More rigorous treatment with acetylcysteine is underway, with favorable initial responses. We propose the name of *TRAP1*-Related Disease (T1ReD) for this likely-treatable novel condition, and suggest that genes involved in antioxidant defense be investigated more thoroughly as candidates in functional disease pathogenesis.

1344T

GENETIC CAUSES OF INTELLECTUAL DISABILITY IN 22Q11.2 DELETION SYNDROME PATIENTS. *T. Guo[1], T. Wang[2], J. Chung[1], J. Cai[1], D. McDonald-McGinn[3], A. S. Bassett[4,5], E. Chow[4], M. Bowser[2], K. Devriendt[6], A. Swillen[6], J. Breckpot[6], A. Marie Higgins[7], N. Philip[8], C. Bearden[9], K. Coleman[10], D. Heine-Suner[11], J. Rosell[11], E. Zackai[3], M. Schneider[12], S. Dahoun[12], S. Eliez[12], M. Armando[13], S. Vicari[13], M. Cristina Digilio[14], A. Weizman[15], D. Gothelf[15,16], R. Shprintzen[7], W. Kates[17], B. Emanuel[3], B. Morrow[1], The International Chromosome 22q11.2 Consortium.* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York; 3) Division of Human Genetics, Children's Hospital of Philadelphia and Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania; 4) 4.Clinical Genetics Research Program, Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 5) Dalglish Family Hearts and Minds Clinic, Toronto General Hospital, 200 Elizabeth Street, Toronto, ON M5G 2C4 Canada; 6) Center for Human Genetics, University of Leuven, Leuven, Belgium; 7) The Virtual Center for Velo-Cardio-Facial Syndrome, Inc., Manlius, NY; 8) Department of Medical Genetics, University of Mediterranee and AP-HM, Timone Children's Hospital, Marseille, France; 9) Department of Psychiatry and Biobehavioral Sciences and Department of Psychology, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA 90095, USA; 10) Children's Healthcare of Atlanta, Nell Hodgson School of Nursing and Department of Human Genetics School of Medicine, Emory University, Atlanta, Georgia; 11) Genetics Department, Hospital Universitari Son Espases, Palma de Mallorca, Spain; 12) Office Médico-Pédagogique Research Unit, Department of Psychiatry, University of Geneva School of Medicine, Geneva 1211, Switzerland; 13) Child Neuropsychiatry Unit, Department of Neuroscience, Bambino Gesù Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Rome 00165, Italy; 14) Medical Genetic Unit, Department of Pediatrics, Bambino Gesù Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Rome 00165, Italy; 15) The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; 16) The Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer 52621, Israel; 17) Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, New York.

Velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome (22q11DS) is the most common microdeletion disorder, occurring in ~1:4,000 live births. Clinical features are quite variable but can often include conotruncal heart defects (CTDs), palatal, immune, endocrine and behavioral problems and intellectual disabilities (ID). We are investigating genetic risk factors to explain this variable intellectual functioning. This variability could be due to genes on the remaining allele of 22q11.2 or elsewhere in the genome. We obtained single time point Verbal, Performance and Full Scale IQ measurements from 559 subjects with 22q11DS. Average full-scale IQ scores (FSIQ) in 22q11DS subjects (73±14) was significantly lower than the average FSIQ score in the general population (100±15). Within 22q11DS subjects mean verbal IQ (VIQ 77±15) was significantly higher than mean performance IQ (PIQ 73±12). Also relevant is the comparison of IQ score to present or absence of cardiac anomalies. Subjects with CTDs had on average, a lower FSIQ score compared with those with normal heart or just a minor aortic arch anomaly (n = 322 with a CTD and 220 without; 72 vs 75, p=0.03). Although most individuals with the syndrome have a typical 3 Mb deletion, a subset have smaller deletions, thus we further investigated the possible association of deletion size on IQ. Among 559 DNA samples from 22q11DS subjects with IQ scores, 281 have Affymetrix SNP 6.0 array data. From this data, we found 85% had a 3 Mb deletion (LCR A-D), 8% had a 1.5 Mb deletion (LCR A-B) and 0.8% of the patients had a 2 Mb deletion (LCR A-C). Notably, individuals with a 3 Mb deletion have significantly lower FSIQ (72 vs 78; p=0.03) and VIQ (75 vs 84; p=0.009) versus those with the smaller deletion. This suggests that genes in the distal 1.5 Mb region (LCR C-D) may contribute to this difference. Among 14 genes in this region, CRKL and SNAP29 are expressed in brain and could contribute to ID. Analysis is underway for an additional 278 22q11DS DNA samples with IQ and Affymetrix 6.0 array data. In addition, whole exome sequence was performed on DNA samples from 69 22q11DS subjects including 28 samples with ID (IQ<70), 16 with borderline IQ (~75) and 25 with near normal IQ (IQ>80) using the Illumina HiSeq 2000 (NHLBI Resequencing Contract). Analysis of this data is underway. The results will help to identify candidate genes for ID, both on the remaining allele of 22q11.2 and outside of the deleted region.

1345F

Novel dominant associations with PANS, autism, anxiety, pain, fatigue and GI dysmotility identified by NextGen sequencing the 1,100 Mito-Carta genes in 270 probands. *L.R. Susswein[1], K.M. Sheldon[1], T. Foss[1], A. Zare[1], K.J. McKernan[1], S.A. Wong[1], R.R. Trifiletti[2], R.G. Boles[1].* 1) Courtagen Life Sciences, Inc., Woburn, MA; 2) Private practice, Ramsey, NJ.

Often referred to as 'functional', 'psychosomatic' or 'non-organic', disease composed of subjective symptomatology is a frequent cause of disability and poses a major challenge for physicians and researchers. Since these conditions are very common among patients with mitochondrial disorders, mitochondrial genes are good candidates in functional disease pathogenesis. Courtagen has NextGen sequenced all ~ 1100 genes encoding mitochondrial proteins (MitoCarta) in 270 unrelated individuals referred for clinical testing because of a suspicion of possible mitochondrial disease. Herein, we present novel disease associations. To limit type II errors due to multiple comparisons, candidates are identified based on family studies or an increased prevalence of deleterious-predicted variants among our patients in comparison to controls, and candidates are assigned based on the presence of a unique associated phenotype versus in a group representative of all patients referred for our testing. Below are three among many candidates. ***FPGS***: This gene encodes folylpolyglutamate synthetase, an enzyme that has a central role in folate homeostasis. R466C is the only suspicious variant in a mother and three-affected children with Pediatric Autoimmune Neurological Syndrome (PANS), and is found in 8/270 (3%) probands v. 0.5% in 1000 Genomes (p=0.0018). 7/8 probands share a PANS-like phenotype defined by the presence of autism, OCD or tics, v. 24/99 in our referral group (p=0.02). ***ABCB6***: This gene encodes a protein that transports porphyrins across the inner mitochondrial membrane. A deleterious-predicted variant was identified in 13 probands v. in 0/41 negative controls (p=0.2). Ten of the 13 probands (77%) were reported to have a PANS-like phenotype including anxiety, tics, OCD and/or loss of milestones v. 35/100 of our referral control group (p=0.006). The 3 remaining probands have developmental delay, whereas this phenotype may be unrecognized. ***TRAP1***: This gene encodes a mitochondrial chaperone. In 10 probands, three deleterious-predicted variants in the ATPase domain were detected, demonstrating a clinical triad of chronic fatigue, pain and GI dysmotility (P=3×10-7). **Conclusion**: Comprehensive data analysis among MitoCarta genes has revealed multiple candidates in disease pathogenesis for which additional study is ongoing. None of the above genes were previously associated with disease. All are potentially treatable, and anecdotal attempts at treatment have been successful.

1346W

Major Depression Susceptibility Loci Identified Using Whole Genome Sequencing in Extended Pedigrees. *E.E.M. Knowles[1], M.A.A. De Almeida[2], J.W. Kent[2], T.D. Dyer[2], J.E. Curran[2], H.H. Goring[2], M.A. Carless[2], R.L. Overa[3], D.R. McKay[1], E. Sprooten[1], J. Blangero[2], D.C. Glahn[1].* 1) Department of Psychiatry, Yale University, New Haven, CT; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 3) Department of Psychiatry, University of Texas Health Science Center, San Antonio, TX.

Background MDD is a common and costly disorder. While effective treatments exist many patients fail to receive them. Identifying genetic markers for depression may provide a reliable indicator of depression risk, which would substantially improve detection, and in so doing enable earlier more effective treatment. The aim of this study was to identify genes for depression modelled as a continuous trait, using whole genome sequence data. The sample comprised 530 Mexican-American individuals from extended pedigrees. Whole genome sequence (WGS) data with a 60-fold coverage were available for the entire sample amounting to ~3.4M SNPs per individual. Association testing was performed using a unitary factor score derived by applying CFA to all items from the Major Past Depressive Episode section of the Mini-International Neuropsychiatric Interview. Genome wide-significant hits were followed up using gene-specific analyses. WGS analysis revealed two variants on chromosome 3 that were significantly associated with depression, plus a number of variants that reached a suggestive level of significance. The first was located at ~188.0 Mb ($\Psi 2$ = 40.15, p = 2.35 × 10–10) and the other at ~67.0 Mb ($\Psi 2$ = 35.21, p = 2.96 × 10–09). Post-hoc analysis revealed a number of interesting candidate genes. Using a continuous measure of depression combined with WGS data we have identified genetic influences for depression. These genetic influences are located on chromosome 3 and overlap, in part, with identifications from linkage studies and also from the GWAS for depression carried out by the PGC.

1347T

Deletions Between 1 and 30 kb Associate with Risk for Autism. *J.D. Buxbaum, C.S. Poultney, A.P. Goldberg, M. Fromer, E. Drapeau, H. Harony-Nicolas, Y. Kajiwara, S. de Rubeis, S. Durand.* Icahn School of Medicine at Mount Sinai, New York, NY.

Genetic studies have shown a strong association between autism spectrum disorders (ASD) and both inherited and de novo copy number variation (CNV). Studies to date focused on CNV that were >30kb. In the current study, we investigated the role of smaller exome-targeted CNV.

Methods: Whole exome sequencing (WES) was performed on 811 subjects (432 ASD cases, 379 controls) using the Agilent Whole Exome 1.1 capture kit Illumina sequencing to generate 75 bp paired-end reads (Neale et al. 2012, Nature, 485:242–5). XHMM (eXome Hidden Markov Model; Fromer et al. 2012, Am J Hum Genet, 91:597–607) was used to call CNV from exome read depths after removing batch effects.

XHMM calls were filtered to retain CNV with XHMM quality score ≥ 65, number of exons spanned ≥ 3, CNV length ≥ 1 kb, per-sample number of CNV ≤ 55, per-sample total kb CNV ≤ 18 Mb, and MAF ≤ 1%. The filtered set contains 1386 CNV calls in 559 samples. These results were further stratified on type (deletion/duplication) and CNV size (1-10 kb, 10–30 kb, 30+ kb). Burden analyses were performed with PLINK on each subset to assess the CNV called per sample, the proportion of samples with CNV, and the number of genes hit by CNV per sample. After finding increased burden for all three measures in the 1–30 kb deletions, we chose a subset of those to validate using qPCR and/or Sanger sequencing.

Results: Strong enrichment was found in the 1-30 kb deletion subset, as measured by the number of CNV calls per sample (p=0.0037), proportion of samples with CNV (28% in cases vs. 21% in controls, p=0.017), and number of genes hit by CNV per sample (p=0.041). We used qPCR to attempt to validate 66 of the 219 deletions in the 1-30 kb range. Of these, qPCR confirmed 55 deletions overlapping the deletion predicted by XHMM. Further validation of 5 deletions via Sanger sequencing validated three deletions, with two technical failures.

Conclusion: We found a significant (p=0.017) burden of small (1–30 kb) deletions in ASD cases that represents a genetic finding that would be made in 7% of individuals with ASD.

1348F

Contribution of LIN7A to developmental disorder. *A. Matsumoto[1], M. Mizuno[2], N. Hamada[2], E. Jimbo[1], K. Kojima[1], M.Y. Momoi[1], K. Nagata[2], T. Yamagata[1].* 1) Pediatrics, Jichi Medical University, Shimotsukeshi, Tochigi, Japan; 2) Department of Molecular Neurobiology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan.

Disorder of synaptic proteins was a major pathophysiology of autism spectrum disorder (ASD). Among them, deletion, duplication and base substitution of SHANK3, one of the scaffolding protein working in the synapse, were detected as causes of autism. We focused on LIN7A and LIN7B that were another scaffolding proteins, and analyzed their contribution to intellectual disability (ID) and ASD. (Patients) Patient 1 is one-year-old boy with ID, spastic diplegia, facial anomalies and thin corpus callosum. Patient 2 is a boy with Asperger syndrome. Patient 3 was a girl with ASD. (Materials and Methods) Their DNA was extracted from lymphocytes after obtaining informed consent from their parents. Array CGH analysis was performed using Agilent Human genome CGH 180K. On the screening for mutation of LIN7B, each exon and intron nearby was amplified and subjected to direct sequencing. Biochemical fraction followed by Western blotting was performed to determine the subcellular localization. In utero electroporation, LIN7A or LIN7B was suppressed using RNAi on on embryonic day 14.5 and the brain was analyzed afer birth. (Results) On array CGH analysis of patient 1, a 14 Mb deletion on 12q21 was detected, and LIN7A was located on it. Patient 2 had a 73kb duplication on 19q31.33, where LIN7B was included. In patient 3, G to C base substitution on the donor site of exon 5 was detected. This base change induced exon 5 skipping and made the truncated protein missing a part of PDZ domain. Lin7A and Lin7B expressed in all fraction, and Lin7A slightly enriched in crude synaptosomal pellet and synaptosome fraction. Lin7B enriched in synapt, postsynaptic density fraction I(PSD-I), postsynaptic vesicle fraction II(PSD-II). Suppression of LIN7A or LIN7B in the embryonal brain resulted the delayed migration of cortical neurons. (Discussion) LIN7 was reported to bind with CASK that is a gene for ID, or NMDA type 2 receptor. We confirm the localization of LIN7A and LIN7B on the synapse. Our results suggested that LIN7A and LIN7B had an important role on neuronal cells and contributed to ID and ASD.

1349W

Cross-disorder copy number variation analysis of Tourette syndrome and obsessive-compulsive disorder. *J.M. Scharf on behalf of the TSAICG and IOCDFGC.* Psychiatric Neurodevelopmental Genetics Unit, Massachusetts Gen Hosp, Boston, MA.

Background: Tourette syndrome (TS) and Obsessive-compulsive disorder (OCD) are highly heritable neuropsychiatric disorders with evidence for shared genetic risk factors from twin and family studies. TS/OCD segregation patterns also suggest the presence of within-family pleiotropy in the expression of the disorders. In general, genetic pleiotropy has been noted in the copy number variation (CNV) literature, where several large genomic regions have been identified repeatedly to be associated with multiple neurodevelopmental disorders, such as autism, epilepsy, schizophrenia and intellectual disability (ID). This study, which represents the first genome-wide analysis of CNVs in OCD and the largest analysis to date in TS, addressed whether large (>500kb), rare (<1%) CNVs contribute to the genetic architecture of TS and OCD. **Methods:** Given prior evidence favoring a shared genetic relationship between TS and OCD, a cross-disorder design was employed to maximize power to detect rare events. The sample consisted of 2699 cases (1086 TS, 1613 OCD) and 1789 controls, including a subset of 348 OCD cases recruited as parent-proband trios to allow a *de novo* CNV analysis. CNVs were called with two algorithms (PennCNV and iPattern) to ensure reliability. **Results:** There was no increased burden of large, rare CNVs in the cross-disorder analysis or in secondary, disorder-specific analyses. However, a trend (p=.06) was noted for a >3-fold increase in deletion burden in regions previously associated with other neurodevelopmental disorders. Further examination revealed that a single locus, 16p13.11, harbored most of the neurodevelopmental burden (5 case deletions: 0 control deletions, p=0.09 in current study, p=0.025 compared to published control rates). Furthermore, three of these 16p13.11 deletions were confirmed as *de novo*, supporting the etiological significance of this region. **Discussion:** These results demonstrate that TS and OCD are associated with known neurodevelopmental deletions, most notably at 16p13.11 which previously was associated with ID and autism. 4/5 cases with a 16p13.11 deletion had OCD (3 OCD only, 1 OCD + tics) whereas one case had TS only; none had ID or ASD. These data are consistent with pleiotropy at the 16p13.11 locus with TS and OCD as part of the phenotypic spectrum. Further study of large, rare CNVs in OCD and TS will be important to expand the range of phenotypes associated with CNVs implicated across neurodevelopmental disorders.

1350T

Large deletion of the PRRT2 gene frequently involves paroxysmal dyskinesia in Korea. *SJ. Lee[1], MW. Seong[1], HJ. Kim[2], G. Ehm[2], HJ. Yang[2], YE. Kim[2], KT. Choi[1], SS. Park[1], BS. Jeon[2].* 1) Department of Laboratory Medicine, College of Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of Neurology and Movement Disorder Center, Parkinson study group, and Neuroscience Research Institute, College of Medicine, Seoul National University, Seoul, Korea.

Introduction: Recently, it has been found that mutations of PRRT2 cause paroxysmal kinesigenic dyskinesia (PKD) and other paroxysmal dyskinesias (PxDs). However, detailed clinical features of the patients with PRRT2 mutation in comparison with those without mutation are not well described. Furthermore, 16p11.2 microdeletions including PRRT2 also have been reported in patients with PKD but it is unknown to what extent PRRT2 deletion contributes to development of PKD and other PxDs. Methods: To address these issues, we performed mutation screening in 30 Korean patients with PxDs by sequence analysis and gene dosage analysis of PRRT2, then analyzed their clinical features. Results: Overall, genetic abnormality in PRRT2 was identified in 9 patient (30%), 3 of the 7 familial cases (43%) and 6 of the 23 sporadic cases (26%). The previously reported c.649dupC and c.649delC was found in 5 and 1 patient, respectively, and a novel mutation c.323_324delCA was found in 1 patient. The other two patients were found to have PRRT2 deletion involving at least exon 2 and 3. Compared with mutation-negative cases, age at PxDs onset was earlier in mutation-positive cases. However, there was no difference in other clinical features. Contrary to common belief that patients with PKC show excellent response to carbamazepine, 4 mutation-positive patients taking carbamazepine reported only partial responses. Conclusion: Our result shows that PRRT2 is the common causative gene for PxDs in Korea. The PRRT2 deletion might frequently involve development of PxDs and gene dosage analysis should be included in the genetic test for PxDs. The mutation detection rate in familial cases is lower than expected and compared with those in previous studies. This result suggests that other genetic, epigenetic or environmental factors are related to paroxysmal dyskinesia. Further studies are needed to identify the possible effect of these factors.

1351F

Epileptic encephalopathies of unknown etiology: de novo mutations and pathway analysis. *S. Esmaeeli Nieh[1], G. Da Gente[1], H.C. Mefford[2], I. Scheffer[3], A. Poduri[4], D. Dlugos[5], E.H. Sherr[1], EPGP and Epi4K investigators.* 1) Department of Neurology, University of California San Francisco, San Francisco, CA; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Department of Neurology, Royal Children's Hospital, Victoria, Australia; 4) Department of Neurology, Children's Hospital, Boston, MA; 5) Department of Neurology, Children's Hospital of Philadelphia, PA.

Background, Approximately 40% of all seizures occurring during the first three years of life are epileptic encephalopathies (EE), including Infantile Spasms (IS) and Lennox-Gastaut Syndrome (LGS), a devastating group of childhood disorders for which the cause is often unknown. We report whole exome sequencing data and related pathway / network-based study from 264 IS/LGS trios (patients and biological parents) as part of the Epi4k project, studying the genetics of EE (http://www.epgp.org/epi4k/). **Patients and methods**, Participants with EE were enrolled in the Epilepsy Phenome/Genome Project (EPGP, an NINDS-sponsored multi-center study). Detailed inclusion criteria are found at www.epgp.org. We sequenced exomes of the probands and their parents, and confirmed de novo mutations by Sanger sequencing. To identify potential biological pathways involved in EE, we conducted pathway/network-based analysis of confirmed *de novo* mutations using GenMAPP-CS interface, incorporating different databases (i.e. PathwayCommons, KEGG, Gene Ontology (GO)). To optimize pathway overrepresentation, pathways were statistically evaluated (Z-score, permutation, gene counts). **Results**, We sequenced the exomes of 264 probands and their parents, and confirmed 329 *de novo* mutations. We found nine genes with multiple *de novo* SNV mutations in two or more trios. We rank all sequenced human genes based on their tolerances to putatively damaging polymorphic gene variants in the human population. Interestingly, we observed an excess of *de novo* mutations to intolerant functional genetic variation, of which some already known to be associated with other neurodevelopmental diseases (Epi4k & EPGP Investigators, 2013, Nature). Pathway analysis pointed to a hotspot with more than 35 nodes for pathways involved in mitosis and cell growth, i.e. G2/M transition, Netrin mediated signalling events, MTOR and MAPK/ERK signalling pathways. Further investigation of pathways and networks is ongoing. **Significance**, We have identified *de novo* mutations in novel candidate genes for both IS and LGS. This large cohort of patients shows significant genetic heterogeneity underlying IS and LGS involving both known and novel genes. Despite this genetic heterogeneity, many of these single-hit mutations appear to converge on specific biologic pathways, suggesting pathway-based analyses of information across multiple genes may offer insight both for EE etiology and disease progression.

1352W

Assessing the potential clinical application of *SCN1A* genetic testing in Dravet and Doose epilepsy syndromes. *M.C. Gonsales[1], C.V. Soler[1], R.R. Kieling[2], A. Palmini[2], D.H. Nakanishi[3], M.L. Manreza[3], P. Preto[4], M.A. Montenegro[4], M.M. Guerreiro[4], I. Lopes-Cendes[1].* 1) Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas, Campinas, Brazil; 2) Porto Alegre Epilepsy Surgery Program, Hospital São Lucas, PUCRS, Porto Alegre, Brazil; 3) Clinical Hospital of Faculty of Medicine, USP, Sao Paulo, Brazil; 4) Department of Neurology, Faculty of Medical Sciences, University of Campinas, Campinas, Brazil.

Purpose Although several genes have been implicated with increased susceptibility to epilepsy, only a few are considered potentially useful for clinical testing. *SCN1A* mutations are associated with phenotypes within the spectrum of generalized epilepsy with febrile seizures plus (GEFS+), but the prognostic value of these mutations and a possible correlation with the different clinical subtypes remain unclear. Therefore, the aim of this study was to expand the knowledge on the clinical use of genetic testing for the most severe phenotypes within the GEFS+ spectrum, namely Dravet and Doose syndromes. **Methods** *SCN1A* mutation screening was performed in 21 patients with Dravet and 15 with Doose syndrome. Potentially deleterious mutations found were investigated in 100 individuals without epilepsy. In addition, eight algorithms were used to analyze the possible impact of missense mutations in protein function. Furthermore, MLPA analyzes were performed to detect copy number variations within *SCN1A*. **Results** We identified 14 potentially deleterious mutations in 15 patients with Dravet: six missense, four splice-site, two frameshift and two in-frame deletions. None of these were found in controls. Furthermore, all missense mutations are predicted to affect protein function. Two patients are monozygotic twins and presented the same alteration, an 18-basepair deletion, which was detected by both sequencing and MLPA. Interestingly, they also presented a missense mutation, which differently from the six mentioned above, was predicted as benign by five algorithms. In addition, this missense change was also present in their unaffected mother and in a control subject, strongly suggesting that the main cause of epilepsy in these patients is the deletion rather than the missense mutation. In contrast, only one patient referred to us as presenting Doose showed a potentially deleterious mutation, although this patient may have indeed a borderline/atypical phenotype. **Conclusion** Patients with Dravet syndrome showed a high frequency of *SCN1A* mutations (71%). Most mutations are missense (43%) and located in functionally important regions. However, we found that *SCN1A* testing in patients with classic Doose syndrome does not seem to be clinically relevant. In addition, we achieve a good discrimination power for identifying potential deleterious sequence variations in *SCN1A*, including missense changes by incorporating prediction algorithms and sequencing of control individuals.

1353T

Trio sequencing in malformations of cortical development. *E.L. Heinzen[1], A. Poduri[2], Epi4K and EPGP Investigators.* 1) Duke University, Durham, NC; 2) Boston Children's Hospital, Boston, MA.

Malformations of cortical development (MCD) are a heterogeneous group of conditions characterized by structural and cellular abnormalities in the cerebral cortex. These conditions arise from errors in cell proliferation, neuronal migration, or cortical organization during cortical development. MCD are strongly associated with severe, intractable epilepsy and intellectual disability. Germline mutations in a number of genes have been identified as the cause of particular subtypes of MCD; however, the cause of MCD in the majority of patients remains unknown. Given the typically sporadic presentation of MCD, we performed exome sequencing of leukocyte DNA of 125 patients with two common subtypes of MCD, polymicrogyria (PMG; n=68) and periventricular nodular heterotopia (PVNH; n=57), and their unaffected parents, to look for de novo mutations responsible for the disease. We identified disease-causing de novo mutations in two previously reported MCD genes: PIK3R2 (n=2) in PMG patients and in FLNA (n=3) in patients with PVNH. Similar to other neurodevelopmental disorders, we see no increase in the total number of de novo mutations in the MCD patient exomes. Comparing de novo mutations identified in this study to those recently published from another Epi4K study of epileptic encephalopathies, and to what was previously reported by other groups for autism spectrum disorders and intellectual disability, we see fewer MCD cases likely explained by germline or very early somatic mutations in known or novel genes. These data suggest that MCD may have a different genetic architecture than epileptic encephalopathies. One possible explanation is a larger role of post-zygotic somatic mutations in MCD that would be overlooked in leukocyte DNA, as has previously been reported in one severe subtype of MCD (hemimegalencephaly). Here we will discuss the results of this study, and discuss future opportunities for somatic gene discovery in MCD making use of fresh tissue specimens available through therapeutic resections.

1354F

Linkage analysis and exome sequencing in a large highly inbred consanguineous kindred to identify idiopathic generalized epilepsy genes. *F. Tuncer[1], S.A. Ugur Iseri[1], M. Calik[2], A.O. Caglayan[3], A. Iscan[4], M. Gunel[3], U. Ozbek[1].* 1) Genetics, Istanbul University Institute for Experimental Medicine, Istanbul, Turkey; 2) Department of Pediatric Neurology, Harran University, Faculty of Medicine,Sanliurfa, Turkey; 3) Departments of Neurosurgery and Genetics, Yale School of Medicine, New Haven, CT, USA; 4) Department of Pediatric Neurology, Bezmialem Vakif University, Faculty of Medicine, Istanbul, Turkey.

Epilepsy is a complex neurological disorder affecting 1% of the world's population. Among different forms of epilepsies, idiopathic generalized epilepsies (IGEs) are characterized by bilateral and synchronous generalized seizures in the absence of detectable brain lesions or metabolic abnormalities. Thus, the primary etiology for this disorder is believed to be genetic. The proposed study includes a large highly inbred consanguineous kindred with multiple IGE affected individuals and an ultimate aim in identifying novel epilepsy gene(s) to delineate the molecular basis of this disorder. Physical, neurological and electroencephalography (EEG) examinations were performed on the subjects recruited with information on family history, revealing 6 affected family members composed of 4 siblings and 2 cousins presenting varying seizure types. Experimental approaches included SNP genotyping of affected and 4 unaffected family members using Illumina Human HumanCytoSNP-12 BeadChip (300K). Genotype data was utilized using easyLinkage Plus software platform, where Mendelian genotyping errors were determined using PedCheck, multipoint lod scores were calculated under the assumption of autosomal recessive inheritance and haplotypes were constructed through GeneHunter. Linkage analyses were performed utilizing 4 affected sibs with their healthy brother and parents. Linkage peaks obtained were investigated through homozygosity mapping, which revealed partitioned homozygous regions in the affected cousins that were homozygous in the affected sibs. To facilitate in finding epilepsy genes, Illumina HiSeq2000 was used to perform exome sequencing on the affected child described to have drug resistance seizures and mental retardation. This child was chosen as the representative phenotype among the affected family members, due to the severest clinical presentation. Exome data obtained from this child was analyzed utilizing the pipeline developed by Yale University's bioinformatics team. Data was filtered for homozygous variants that either reside under the linkage peaks, within the homozygous regions shared by affected members and/or are novel and not caught by previous analyses. Validation of candidate variants obtained from exome data via Sanger sequencing, followed by familial segregation underline the current experiments. Prospective work includes healthy population screening for the candidate variant that shows familial segregation in line with affection status.

1355W

Mutation screening in the mitochondrial D-loop region in FMR1 premutation carriers. *F. Silva[1,2], L. Rodriguez-Revenga[1,3,4], I. Madrigal[1,3,4], MI. Alvarez-Mora[1,3], DM. Elurbe[1,3], M. Milà[1,2,3,4].* 1) Department of Biochemistry and Molecular Genetics, Hospital clinic Barcelona; 2) Fundació Clínic per a la Recerca Biomèdica, Barcelona, Spain; 3) CIBER de Enfermedades Raras (CIBERER), Barcelona, Spain; 4) IDIBAPS (Institut d'Investigacions Biomèdiques August Pi I Sunyer), Barcelona, Spain.

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset neurodegenerative disorder that occurs in FMR1 premutation carriers. Many studies have shown that mitochondrial dysfunction is involved in neurological diseases. The mitochondrial DNA (mtDNA) could be considered as a candidate modifier factor for FXTAS, since mitochondrial oxidative stress caused by mtDNA mutations has been implicated in the neuronal death in neurodegenerative disorders. The main goal of this work was to determine the mutation rate in the D-loop region, since this region has been considered a 'hot spot' for mutations in many diseases. In order to identify mutations, the hypervariable region I (HVR-I) was sequenced in 44 FMR1 premutated carriers, 20 of which presenting with FXTAS disease (11 females and 9 males) and 24 without FXTAS (18 females and 6 males), and 22 control individuals (22 males). Alignment was made with the revised Cambridge reference sequence. All FXTAS present one or more changes in the D-loop region, 22 of the 24 non-FXTAS and 21 of the controls present mutations. The T16519C change was the most prevalent mutation in all groups (55% FXTAS; 37,5% non-FXTAS and 54,5% controls). Our results do not show differences on the rate of D-loop changes between FXTAS, non-FXTAS and controls.

1356T

"Game of Exomes" and Autism Spectrum Disorder. *M. Cuccaro[1], N. Dueker[1], E.R. Martin[1,2], A.J. Griswold[1], H.N. Cukier[1], S. Slifer[1], J. Jaworski[1], I. Konidari[1], P.L. Whitehead[1], M. Schmidt[1], J.R. Gilbert[1,2], J.L. Haines[3], M.A. Pericak-Vance[1,2].* 1) Hussman Institute for Human Genomics, Univ Miami Sch Med, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN.

Autism spectrum disorder (ASD) is a highly prevalent developmental disorder, affecting an estimated 1 in 88 individuals, and is associated with significant morbidity. Despite demonstrating high heritability estimates, only a small proportion of the genetic risk for ASD is explained. While previous research, including genome wide association studies, has focused on the effects of common variants, recent exome sequencing studies suggest that rare variants (RVs) contribute significantly to ASD risk. Therefore, we analyzed RV data to identify loci associated with ASD. Participants were drawn from a large, family-based study of ASD and included 995 unrelated cases and 650 controls that were genotyped using the Illumina HumanExome-12v1 Array (HIHG sample). In addition, whole exome sequencing data on 1,014 cases and 866 controls from the ARRA autism sequencing consortium was utilized for replication and joint analyses (ARRA sample). To identify individual SNPs associated with ASD, we performed Fisher's Exact Test on all SNPs with MAF<5% and Fisher's method for combining p-values to perform meta-analyses. To identify genes associated with ASD we performed gene-based analyses testing autosomal genes genotyped on the Array for association with ASD. The sequence kernel association optimal test (SKAT-O), Cochran-Armitage (CA) Sum and CA Max tests were used for these analyses. A total of 108,045 SNPs were included in single-SNP analyses and none met chip-wide significance in our HIHG analyses or meta-analyses. Of the 12,612 genes analyzed in our HIHG sample, 27 genes were associated with $p<0.001$ in at least one test. A total of 13,064 genes were included in our joint analyses and 17 were associated with $p<0.001$ in at least one test. We also identified 19 ASD candidate genes with $p<0.01$, the most significantly associated being adenylate cyclase 5 (*ADCY5*) (SKAT-O $p=0.002$, CA Sum test $p=1.84\times10^{-4}$, CA Max test $p=3.9\times10^{-4}$). *ADCY5* is an excellent ASD candidate as it plays an important role in G-protein signaling and two separate reports have identified cases with de novo mutations in this gene. While our study failed to identify RVs of large effect size (individual RVs or RVs within genes) to be significantly associated with ASD, we did identify several nominally significant associations providing suggestive evidence for a role of RVs in ASD risk.

1357F

Role of *MAPT* variation in neurodegenerative disorders. *C. Labbe[1], A. Ortolaza[1], S. Rayaprolu[1], R. Uitti[2], D. Dickson[1], Z. Wszolek[2], O. Ross[1].* 1) Department of Neuroscience, Mayo Clinic , Jacksonville, FL; 2) Department of Neurology, Mayo Clinic, Jacksonville, Florida, USA.

Tau inclusions define the neurodegenerative diseases called tauopathies. Rare variants in *MAPT*, the gene encoding protein tau, cause tau dysfunction leading to neurodegeneration. A common non-recombining *MAPT* haplotype (*MAPT* H1) has been associated to several tauopathies including progressive supranuclear palsy, corticobasal degeneration, and Pick's disease; yet *MAPT* H1 has also been implicated in the risk to Parkinson's disease (which is not a traditional tauopathy). Preliminary sub-haplotype analyses suggest that different genetic variants on the *MAPT* H1 haplotype associate with each of these disorders. To date it remains unclear which variant(s) at the *MAPT* locus is(are) responsible for the risk and what is the underlying pathomechanism of disease. The objective of this study is to identify causal variants for Parkinson's disease and progressive supranuclear palsy (and other tauopathies) within the *MAPT* region using next generation sequencing technologies. We captured the entire *MAPT* gene and 10kb on each side, a 154kb genomic region that was sequenced in 300 patients with Parkinson's disease, 300 patients with progressive supranuclear palsy, 20 patients with Pick's disease, 30 patients with corticobasal degeneration and 300 controls using a pooling strategy (10 DNA samples/pool). We used the Haloplex system and designed 4248 amplicons for a total coverage of 96.3%. We selected: (a) common variants tagging each of the *MAPT* sub-haplotypes (frequencies >5%); and (b) rare variants most likely to be functionally relevant. Common and rare genetic variants identified are being genotyped in our independent case-control series (747 progressive supranuclear palsy patients vs 727 controls; Parkinson's disease: 692 clinically diagnosed patients vs 689 controls), with linear regression analyses to study all variants with a minor allele frequency (MAF) ≥1%. A collapsed marker approach will assess joint effects of variants with MAF<1%. We will present results using the latest sequencing and genotyping technologies to comprehensively define the *MAPT* locus associated with Parkinson's disease and progressive supranuclear palsy and thus identify novel targets for both neuroprotective and symptomatic therapies.

1358W

Whole genome sequencing in adults with 22q11.2 deletion syndrome: A pilot study. *N.J. Butcher[1,2], C.R. Marshall[3], A.S. Bassett[1,2,4].* 1) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada.

Background: 22q11.2 deletion syndrome (22q11.2DS) is a common multisystem genetic syndrome (~1/4000) associated with a hemizygous 22q11.2 deletion. Major features of 22q11.2DS include congenital and later onset conditions such as birth defects, intellectual disability, schizophrenia and most recently, Parkinson's disease. Despite the similar deletion between patients, the reduced penetrance and variable expressivity suggest mediation of the phenotype by variants on the intact chromosome 22q11.2 region and/or elsewhere in the genome. Methods: We conducted whole genome sequencing (WGS; Complete Genomics) and high-resolution SNP microarray genotyping in a pilot study of nine individuals with 22q11.2DS (aged 22.2 to 58.9 years; 5 male, 4 female) using DNA extracted from blood. Major phenotypes included serious congenital heart defects, n=4; schizophrenia, n=6; and Parkinson's disease, n=3. We examined each case for very rare sequence and copy number variants in the coding regions of the intact 22q11.2 chromosome. Results: Eight subjects had the typical ~2.5 Mb hemizygous 22q11.2 deletion; one had a typical nested proximal 1.5 Mb deletion. After filtering for segmental duplications, the total number of very rare coding sequence variants in the 22q11.2 deleted region ranged from 2 to 10 per case. Three of the nine individuals had a single very rare coding variant predicted to be damaging by SIFT and Polyphen (genes: MED15, TRMT2A, GNB1L). Four subjects, all with schizophrenia, had a very rare coding variant in CLDN5 (n=1 stop mutation; n=3 frameshift). Notably, there were no deleterious mutations in genes previously implicated in 22q11.2DS phenotypes, including COMT, PIK4CA, and SNAP29. Only one subject, without congenital cardiac disease, had a very rare coding variant in TBX1. There were no very rare copy number variations. Conclusions: These pilot findings suggest that expression of specific major conditions associated with 22q11.2DS, including congenital cardiac disease, schizophrenia, and Parkinson's disease, may not often be related to the unmasking of deleterious variants on the intact 22q11.2 chromosome. Replication in larger samples would be needed to confirm these findings. Studies of coding and non-coding variants in the rest of the genome, together with expression studies, may help reveal the genetic underpinnings of the variable 22q11.2DS phenotype and the aetiology of its major developmental, psychiatric, and neurodegenerative features.

1359T

Complete Genome Sequence Based Family Analysis of Monozygotic Twins Discordant For Schizophrenia. *C. Castellani[1], R. O'Reilly[2], S. Singh[1].* 1) Department of Biology, The University of Western Ontario, London, Ontario, Canada; 2) Department of Psychiatry, The University of Western Ontario, London, Ontario, Canada.

The reality of individual genome sequencing now offers a new hope in the search of the cause of complex diseases. When combined with genetic relationships, individual sequences add an unrivaled proficiency. Given the near identical genetic structure of monozygotic twins, any difference between monozygotic twins discordant for a disease will have a high likelihood of being causal. With this in mind we have sequenced the DNA of two pairs of MZ twins discordant for schizophrenia and one set of parents using the Complete Genomics system. The sequences were further assessed for accuracy in relation to Affymetrix SNP Array 6.0 results on the same samples. Genome wide variations including SNPs, indels, and CNVs were assessed. Our approach has allowed us to evaluate the similarities and differences across unrelated individuals, parents and children, as well as between MZ twins and to identify variants unique to affected individuals. Variant comparisons were performed using Golden Helix SNP and Variation Suite. The results show that an individual carries approximately 3.7 million SNPs, 400,000 indels, and 150 CNVs. Also, two unrelated individuals differed for 1.5-1.8 million SNPs (45 percent), a parent and child differed for 0.9-1.0 million SNPs (30 percent) and a pair of MZ twins differed for 100,000 (3 percent) SNPs. In our family analysis, a total of 968 variants were found in the affected twin that were not found in their Mother or Father. Of these variants, 138 were also present in the unaffected co-twin. Of the 830 unique variants to the affected twin, 6 variants were found to be in coding regions. In our pair-wise analysis, 24 and 40 coding variants respectively were found in the affected patient of twin pair 1 and twin pair 2 that were not found in their co-twins. It should be noted however that a number of these differences will be the result of sequencing errors and confirmation of variants of interest will serve as the next step of this analysis. The results support our strategy and identify patient specific genetic changes that may lead to schizophrenia. The novel results re-enforce that individual genomes harbor extensive variability, some inherited and others acquired during parental meiosis and/or mitosis during ontogeny. Even monozygotic twins are not identical and each individual may be a mosaic. This is supported by a high mutation rate and the persistence of genetic diseases with a severely reduced fecundity in all human populations.

1360F

Whole genome DNA sequencing identifies variants showing allelic association with bipolar affective disorder. *A. Fiorentino[1], N. O'Brien[1], A. McQuillin[1], A. Narula[1], A. Anjorin[1], R. Kandaswamy[1], R. Blizard[1], D. Curtis[2], H. Gurling[1].* 1) Molecular Psychiatry Laboratory, UCL, Rockefeller building, London, United Kingdom; 2) Department of Psychological Medicine, Queen Mary University of London, London, UK.

A number of different loci and chromosomal regions have been reported to be implicated in bipolar disorder (BD) through GWAS. These studies have succeeded in identifying specific regions which could be implicated in disease aetiology. It is therefore probable that rare aetiological variants exist in these regions. Variants increasing genetic susceptibility to bipolar disorder may be found in non-protein coding control regions of genes as well as in exons. We have sought to detect these variants using whole genome sequencing data from 99 BD subjects. Two of the best implicated BD susceptibility genes are CACNA1C (L-type voltage gated calcium channel alpha-subunit) and ANK3 (ankyrin 3). The region of CACNA1C that shows the strongest linkage disequilibrium signal from markers showing allelic association with BD is entirely within intron 3. For ANK3 there are two regions that show the strongest evidence for association and these are in intron 2 and intron 26 (isoform NM_001204403). We have selected variants in coding and noncoding cDNA regions, splicing regions, promoter regions and in potentially functional regions of CACNA1C intron 3. Variants with markedly different allele frequencies in the bipolar samples compared to European samples from the 1000 Genome Project were then genotyped in 1,510 bipolar subjects and 1,095 controls. We identified an A to G base pair change in ANK3 (rs139972937) at position 8266 resulting in a missense amino acid change from an asparagine to a serine at protein position 2643. The G allele encoding for serine was associated with BD at p=0.042. A base pair change (rs79398153) in the 3rd intron of CACNA1C likely to affect an ENCODE-defined control region also showed allelic association with BD (p=0.015). SNPs at the ANK3 and CACNA1C loci previously associated with BD in previous GWAS were not found to be in linkage disequilibrium with the two new possible aetiological base pair changes that we have found.

1361W

Whole genome sequencing for bipolar disorder. *A.E. Locke[1], M. Flickinger[1], T. DeAngelis[6], D. Absher[6], S. Vrieze[1], S. Ramdas[1,2], B. Li[2], J. Li[2], M. Burmeister[2,3,4], L. Legrand[10], M. Rivera[8], J.L. Sobell[9], W.G. Iacono[10], M. McGue[10], C.N. Pato[9], M.T. Pato[9], J.A. Knowles[9], G. Breen[8], J.B. Vincent[7], M. McInnis[3,5], G. Abecasis[1], H.M. Kang[1], L.J. Scott[1], S. Levy[6], R.M. Myers[6], M. Boehnke[1] on behalf of the Bipolar Research in Deep Genome and Epigenome Sequencing (BRIDGES) Study.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Psychiatry, University of Michigan, Ann Arbor, MI; 4) Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI; 5) University of MIchigan Depression Center, Ann Arbor, MI; 6) HudsonAlpha Institute for Biotechnology, Hunstville, AL; 7) Center for Addiction and Mental Health, Department of Psychiatry, University of Toronto, ON, Canada; 8) Medical Research Council Social, Genetic and Developmental Psychiatry Centre, Institute for Psychiatry, Kings College London, London, UK; 9) Department of Psychiatry and Behavioral Sciences, University of Southern California, Los Angeles, CA; 10) Department of Psychology, University of Minnesota, Minneapolis, MN.

Bipolar disorder is a psychiatric condition characterized by alternating periods of mania and depression that affects >1% of individuals worldwide. Despite estimated heritability between 45 and 70% and sibling recurrence risk estimates between five and ten, relatively little is understood about the genetics of this debilitating disease. Here we present results from whole genome sequencing at a mean depth of ~10x for 1,583 unrelated individuals (788 clinically diagnosed bipolar I cases and 795 controls with no diagnosed mental illness) of northern European descent, an initial data freeze for a larger study aiming to sequence 4,000 individuals. Using the gotCloud sequence analysis pipeline, we identify >32.3M SNP variants, including 14.3M singletons (44.4% of SNPs) after quality control filtering. 53% of these SNPs and 86% of the singletons are not catalogued in dbSNP build 135. Comparison to existing GWAS array data yielded overall genotype concordance of >99.9%. LD-aware genotype refinement significantly improved heterozygote concordance from 95% to 99.7%. In addition to SNP variants, we are currently calling indels and larger structural variants to generate an integrated map of genetic variation for use in testing association with bipolar disorder. In preliminary case/control association tests no single variant reached genome-wide significance. Low frequency variants upstream of *SLC24A2*, a calcium/potassium/sodium transporter expressed in the brain, had the strongest association signal ($P=7\times10^{-7}$). We replicate the same direction of effect for all five previously published bipolar disorder risk variants (rs12576775 at *ODZ*, rs4765913 at *CACNA1C*, rs9371601 at *SYNE1*, rs10994397 at *ANK3*, and rs736408 at *NEK4*), three at nominal significance ($P\leq0.05$). We will perform aggregate tests using SKAT and variable threshold tests to combine low frequency and rare variants across protein coding genes or annotated non-coding regions, with a particular focus on gene regulatory regions in brain tissues.

1362T

Identifying Functional Variation in Schizophrenia GWAS Loci by Pooled Sequencing. *E.K. Loken[1], S. Bacanu[1], D. Walsh[2], F.A. O'Neill[3], K.S. Kendler[1], B.P. Riley[1].* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, VA; 2) Health Research Board, Dublin, Ireland; 3) Department of Psychiatry, Queens University, Belfast, United Kingdom.

Mounting evidence supports excess rare variants in cases in GWAS-identified loci. Rare functional variants in these regions are independent of the common variant signal and provide independent support for the gene as part of the disease pathway. Pooled sequencing allows efficient target capture and library construction of this large sample. A large study of this type has not yet been pursued for the GWAS loci in schizophrenia. We use this method to interrogate 500 kb of schizophrenia GWAS loci exons from the first Psychiatric GWAS Consortium (PGC) schizophrenia analysis and additional evolutionarily constrained regions in our sample of 912 cases from the Irish Case/Control Study of Schizophrenia (ICCSS) sample and 1296 Irish population controls from the Trinity Biobank in Dublin. We sequenced 38 case pools and 54 control pools (total N=912 cases and 1296 controls, 24 per pool). Samples were quantified and normalized twice using PicoGreen followed by equimolar pooling. Pools were prepared for sequencing using the Agilent SureSelect targeted sequencing protocol and sequenced on the Illumina HiSeq platform using 2×100 bp reads. We selected exons and regions with significant mammalian conservation in linkage disequilibrium with the significant SNPs in the PGC analysis. We used Syzygy to call allele frequencies (AFs) from the sample pools. Variants will be validated using Sequenom iPLEX Gold. We compared called variant AFs with imputed AFs for the pool. We considered a correlation above 0.99 to be good evidence for successful equimolar pooling. We use the SNP-set Kernel Association Test (SKAT) to test all functional variants in a gene with MAF < 2%;. We also individually test variants >2%; MAF which are rated by PolyPhen as likely to be damaging. Preliminary sequencing at ~50x coverage per sample per pool for eight pools shows a high correlation (>0.99) between called AFs and 1774 known AFs from imputed genotype array data. We observe low false positive (2%) and false negative rates (<1%). These values should improve in the final ~100x coverage sequencing run. Our goal is to detect and validate causal rare variation in the PGC defined schizophrenia GWAS loci (MIR137, PCGEM1, TRIM26, CSMD1, MMP16, CNNM2, NT5C2, STT3A, CCDC68, TCF4, ITIH3/4, ANK3, and CACNA1C) and establish patterns of biased rare variation between cases and controls in GWAS loci using the region based test SKAT.

1363F

Exome sequencing identifies novel rare variants in human T-cell leukemia virus type-1-associated myelopathy/tropical spastic paraparesis. *S. Nozuma[1], E. Matsuura[1], Y. Higuchi[1], J. Yuan[1], Y. Sakiyama[1], A. Hashiguchi[1], Y. Okamoto[1], A. Yoshimura[1], T. Matsuzaki[2], J. Mitsui[3], H. Ishiura[3], Y. Takahashi[3], J. Yoshimura[4], K. Doi[4], R. Kubota[2], S. Morishita[4], S. Tsuji[3], S. Izumo[2], H. Takashima[1].* 1) Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan; 2) Molecular Pathology, Center for Chronic Viral Diseases, Kagoshima University, Kagoshima, Japan; 3) Neurology, The University of Tokyo Graduate School of Medicine, Tokyo, Japan; 4) Computational Biology, Graduate School of Frontier Sciences, the University of Tokyo, Tokyo, Japan.

The majority of human T-cell leukemia virus type-1 (HTLV-1)-infected individuals remain lifelong asymptomatic carriers, but some infected individuals develop HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP). HAM/TSP is a chronic inflammatory disease of the central nervous system that presents as slowly progressive spastic paraparesis with neurogenic bladder. Previous studies revealed that several host genetic factors were associated with the development of HAM/TSP. To identify the genetic factors contributing to HAM/TSP, we analyzed familial HAM/TSP. We performed whole-exome sequencing in 32 patients with familial HAM/TSP, 20 patients with sporadic HAM/TSP, and 20 HTLV-1 carriers. Rare variants with minor allele frequencies of less than 5% in 1000 genomes were considered for analysis. To predict loss-of-function genes, we determined whether missense mutations predicted by PolyPhen2 were damaged or frameshift indels. A total of 28,052 SNPs and 2097 indels were identified among patients with familial HAM/TSP. Rare variants identified in patients with familial HAM/TSP but not in HTLV-1 carriers were analyzed. The maximum number of patients with familial HAM/TSP who commonly shared the variants was 11. The affected rare variants shared by a minimum of 5 patients were investigated, and refining the variants to those associated with a predicted loss of function resulted in the selection of 23 genes. Among these, 5 genes were functionally related to the immune system, 2 genes were related to cell communications, and 1 gene was related to the nervous system. Some variants were detected in patients with sporadic HAM/TSP. These results demonstrate that rare variants are related to the development of HAM/TSP. Exome sequencing is a useful method to identify rare variants in common diseases.

1364W

Discovery and Analysis of Rare Variants for Bipolar Disorder by Exome Sequencing in Multiplex Families. *S. Ramdas[1], A.B. Ozel[2], J. Li[2].*
1) Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Bipolar disorder (BPD) is a severe psychiatric disease marked by alternating manic and depressive episodes. Twin and family studies have shown that BPD is highly heritable; but extensive genetic and phenotypic complexity remains a major barrier to identifying its genetic basis. In this study we focus on multi-generation multiplex BPD families in the NIMH repository. We hypothesize that these families manifest a nearly Mendelian or at least highly penetrant subset of bipolar cases and may transmit one or more high-impact coding variants that alter the function of key neurodevelopmental or neural signaling genes. We genotyped and performed exome sequencing for 82 affected individuals in 29 families representing first cousin pairs or more distant relatives, aiming to detect functionally damaging rare variants in regions shared by affected relatives. Comparison of exome sequencing and HumanExome Beadchip genotype data revealed high levels of concordance: on variant sites that were both genotyped and had a called genotype in the exome data, overall concordance is 0.98 and heterozygote concordance is 0.91. Variant filtering were applied to identify those that are bi-allelic, have a minor allele frequency of <5% in European samples from the Exome Sequencing Project, and are damaging missense, nonsense, or splicing site variants, resulting in a median of 336 variants per family. As Exome chip genotype data allow inference of genomic segments shared among family members identically by descent (IBD), we used Beagle to identify IBD regions shared by BPD cases within each family. This led to a further reduction in the number of candidate variants. In one family, only one gene, FOXD4 (Forkhead box protein D4), was identified after filtering. This gene has previously been implicated in suicidality and Obsessive Compulsive Disorder (OCD). Members in another family carry damaging variants in DRD5 (Dopamine receptor D5) and GRIK3 (Glutamate receptor, ionotropic kainate 3), which were reported in association studies of ADHD and schizophrenia respectively. Additional filtering and validation are underway for the candidate variants across all families. (This study is supported by the IMHRO - Johnson & Johnson Rising Star Translational Research Award.).

1365T

Gene based analysis on major depressive disorder using 12,000 case control samples sequenced at one fold coverage. *Y. Li[1], J. Wang[2], S. Shi[3], Y. Chen[4], J. Marchini[1], K. Kendler[5], R. Mott[1], J. Flint[1] on behalf of the CONVERGE Consortium.*
1) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, Oxfordshire, United Kingdom; 2) Beijing Genomics Institute, Floor 9 Complex Building, Beishan Industrial Zone, Yantian District, Shenzhen 518083, P.R.China; 3) Huashan Hospital of Fudan University, No.12 Middle Wulumuqi Road, Shanghai, P.R.China; 4) CTSU, Richard Doll Building, Old Road Campus, University of Oxford, Headington, Oxford OX3 7LF, United Kington; 5) Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, PO Box 980126 Richmond VA 23298, USA.

We present results for low pass sequencing from the CONVERGE project, a case-control study of major depressive disorder (MDD) using 12,000 Han Chinese women. To date, no genetic loci have been robustly associated with MDD using genome-wide association based on genotypes called from arrays. Low pass sequencing allows novel approaches to genotyping, particularly in accessing low frequency variants. We applied gene enrichment tests on functional SNPs with different levels of filtering and allele frequency cut-offs to test whether cases were enriched with disease causal SNPs than controls in any genes, and if so, where they are on the allele frequency spectrum. The filtering includes base quality, mapping quality, exclusion of non-pair end reads and alternative allele supported by at least two reads. The gene enrichment analysis has given us candidate genes, which can now be verified by targeted re-sequencing. Furthermore we report higher copy number of the mitochondrial genome in cases with MDD. This is the largest GWAS using whole genome low coverage sequencing data. Although we face the challenge of calling and imputing low frequency variants accurately, my pipeline allows different levels of filtering at an individual level, which enables us to examine rare variants aggregately down to a singleton level. The results help to answer an important question about whether rare functional variants are responsible for the genetic susceptibility to MDD.

1366F

Exome sequencing identifies de novo missense mutation in *KCND2* in identical twins with autism and seizures that results in slow potassium channel inactivation. *S. Nelson[1,2], H. Lee[2], M. Lin[3], H. Kornblum[4], D. Papazian[3].*
1) Dept Human Genetics, UCLA Medical Ctr, Los Angeles, CA; 2) Dept Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 3) Dept Physiology, UCLA, Los Angeles, CA; 4) Dept Pediatrics, Los Angeles, CA.

There have been numerous studies and case reports showing co-existence of autism and epilepsy suggesting significant overlap between the two phenotypes. However, the relationship between the two on the molecular level remains unclear. Here, we performed whole exome sequencing on a family with identical twins affected with autism and severe, intractable seizures. A de novo variant p.Val404Met was identified in *KCND2* gene that encodes A-type potassium channel Kv4.2 which forms the pore of the somatodendritic subthreshold A-type potassium current (I_{SA}) channels. The p.Val404Met variant is a novel variant and occurs at a highly conserved residue within the C-terminal end of the transmembrane helix S6 region that makes up the ion permeation pathway. In available exome data, the transmembrane domains in Kv4.2 protein have a low frequency of predicted-to-be damaging variants relative to the cytoplasmic domains indicating strong selection against variation in the pore domains. To investigate the functional effect of the heterozygous p.Val404Met mutation on the properties of Kv4.2 channels, reference and mutant proteins were expressed together and individually in Xenopus oocytes and analyzed using a two electrode voltage clamp. The mutant channel demonstrated significantly reduced rate of both the opening and inactivation of the channel. Co-expression of the reference with the mutant channels was mostly indistinguishable from the mutant channel alone demonstrating the autosomal dominant gain of function mutational effect. The effect of the mutation on closed-state inactivation was evident in the presence of the auxiliary subunits that associate with Kv4 subunits to form I_{SA} channels in vivo. Our results taken together with previous reports strongly suggest that the size of the residue at position 404 is a major factor in determining the likelihood of inactivation and given the key role of closed-state inactivation in determining the steady state availability of I_{SA} channels in neurons, the dominant phenotype of p.Val404Met is likely to disrupt the normal physiological functions of I_{SA}. Discovery of a predicted to be damaging, novel de novo variant in *KCND2* at a conserved residue, coupled with physiological evidence of the mutant protein disrupting the potassium current inactivation impose a strong indication that *KCND2* is the causal gene for the epilepsy in this family, and provides suggestive evidence of a role in the autistic features that they exhibit.

1367W

Amish revisited: A next-generation sequencing study of bipolar disorder among the Plain people. *L. Hou[1], N. Akula[1], L. Kassem[1], D. Chen[1], Y. Yao[2], J.L. Haines[3], M. Pericak-Vance[4], T.G. Schulze[1,5], F.J. McMahon[1].* 1) Human Genetics Branch, National Institute of Mental Health Intramural Research Program, Bethesda, MD USA; 2) Unit of Statistical Genomics, National Institute of Mental Health Intramural Research Program, Bethesda, MD, USA; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN USA; 4) Hussman Institute for Human Genomics, University of Miami, Miami, FL USA; 5) University of Goettingen, Goettingen, Germany.

Bipolar disorder (BP) is a common, highly heritable mental illness. A number of risk loci have been identified by GWAS, but the large proportion of unexplained genetic variance suggests that rare variants might contribute to risk of the disorder. Isolated populations with large families, like the Amish, have many advantages for discovering rare genetic susceptibility loci, such as lower allelic and locus heterogeneity, higher environmental and phenotypic homogeneity, and enrichment of rare disease alleles. We are ascertaining distantly related cases of BD and their relatives from Plain (Amish and Mennonite) communities in the US. So far we have 170 study participants, 94 of whom have been assigned a diagnosis of BD with high confidence. Identity-by-descent (IBD) analysis of SNP array (Illumina OmniExpress) genotypes (by use of Beagle v 3.3.2) demonstrates sharing of many long (>5 cM) chromosomal segments, consistent with known pedigree relationships, state of residence, and Amish vs. Mennonite affiliation. To identify IBD segments that were shared more often by case-case pairs than others, we combined the OmniExpress data with SNP array (Affymetrix 6.0) data from 893 Midwestern Amish individuals (collected by JLH & MPV), extracted 154,271 overlapping markers, excluded closely related pairs (pi-hat >0.1), and carried out a shared segment analysis with PLINK (v 1.07). Several highly shared IBD segments were identified on chromosomes 1p34.3, 2q31.1, 7p22, 7q36.3, 9p24.2-p24.1, and 10q22.1 at permutation p-values ranging from 0.01 to 2E-05. Interestingly, SNPs in several of these segments have been reported to be associated with bipolar disorder by previous GWAS. Exome sequencing has so far been carried out on 21 distantly related cases using the Illumina HiSeq system. The BWA/GATK package was applied for single nucleotide and insertion-deletion variant calling, with filtering of common SNVs (MAF > 5% in the Exome Sequencing or 1000 Genomes Project). Five novel, nonfunctional SNVs and 9 potentially functional SNVs, including 8 non-synonymous and 1 frame-shift insertion, were found within the highly shared regions and were each detected in at least 4 cases. While additional shared segment analysis and sequencing in larger samples is needed, these preliminary results suggest that rare variants may contribute to BD in Plain populations.

1368T

Genetic etiologies of speech sound disorders. *B. Peter[1], E.M. Wijsman[2,3], M. Matsushita[2], K. Oda[2], K. Chapman[4], U.W. Center Mendelian Genomics[5], I. Stanaway[6], W.H. Raskind[1,7].* 1) Dpt. of Speech & Hearing Sciences, University of Washington, Seattle, WA; 2) Dpt. of Medicine, University of Washington, Seattle, WA; 3) Dpt. of Biostatistics, University of Washington, Seattle, WA; 4) Dpt. of Communication Sciences & Disorders, University of Utah; 5) Dpt. of Genome Sciences, University of Washington; 6) Dpt. of Environmental and Occupational Health, Toxicoloy, University of Washington; 7) Dpt. of Psychiatry and Behavioral Sciences.

Speech sound disorders (SSDs) are disorders of the speech output mechanism, often diagnosed before the age of 4. Children with SSD fail to develop speech that is easily understood by others, due to omitted, distorted, or substituted speech sounds. SSD runs in families and there is evidence of a genetic etiology but the genetic pathways are poorly understood. One proposed SSD subtype is childhood apraxia of speech (CAS), a severe SSD subtype that is thought to result from deficits in motor programming. In some forms of syndromic CAS, the genetic etiology is known, e.g., GALT mutations in galactosemia and FOXP2 mutations in severe speech and language disorder. The genetic etiology of idiopathic CAS has not been delineated. We study multigenerational families with familial idiopathic CAS from phenotypic and genetic perspectives. Our results indicate that motor programming deficits are not limited to the speech production system but can also be observed in hand motor tasks, primarily when the task involved alternating-sequential movements, not repetitive ones. An even more comprehensive deficit in sequential processing emerged from group differences in adults with, and without, a childhood history of CAS in tasks involving sequential processing on linguistic and cognitive levels, for instance nonword reading and spelling. To investigate the genetic etiologies of CAS, we use approaches made possible by our multigenerational family design. A combination of linkage analysis, exome sequencing, and identity-by-descent studies are leading to the identification of new candidate regions and genes for CAS. In one family consisting of 23 individuals in three generations, two new candidate genes, expressed in relevant brain regions, were identified on chromosomes 3 and 5. Similar genetic analyses in other multigenerational families are underway. The long-term goal of this study is to create a catalogue of genetic etiologies in SSD to aid in early identification of infants at genetic risk for SSD and development of effective intervention and prevention approaches to avoid or ameliorate the devastating social, psychological, and educational trauma of unintelligible speech in the preschool and early school years. Sequencing was provided by the University of Washington Center for Mendelian Genomics funded by NIH grant 1U54HG006493 to Drs. D. Nickerson, J. Shendure and M. Bamshad.

1369F

Comprehensive annotation and analysis of noncoding autism spectrum disorder risk loci by targeted massively parallel sequencing. *A.J. Griswold[1], D. Van Booven[1,2], J.M. Jaworski[1], S. Slifer[1], M.A. Schmidt[1,2], W. Hulme[1], I. Konidari[1], P.L. Whitehead[1], J.A. Rantus[1], S.M. Williams[3], R. Menon[4], M.L. Cucccaro[1,2], E.R. Martin[1,2], J.L. Haines[5], J.R. Gilbert[1,2], J.P. Hussman[6], M.A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN, USA; 4) Rollins School of Public Health, Emory University, Atlanta, GA, USA; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN, USA; 6) Hussman Foundation, Ellicott City, MD, USA.

Genome-wide association studies (GWAS) of autism spectrum disorder (ASD) show that common variants with small effects do not greatly impact ASD risk. Exome sequencing studies, while implicating rare, *de novo* protein coding variations in a few genes, suggest that many rare variants in hundreds of genes influence ASD. A largely unstudied hypothesis is that rare, functional variants in non-protein coding regions of the genome contribute to ASD. To identify such variants, we sequenced 17Mb of ASD associated genes and genomic loci (Hussman et al, 2011) in 919 cases and 854 controls. We targeted exons, including UTRs, of 681 genes, and conserved regions in their introns, 5kb from the 5' and 3' ends, and ASD associated intergenic regions. This is the first large scale study of noncoding variants in ASD. We identified 427,676 single nucleotide variants (SNVs), of which only 20,293 were in coding exons. We annotated each noncoding SNV with databases including ENCODE (DNaseI sensitivity and transcription factor binding sites), VISTA (noncoding segments with demonstrated gene enhancer activity), and GENCODE (evidence based noncoding transcribed loci). SNVs without predicted functional effects are assessed using predictions for splice enhancing ability, RNA secondary structure effects, and transcriptional regulatory potential. Restricting our analysis to the 358,432 rare, noncoding SNVs (MAF≤0.01 in our dataset), 113,896 are predicted to have a functional role in at least one database. Of these, 11 occur at least once and uniquely in cases and had a score of 2a in the RegulomeDB, suggesting they affect a known transcription factor binding site motif. Two of these are in the potassium channel gene *KCNN3*, in AR and POU2F2 binding sites. We used the sequence kernel association test (SKAT) to detect association between ASD and sets of SNVs in noncoding elements while adjusting for population stratification. SKAT did not identify multiple testing corrected association of ASD with any noncoding feature. However, a promising nominal association was detected in enhancer hs1316 (p=0.03) that enhances expression in the midbrain of mouse embryos. To increase statistical power and identify new variants, sequencing of an additional 1525 cases and 425 controls is underway. This work adds noncoding SNVs to the growing list of ASD risk loci and may implicate biological mechanisms beyond protein coding changes contributing to ASD etiology.

1370W

Genome-wide frequencies of phase-specific functional variant combinations in a large set of families with bipolar disorder. *D.X. Quarless[1,4], K. Bhutani[2,4], V. Bansal[5], A. Torkamani[5], N. Rens[9], S.A. Ament[6], H. Li[6], M.E. Brunkow[6], R. Gelinas[6], J. Nurnberger[7], H. Edenberg[7], E. Gershon[8], G. Glusman[6], J.C. Roach[6], J.R. Kelsoe[3], N.J. Schork[4,5], The Bipolar Genome Study Consortium.* 1) Biomedical Sciences Department, University of California San Diego, La Jolla, CA; 2) Bioinformatics Graduate Program, University of California San Diego, La Jolla, CA; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA; 4) The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA; 5) Department of Molecular and Experimental Medicines, The Scripps Research Institute, La Jolla, CA; 6) Computational Biology Genetics Genomics Immunology, Institute for Systems Biology, Seattle, WA; 7) Department of Psychiatry, Indiana University, Indianapolis, IN; 8) Department of Psychiatry, University of Chicago, Chicago, IL; 9) Johns Hopkins University, Baltimore, MD.

There is growing interest in the identification of complex combinations of inherited variants that influence phenotypic variation. These combinations can implicate any number of genomic functional elements. In addition, many potential phenotypically relevant variant pair combinations can operate either in cis or trans and thus their identification requires phase information. For example, protein coding compound heterozygous variant pair combinations acting in trans and coupled, epistatic regulatory and protein coding variant paired combinations acting in cis could not be identified if the nucleotide content of the two homologous chromosomes an individual possesses has not been differentiated. We characterized the number of phase-specific functional variant combinations in the genomes of 114 individuals within 35 different families using a suite of computational tools for both genome phasing and functional variant annotation. We find that a typical human genome has a large number of likely phase-specific functional variant pairs. We comment on the implications of our findings for sequencing and gene mapping studies focusing on rare and common conditions, as well as clinical studies seeking to identify functional variants for idiopathic diseases.

1371T

Investigation of the role of microRNAs in autism by deep RNA-sequencing. *J.A. Lamb[1], F. Marriage[1], E. Tsitsiou[2], P. Wang[2], L. Zeef[2].* 1) Centre for Integrated Genomic Medical Research, Faculty of Medical and Human Sciences, University of Manchester, Manchester, United Kingdom; 2) Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom.

Background. Numerous heterogeneous rare intragenic variations and copy number variants have been recently identified in autism, with predicted effects on protein function. Non-coding RNA has an important function in genome regulation, and there is evidence for a role of microRNAs in development and in several neuropsychiatric and neurological disorders. MicroRNAs act as pleiotrophic regulators of gene expression, targeting a large number of neuronal genes, suggesting that genetic variation in microRNAs may underlie the clinical heterogeneity of autism. **Objectives**. To identify novel genetic variation in microRNAs and differentially expressed microRNAs in individuals with autism compared to healthy controls. **Methods**. Genomewide microRNA expression profiling was carried out on lymphoblastoid cell line RNA from 60 individuals with autism and 10 healthy controls on the SOLiD 4.0 analyzer. Sequence data was processed and mapped to the human genome reference sequence (SeqTrimMap and Bowtie), and against all known human mature microRNAs (miRBase v19). Differentially expressed microRNAs were identified using DESeq and false discovery rate correction. Dysregulated microRNAs were validated by qRT-PCR. Variant calling was carried out using Samtools and filters including minimal base and map quality. **Results**. Seven differentially expressed microRNAs were identified in samples from individuals with autism compared to healthy controls (6 up-regulated, 1 down-regulated; FDR $P<0.05$). Altered expression profiles were validated by qRT-PCR. Data for hsa-miR-130b are indicative of possible arm switching in individuals with autism. Seven novel sequence variants were identified in the seed region of the mature microRNA in individuals with autism that were not present in control samples or in dbSNP (137). These may alter target mRNA binding and are therefore predicted to be functionally deleterious. Validation of these novel variants and segregation analysis within families of affected individuals is ongoing by Sanger sequencing. The putative mRNA targets of differentially expressed microRNAs and microRNAs containing novel variants are being investigated using *in silico* target prediction programmes (Ingenuity Pathway Analysis). **Conclusion**. A number of differentially expressed microRNAs and novel variants within microRNA genes have been identified in individuals with autism. These may affect target mRNA binding, suggesting a possible pathogenic role in individuals with autism.

1372F

Somatic Instability in Sporadic Amyotrophic Lateral Sclerosis? *C.S. Leblond[1,2], JB. Rivière[3,4], M.J. Strong[5,6], K. Volkening[5,6], P. Hince[2], D. Spiegelman[2], A. Dionne-Laporte[2], J. Robertson[7], L. Zinman[8], P.A. Dion[2,9], G.A. Rouleau[2].* 1) McGill university, Montreal, Canada; 2) Montreal Neurological Institute and Hospital, Neurology and Neurosurgery departement, Mc Gill University, Montreal (Qc) Canada; 3) Equipe Génétique des Anomalies du Développement (EA 4271 GAD), Université de Bourgogne, France; 4) Laboratoire de Génétique Moléculaire, Centre Hospitalier Universitaire Dijon, France; 5) Molecular Brain Research Group, Robarts Research Institute, University of Western Ontario, London (On) Canada; 6) Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, University of Western Ontario, London (On) Canada; 7) Tanz Centre for Research in Neurodegenerative Diseases, and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto (On) Canada; 8) Sunnybrook Health Sciences Centre, Toronto (ON) Canada; 9) Pathology and Cellular Biology department of Montreal University, Montreal (Qc) Canada.

Amyotrophic lateral sclerosis (ALS) is a devastating neurological disease characterized by the degeneration of upper and lower motor neurons of the motor cortex, brainstem and spinal cord. The majority of ALS cases are sporadic (sALS) and only 5-10% have a family history (fALS). SALS cases are clinically and pathologically very similar to FALS cases. Whereas 60-80% of fALS cases are explained by germinal mutations, the genetic etiology of sALS remains unclear. The lack of genetic evidence in sALS suggests that inherited or germinal de novo mutations are very rare events in this disease. Nonetheless it is important to keep in mind that current genetic evidence for sALS comes from genetic examinations made using blood DNA for the most part and to a lesser extent from available post-mortem neuronal tissues. Here, we hypothesize that somatic mutations appear early during the embryogenesis events that underlie the development of the spinal cord. These somatic events could later trigger the development of ALS in sporadic cases, as observed in patients with germline ALS-predisposing gene mutations. Given ALS is a motor neuron disease that essentially affects the corticospinal tract, we are in the process of preparing a total of one thousand sections isolated from flash frozen SALS patient spinal cords. These sections will be individually screened for the presence of mutations in the most commonly observed ALS genes. At first, we will look for repeat expansions in C9ORF72 by repeat-primed PCR (RP-PCR) since repeat mutations have been shown to be unstable during replication and DNA repair. Following this, we will look for point mutations in SOD1, TARDBP and FUS using a targeted deep sequencing method 'single molecule Molecular Inversion Probe (smMIP)' specifically designed to identify rare somatic events. The smMIP method has been shown to be highly accurate and sensitive for the detection of sub-clonal variations based on the tagging of amplified DNA molecules that originated from one of the various inputted DNA molecules from the different cells found in a tissue or sample. Preliminary results showed that our RP-PCR conditions are sensitive enough to detect expansions at a low level of mosaicism (20%). To date, we have examined 499 spinal cord sections that were derived from 13 sALS patients for the presence of expanded C9ORF72 alleles but have thus far detected no expansion.

1373W

Towards the identification of new genes implicated in recessive early-onset forms of Parkinson's disease. *S. Lesage[1], A.S. Cocquel[1], A.L. Leutenegger[2], A. Honoré[1], C. Condroyer[1], A. Dürr[1,3], A. Brice[1,3].* 1) INSERM UMR_S975, ICM, Pitié-Salpêtrière hospital, Paris, France, France; 2) INSERM UMR_S946, Fondation Jean Dausset-CEPH, Paris, France; 3) Département de Génétique et Cytogénétique, Hôpital de la Pitié-Salpêtrière, Paris, France.

BACKGROUND: Parkinson's disease (PD) is a progressive neurodegenerative disorder, probably resulting from the interplay between genetic and environmental factors. To date, more than 15 loci have been identified and 6 genes are now confirmed to be causative for monogenic forms of PD, including 4 genes (Parkin/PARK2, PINK1/PARK6, DJ-1/PARK7 and ATP13A2/PARK9) that account for only ~50% of PD cases with early-onset (EO) autosomal recessive (AR) inheritance. Hence, a significant proportion of inherited PD cases still remain unexplained genetically. AIM: We propose to use an integrative approach combining homozygosity mapping/genomic rearrangement detection and targeted exome sequencing in consanguineous PD families with an early age at disease onset (≤55 years) to identify new causative genes for recessive PD. PATIENTS AND METHODS: Homozygosity mapping was performed in a series of 160 PD consanguineous families or isolated patients originating from Europe (n=38), North Africa (n=57), Turkey (n=61) and other countries (n=4) excluded for PARK2, PINK1, DJ-1, and the common LRRK2 G2019S mutation, using SNP genotyping microarrays (Illumina Infinium HD HumanCytoSNP-12 BeadChip) and an original linkage statistics program that allows the inclusion of individuals for whom genealogical information is lacking. Using the same SNP microarrays, rare large deletions and duplications were searched for. To identify causative genes, whole exome sequencing using the Agilent SureSelect 50 Mb Human Exome target enrichment technology was performed in 60 EO PD families confirmed to be consanguineous by their Genomic inbreeding coefficient F values. RESULTS: Linkage analyses using the Genomically Controlled Homozygosity Mapping Statistic under heterogeneity (HLOD) and stratified according to geographical origin have revealed 5 linked regions (7p, 7q, 10p, 1p, 1q) across the genome with HLOD scores ≥ 2. Using the Illumina cnvPartition module, we did not detect any large rare rearrangements in the series of 160 consanguineous families. First analyses of the whole exomes of index cases have led to the identification of homozygous mutations in ATP13A2, FBX07. CONCLUSION: As a proof of concept, the strategy combining homozygosity mapping/ genomic rearrangement detection and exome sequencing using a unique series of EO PD consanguineous families has led to the identification of homozygous mutations in ATP13A2, FBX07 and probably more novel causative genes for recessive PD.

1374T

Rare variant discovery of progressive supranuclear palsy using whole-exome sequencing. *C-F. Lin[1,2], E.T. Geller[1], O. Valladares[1,2], Y-H. Hwang[2,3], L. Stutzbach[1], L. Cantwell[1], L-S. Wang[1,2], G.D. Schellenberg[1].* 1) Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Genomics and Computational Biology Graduate Group, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Progressive supranuclear palsy (PSP) is a rare neurodegenerative disorder that causes movement disability and cognitive decline, and hence is often misdiagnosed as Parkinson's disease or Alzheimer's disease. MAPT is a well known genetic factor for PSP with strong effect size. In 2011 a GWAS of 1114 cases and 3247 controls identified SNPs within three genes (STX6, EIF2AK3 and MOBP) with genome-wide significance. However, these common variants do not fully explain the known heritability of PSP. Whole-exome sequencing (WES) technologies provide a cost-effective approach to genotype more than 44Mb human exonic regions of the human genome and represents an exciting direction for rare variant discovery. We performed WES of 278 PSP patients using Illumina Hi-Seq 2000 (100nt pair-end) and Nimblegen SeqCap EZ Human Exome library v2 for target capture. Sequencing data were analyzed using the NIAGADS DRAW workflow, following the best practices suggested by the Broad Institute Genome Analysis Toolkit (GATK) site. All samples have at least 80% of the targeted regions at >20x and 90%; of the regions at >10x coverage. Overall we observed 165,768 exonic variants (including core splice sites), 37,731 (23%) of which were not reported in dbSNP 137; ~95,000 were not synonymous SNVs. On average each sample has 15371 (CI:15202-15541) single nuclear variations (SNVs) and small insertions/deletions. We next examined the variants in a set of 61 genes known to be related to neurodegenerative disorders. We found that a missense SNV in the GAK gene shows significant association: the minor allele is present in five out of 566 alleles (five out of the 278 PSP subjects). Four of the five heterozygotes were confirmed through Sanger sequencing. A fisher exact test with four out of 8598 alleles in the NHLBI GO Exome Sequencing Project (ESP) European American population results in a P-value of $7.7 \times 10{-}4$ (OR=15.6). We also observed an elevated number of loss of function variants in the FIG4 gene. Our initial analysis is encouraging as it suggests that rare variants in the exome may contribute to PSP risk. Our next step is to complete whole-exome sequencing of 750 PSP subjects, perform exome-wide association analysis as well as gene-wise burden tests using additional controls, and validate called rare variants using independent molecular methods such as Sanger sequencing or TaqMan PCR. We will report on the replication analyses of these preliminary findings in this presentation.

1375F

A recurrent microdeletion at 20p13 unmasks a recessive mutation in PLCB1 in a patient with severe infantile epileptic encephalopathy. *I.M. Wentzensen[1], A. Mctague[2], E. Meyer[2], C. Applegate[1], A. Ngoh[2], E. Kossoff[3,4], D. Batista[5,6], T. Wang[1,3].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Neurosciences Unit, University College of London, Institute of Child Health, London, UK; 3) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Kennedy Krieger Institute, Baltimore, MD.

Phospholipase C beta 1 (PLCB1) is a G protein-coupled phosphodiesterase which catalyzes the formation of second messenger molecules inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. PLCB1 plays an important role in several intracellular signaling pathways in the central nervous system. A recent report described a homozygous 20p13 microdeletion involving the promoter and first three exons of PLCB1 to cause infantile epilepsy in consanguineous families. We report a 10 month-old African American female with intractable epilepsy, severe global developmental delay, and regression. This infant was born at 37 weeks of gestation after a normal pregnancy and delivery. No significant dysmorphic features or congenital malformations were present on physical exam and imaging studies. Metabolic studies were unremarkable. Brain MRI showed mild cerebral volume reduction, mildly hypoplastic corpus callosum, age-appropriate myelination, and normal neuronal migration patterns. EEGs showed right temporal lobe sharp waves and spikes followed by generalization of epileptiform activities. A SNP microarray identified a heterozygous microdeletion of 476kb (8,094,442-8,580,722; GRCh37/hg19) on chromosome 20p13 involving the promoter region and first three exons of the PLCB1 gene. This deletion was inherited from her mother who has no history of seizures. Long-range PCR was performed to establish the exact breakpoints of the deletion. These are similar to the previously reported first cases of microdeletion at 20p13. Interestingly, these recurrent microdeletions are flanked by two long interspersed (LINE) elements suggesting nonallelic homologous recombination at this location. Sanger sequencing of all exons of PLCB1 identified a single base substitution (c.99+1G>A) which involves the conserved splice site of intron 1 in the proband and her unaffected father. These results support that loss-of-function mutations in PLCB1 cause a rare recessive from of early infantile epileptic encephalopathy and that flanking LINE elements may play a role in recurrent microdeletions in human disease.

1376W

A Follow-Up of Whole-Exome Sequencing in Multiplex Families. *A.H. Beecham[1], J.L. McCauley[1], A. Hadjixenofontos[1], P.L. Whitehead[1], W.F. Hulme[1], I. Konidari[1], S.L. Hauser[2], J.R. Oksenberg[2], J.M. Vance[1], J.L. Haines[3], M.A. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Department of Neurology, School of Medicine, University of California, San Francisco, CA, USA; 3) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, USA.

Multiple sclerosis (MS) is a common neurodegenerative disease, affecting more than 1.3 million individuals worldwide. Given the number of multiplex families found in MS, it is plausible that rare variants may contribute to MS in these families. We have used next-generation sequencing to scan multiplex families for rare variants. Filtering 127 individuals (91 affected and 36 unaffected) from 26 multiplex families, assuming a dominant model with incomplete penetrance, we identified 578 high quality coding (nonsense, missense, or splice), conserved, potentially damaging, and rare (MAF $\leq$ 0.01) variants in 550 genes segregating completely in all affected of at least one family. Three of the 578 variants are in confirmed MS genes including CLEC16A, EOMES, and SLC9A8. Assuming a recessive model, we identified 68 high quality coding, conserved, potentially damaging, and rare (homozygote frequency $\leq$ 0.05) variants in 61 genes segregating completely in at least one family. Using a gene focused approach; we found 20 genes having at least two different filtered variants segregating in at least two different families. Eight of the 20 genes showed significant burden ($p < 0.05$ from Fisher's exact test) in our families when compared to 141 non-diseased controls including: PNPLA6, PLXNA4, DNAH1, PDHA2, SAMD3, COL11A2, SMAD5, and HIVEP2. Lastly, we identified 510 rare (MAF $\leq$ 0.01) and 10 moderately rare (0.01 < MAF $\leq$ 0.05) high quality coding, conserved, potentially damaging variants in known MS genes, including the Major Histocompatibility Complex (chromosome 6 from 26-36 MB) that were present in at least one MS affected individual. We submitted 1,154 unique variants from our three filtering approaches (dominant, recessive, known MS genes) for inclusion in a custom Illumina Infinium HumanExome +BeadChip designed by the International Multiple Sclerosis Genetics Consortium. In total, 1082 of our nominated variants made it onto the chip after manufacturer quality control, including all priority variants from the eight genes demonstrating significant burden. We are currently genotyping 114 affected individuals from 51 additional multiplex families to screen for these variants and other rare variation in these genes included as part of the exome chip content.

1377T

Exome Sequencing for Schizophrenia and Alcohol Dependence in an Irish Cohort Using Low-cost Library Prep and Target Capture. *D.G. Brohawn[1], T. Hendon[1], T. Bigdeli[1], E. Loken[1], D. Walsh[2], F.A O' Neill[3], S. Bacanu[1], K. Kendler[1], B.T Webb[1], B. Riley[1].* 1) Virginia Commonwealth University, Richmond, VA; 2) Health Research Board, Dublin, Ireland; 3) Queens University, Belfast, Ireland.

Common and rare alleles both influence common disorder risk. Identifying specific common alleles and elucidating their function remain difficult due to the small effect sizes. Rare functional variation in relevant genes may have more direct impact on trait risk and may elucidate the role of variation in the gene on risk. However, these variants are relatively rare and large sample sizes are required to detect these effects. Library preparation, target capture, and sequencing remain costly for the sample sizes necessary. We adapted a recently published protocol for reduced cost library preparation and target capture by at least 30%; in order to efficiently collect exome data. DNA is quantified, sheared to a mean size of 170 bp, and indexed partial Illumina adapter sequences are added and target capture using the Agilent 71 Mb exome + UTR kit is performed on equimolar pools of 6 samples. Sequencing is performed on the Illumina HiSeq 2500 platform using 2×100 bp reads. Reads are aligned, base quality scores are generated, and variants are called using STAMPY and GATK, and analyzed using the SNP-set Kernel Association Test (SKAT) for burden and gene-based association. Variants with MAF >2%; will be tested individually. We sequence both cases of alcohol dependence (AD, N=500) and schizophrenia (SCH, N=250) to 50X average depth using this approach. Sample sizes are increased by imputing sequence variants into unsequenced samples with GWAS data. Variants from cases and the UK10K controls (N=2432) will be used to produce a single reference panel for imputing variants into additional unsequenced cases and controls (AD total N=4897, SCH total N=6478). We compared the standard and adapted library preparation, target capture, and sequencing protocols. The standard protocol involved single sample exome capture, while the adapted protocol involved pooled sample exome capture. Our baseline is the standard Truseq Illumina library prep, capture and sequencing of the 71 Mb exome + UTR target sequenced to 50X average cover of unique reads yielding 10x coverage for >93% of target. The baseline protocol had comparable QC metrics to initial runs of the pooled capture protocol for unique reads (87.2% baseline vs. 86.1% pooled) and % of bases on or near bait (89.2% baseline vs 91.5% pooled). We have completed prep for 73 additional samples and we anticipate that >100 exomes will be complete and analyzed by the time of presentation.

1378F

Reduced mRNA expression and aberrant intron DNA methylation of EGR2 in female schizophrenia patients. *M. Cheng[1], Y. Chuang[1], S. Hsu[1], C. Chen[2].* 1) Department of Psychiatry, Yuli Veterans Hospital, Hualien, Taiwan; 2) Department of Psychiatry, Chang Gung Memorial Hospital at Linkou and Chang Gung University School of Medicine, Taoyuan, Taiwan.

Background: Abnormal myelination has been considered as part of the pathophysiology of schizophrenia. Early growth response 2 (EGR2) has a specific function in the regulation of hindbrain development, myelination of peripheral nervous system, and the stabilization of long-term potentiation. And it has been considered as a candidate gene for schizophrenia. Methods: We compared the EGR2 mRNA levels in lymphoblastoid cell lines between 119 schizophrenic patients and 114 controls using real-time quantitative PCR. We attempted to detect genetic variants of the EGR2 gene in 500 patients with schizophrenia and 500 non-psychotic controls from Taiwan using direct sequencing and conducted a case-control association study. We measured the EGR2 mRNA in SH-SY5Y cells treated with a DNA methyltransferase inhibitor, 5-azacytidine, for 24 hours. We also compared DNA methylation status of EGR2 gene in the lymphocytes in between 61 patients with schizophrenia and 64 control subjects using pyrosequencing. Results: Female but not male schizophrenic patients had significantly reduced EGR2 mRNA levels compared to controls (p=0.024). We identified 6 known single nucleotide polymorphisms (SNPs) in this sample. However, SNP-based analysis showed no association of these SNPs with schizophrenia. We also detected 24 rare mutations in this sample. Notably, a c.518-519insTCC variant which causes a proline insertion was found in 2 unrelated patients, but not detected in 500 controls. EGR2 expression was up-regulated in human SH-SY5Y after being treated with 5-azacytidine. One CpG site of EGR2 intron region had significantly increased DNA methylation in peripheral blood cells of female schizophrenia compared to female controls. Conclusions: Our results suggest that rare variants of the EGR2 gene might contribute to the pathogenesis of schizophrenia in some patients, but further functional assays are needed to verify their relevance to the pathogenesis of schizophrenia. The increased DNA methylation in the EGR2 gene in female schizophrenia patients may partly account for the reduced EGR2 gene expression in female patients.

1379W

Association of rare variants with anorexia nervosa by whole exome sequencing. *J. Connolly[1], D. Li[1], H. Hakonarson[1, 2].* 1) Ctr Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

Anorexia nervosa (AN) is a perplexing illness characterized by low body weight and persistent fear of weight gain during period of growth, resulting in extreme emaciation. People with AN usually manifest with symptoms of depression, anxiety, and obsessive-compulsive behaviors that are common features in other neuropsychiatric disorders. Multiple bodies of evidence now suggest the role of genetic influences to AN. Family studies have consistently demonstrated that AN occurs in families and twin studies have revealed the contribution of additive genetic factors to the observed familial aggregation, but despite many studies conducted so far, the genetic architecture underlying AN susceptibility remains largely unknown. Here we have completed analysis of 10 extreme cases samples and multiple control samples with where an average coverage of 65X was established. A total of 133,955 single nucleotide variants (SNVs) and 70,678 small insertions/deletions (indels) were identified in the 10 case samples that underwent whole exome sequencing (WES), of which about 4% of the variants are novel. Two novel variants were identified in previous candidate genes that involved in the mood control and two in previous genome-wide association signals. Also, we performed association analysis using both internal and 1000 genomes data as control after combining several Caucasian populations. Due to excessive rare variants in WES results, burden test was applied. The total number of rare variants across a gene is tabulated in each individual and these totals are compared between cases and controls. The top association signals were detected in C7orf23, DMTF1, COL18A1, PCNT and ERMN. Interestingly, PCNT gene previously associated with an increased risk of developing psychiatric disorders such as schizophrenia and depression, was fourth significant (3.65E-03).

1380T

Detection of de novo mutations in schizophrenia patients and their unaffected siblings. *S. de Jong[1], N. Tran[2], E. Strengman[1], E-J. ten Dam[1], R.S. Kahn[3], R.A. Ophoff[1,3,4].* 1) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience & Human Behavior, University of California Los Angeles, Los Angeles, California, USA; 2) Informatics Center for Neurogenetics and Neurogenomics, University of California Los Angeles, Los Angeles, California, USA; 3) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, the Netherlands; 4) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA.

Schizophrenia is a common disease, affecting ~1% of the population with an estimated heritability of up tot 80%. Recent studies have shown that in addition to a polygenic model of disease, single de novo mutations may also contribute to disease susceptibility. In order to test whether there is an increased de novo mutation rate in patients compared to unaffected siblings, we applied whole exome sequencing in a family-based design. We included 19 families, consisting of proband with both parents, and at least one unaffected sibling. Whole exome capture was performed using Illumina TruSeq kits followed by paired end (100 bp reads) sequencing on the HiSeq2000 platform, resulting in ~30million readpairs per sample. Reads were aligned to hg19 using BWA. Single sample variant calling was performed with SAMtools. Concordance rates with SNP array for a subset of the data were ~98%. De novo variants were called per family using DeNovoGear and Polymutt. The unaffected sibling in each family served as an extra filtering step, excluding calls appearing in both siblings, which are likely to be false positives. DeNovoGear is a Bayesian algorithm that starts with an a priori probability of 1×10−8/bp/generation on haploid germline point mutation rate. Putative variants with a posterior probability of >0.98, mapping quality >20 and sequence read depth >30 across all family members were selected. Variants not overlapping between siblings and also called with Polymutt (~60%) were subsequently filtered for heterozygote genotypes in offspring. We completed the preliminary analyses using these criteria. Based on our results thus far the average predicted de novo variant is 7 per meiosis. The number of predicted de novo SNVs varied greatly by subject and family. Validation using Sanger sequencing needs to be performed on these predicted variants to exclude sequencing artifacts. The preliminary results suggest a slight overrepresentation of the overall de novo variants in coding regions in the affecteds. We did not, however, observe a difference in the distribution of nonsynonymous SNPs (damaging or benign) between probands and unaffected siblings. In both the unaffected and the affected group we find a double hit in the same gene. Our preliminary results indicate that although the number of de novo variants might not differ, the distribution of types of SNVs seems to be different between schizophrenia patients and their unaffected siblings.

1381F

Genome-wide profiling of DNA-protein interactions and multiple histone methylations in the cultured olfactory cells from Schizophrenia.
Z. Deng[1], M. Deng[2], O.V. Evgrafov[1], J.A. Knowles[1]. 1) Psychiatry & Behavioral Sci, Univ Southern California, Los Angeles, CA; 2) Undergraduate, Brown University, Providence, RI.

Schizophrenia (SCZ) is a strongly heritable complex disorder. The discovery of candidate biomarkers and genetic risk factors in schizophrenia has proven challenging. The great difficulty in the search of genes and biomarkers for schizophrenia can be ascribed to a combination of high disease heterogeneity, the absence of prominent and unique alteration, and the complex interplay of genetic predisposition and environmental influences. We hypothesize that epigenetic variations (DNA methylation and histone modifications), together with particular genetic variations (SNPs, insertion/deletions), play roles in the etiology of Schizophrenia. Furthermore, epigenetic modifications are responsive to environmental factors (including antipsychotic treatments) and could be a link between genetic predisposition and environmental influence on the development of schizophrenia. To reduce the complexity and limit genetic heterogeneity, we choose a subgroup of SCZ with olfactory deficits and focused on the epigenetic alteration using a genome-wide ChIP-seq of the neuronal precursor cell culture from SCZ nasal biopsy samples. We have performed seven different genome-wide epigenetic profiles (H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K9ac, CTCF and RNA pol II) in four SCZ patients and two controls. We observed more epigenetic heterogeneity among the SCZ samples, while the patterns were more consistent within the control samples. We find several abnormal epigenetic peaks (genome-wide FDR<0.01) are associated with some new schizophrenia loci identified in a mega-analysis combining genome-wide association study (GWAS) data from 17 separate studies. (Schizophrenia Psychiatric Genome-Wide Association Study Consortium, 2011). These epigenetic profiles have been subsequently analyzed together with genome-wide transcriptional profiles from the RNA-seq results of the cases and controls. We have compiled a list of candidate genes, biomarkers and chromosome regions identified in our epigenetic study and transcriptional study.

1382W

Examining De-novo mutation rates and patterns from whole-exome sequencing of sporadic cases of Schizophrenia from Taiwan.
D. Howrigan[1,2,7], B. Neale[1,2,7], J. Moran[2], K. Chambert[2], S. Rose[2], N. Laird[6], H.-G. Hwu[5], W.-J. Chen[5], C.-M. Liu[5], C.-C. Liu[5], J. Nemesh[2,7], E. Bevilacqua[2], A. Hansen[7], S.V. Faraone[3], S. Glatt[3], M. Tsuang[4], S. McCarroll[2,8]. 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute of Harvard and MIT, Cambridge, MA; 3) SUNY Upstate Medical University, Syracuse, NY; 4) University of California, San Diego, CA; 5) National Taiwan University, Taiwan; 6) Harvard School of Public Health, Boston, MA; 7) Harvard Medical School, Boston, MA; 8) Harvard University, Cambridge, MA.

Recent studies have implicated that exonic de-novo mutations play a convincing role in developmental disorders such as autism, intellectual disability, and epilepsy. For sporadic cases of schizophrenia, early reports have pointed towards promising de novo candidates; however larger numbers of sequenced trios are required to identify specific genes, as de novo events arise at a low rate in the population (roughly one per trio). Under the collaboration of multiple centers, whole-exome sequencing has been performed on 1,135 complete trios from a Taiwanese cohort, making it one the largest de novo sequencing projects to date. Exome sequencing data were generated using the Illumina HiSeq sequencing platform with the Agilent SureSelect exome capture platform, and the Illumina Miseq platform was chosen as the method for validating putative de novo calls. Preliminary results suggest that the overall rate of mutation in affected offspring falls in line with the expected mutation rate. Upon examination of high-quality putative de novo events (those with twenty or more reads in the full trio), multiple de-novo mutations have been found in thirty-nine separate genes, with multiple loss-of-function mutations found in SV2B, a synaptic vesicle glycoprotein gene primarily expressed in the brain. These findings, however, do not surpass exome-wide significance after incorporating gene size and site-specific mutation rates into expectations of de novo mutation. Further analyses examining de novo events in conserved gene sets show a 1.79-fold higher enrichment of loss-of-function events and a 1.19-fold higher enrichment of non-synonymous events. However, these levels only reach suggestive significance (p = 0.08 and 0.1, respectively). Overall, our findings do not clearly identify any single gene as an unequivocal risk factor for schizophrenia when disrupted by de novo mutation, although a number of genes and gene-set analyses show suggestive signals. We also integrate the results from these trios with the other available trio sequencing data from the community to provide the clearest evaluation to date of the role that de novo events play in the etiology of schizophrenia.

1383T

Deep re-sequencing of *TREM2* identifies novel rare variants associated with late-onset Alzheimer's disease.
S. Jin, T. Skorupa, B. Cooper, D. Carrell, J. Norton, B. Benitez, C. Cruchaga, A. Goate. Psychiatry, Washington University in St. Louis, ST. LOUIS, MO.

Recent studies have found heterozygous rare variants in *TREM2* (triggering receptor expressed on myeloid cells 2), which has been reported to be involved in inflammation in the brain, significantly affect risk for late-onset Alzheimer's disease (LOAD). Although several rare variants have been associated with LOAD risk, large scale deep re-sequencing studies have not been conducted to comprehensively identify novel rare variants in *TREM2* that affect LOAD. We performed pooled-DNA sequencing of all exons and flanking regions in *TREM2* in 2,074 LOAD cases and 1,388 cognitively normal elderly controls of European descent from the National Cell Repository for Alzheimer's disease, the Knight Alzheimer's Disease Research Center, and the Alzheimer's Disease Neuroimaging Initiative using Illumina Miseq sequencing. We used the SPLINTER algorithm to perform alignment and call variants in the pooled samples. SIFT2, SeattleSeq and Ensembl databases were used to annotate variants. We then confirmed rare variant calls in all sequenced individuals using Sequenom iPLEX or KASPar genotyping systems. Of the 122.8 million reads generated, 69.7 million (56.7%) mapped back to the reference genome with a 30-fold minimum coverage. Eight novel or potentially functional rare variants passed the filtering quality controls and were selected for direct genotyping. Six of the eight variants were validated (sensitivity=75%), four of which are novel. Five rare variants (4 missense and 1 stop-gained) were present only in cases and not in controls. Among them, four were predicted as probably-damaging by PolyPhen or SIFT2. We also observed one missense mutation present only in controls and predicted to be tolerated. When collapsed, these six rare variants in *TREM2* were associated with a significant increase in LOAD risk (OR=6.04; P=0.04) which suggests that some could be functional and have dramatic effects on LOAD risk. This study suggests deep re-sequencing is an effective and accurate approach to identify novel rare variants associated with LOAD. The identified heterozygous rare variants may cause a loss-of-function of *TREM2*, consistent with the notion that heterozygous *TREM2* mutations contribute to LOAD.

1384F

Mutations in the *CASR* gene in Idiopathic Generalized Epilepsy.
M. Kaur[1], P. Satishchandra[2], K. Radhakrishnan[3], A. Kapoor[1], A. Anand[1]. 1) Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, Karnataka, India; 2) Department of Neurology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India; 3) Department of Neurology, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India.

Idiopathic generalized epilepsy (IGE) is a common type of epilepsy with substantial genetic basis to its etiology. Genetic studies have identified 12 IGE loci, and at 6 of these, causative genes have been found. We conducted detailed sequence analysis of an IGE locus, EIG8 (MIM612899) which was identified in a family from south India with several of its members affected with epilepsy (Kapoor *et al*. *Ann Neurol* 2008). EIG8 maps at chromosome 3q13-q21. We used a combination of Sanger sequencing and next generation sequencing to analyze transcripts for 234 genes in the critical genomic region. New single nucleotide variants were found in the *NSUN3, EPHA6, ABI3BP, TRAT1, KIAA1407, IQCB1, CASR* and *ADCY5* genes. We studied segregation of the variants in the epilepsy family, examined their allele frequencies in the control individuals and conducted *in silico* pathogenicity analysis. These studies suggested role of c.2693G>A (R898Q) in *CASR* (extracellular calcium sensing receptor) in causation of epilepsy in the family. Additionally, we examined *CASR* in 480 IGE/JME patients and 252 control individuals, and detected 22 uncommon variants (m.a.f≤0.005). Of the variants identified, 7 were missense variants leading to E354A, D433H, S580N, I686V, R898Q, A988G, A988V. While *CASR* is known to play a role in maintenance of systemic calcium homeostasis, its role in neuronal cells remains unclear. *CASR* may play a role in regulating neuronal excitability in the human brain. We studied functional correlates of the missense variants identified, using a *trans*-reporting MAPK pathway assay and an IP-One ELISA-based assay in HEK293T cells, transiently transfected with the wild type and mutant receptor constructs. Three mutant receptors exhibited significantly enhanced Ca^{2+} responsiveness as compared to wild type CASR, whereas the remaining ones, showed relatively weak significant differences from the wild type protein. We propose CASR's membrane localization/trafficking or its affinity to calcium may be compromised by the variants.

1385W

Trio-based pathway analysis of bipolar disorder. *N. Matoba[1,2], M. Kataoka[1,3], K. Fujii[4], Y. Suzuki[2], S. Sugano[2], T. Kato[1].* 1) Lab. for Molecular Dynamics of Mental Disorder, RIKEN Brain Science Institute, Wako , Saitama, Japan; 2) Dept. of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan; 3) Dept. of Child Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Hongo, Tokyo, Japan; 4) Dept. of Psychiatry, Dokkyo Medical University, Mibu, Tochigi, Japan.

Bipolar disorder is one of the two major mental disorders and twin studies reported the heritability is around 85%(Cardno et al., 1999, McGuffin et al., 2003). Genome-Wide Association Studies (GWAS) identified a number of common SNPs that are associated with bipolar disorder, but no strong contribution has been shown. Next-Generation Sequencing (NGS) technology enables us to detect causes of rare genetic disorders. However, the way to identify causative genes of complex diseases like bipolar disorder is still under development. To identify molecular pathways related to bipolar disorder, we compared the deleterious mutations transmitted to, or untransmitted to the proband using whole exome sequencing of trio families of bipolar disorder. Fifty trio families were enrolled. Each individuals was interviewed by trained psychiatrists using a structured interview were selected. After target exome were captured from the whole-blood or saliva DNA of participants using the SureSelect Human All Exon V4(Agilent), whole-exome sequencing was carried out using Hiseq2000(Illumina). Reads were mapped to human genome build v37 with BWA and PCR duplicates were removed with Picards. Local re-alignment and variant calling was performed using GATK. All variants were annotated with ANNOVAR. 87.8% of target regions were covered with at least 20 unique reads and we only focused on these regions. The criteria for candidate variants were non-synonymous SNVs, not in dbSNP137(excluding MAF < 0.01 or reported as disease related) and predicted as 'disease causing' by SIFT and 'possibly damaging or probably damaging' by PolyPhen2. All variants observed in each family were divided into two groups, transmitted (found in a proband and one of the parents) or un-transmitted (found in a parent only) variants. In a preliminary analysis of 7 families, 42.86 variants were transmitted to proband and 49.29 variants were un-transmitted to the proband on average. Gene Ontology (GO) term GO:0012502 (induction of programmed cell death) was enriched in transmitted variants while muscle development related GO terms were enriched in un-transmitted variants. Enrichment of apoptosis related genes in transmitted mutations is compatible with the possible role of cellular valunevability in bipolar disorder. Further analysis using other 43 trio families with bipolar disorder and control trio families is ongoing.

1386T

The not-so-silent effect of silence mutations in Autism Spectrum Disorders. *O.A. Moreno-Ramos, M.C. Lattig, LICA (Liga Colombiana de Autismo).* Universidad de los Andes, Laboratorio de Genetica Humana, Bogotá, Colombia.

High locus and allelic heterogeneity found in Autism Spectrum Disorders (ASD) complicate the comprehension of the genetic bases of the disorders. Thousands of clues have emerged from initial exome sequencing studies, identifying many novel de novo mutations in ASD family trio studies. Affected proteins seem to be highly interconnected and expressed mainly in brain. Most of the variants reported to date were found mostly in Caucasian or European descendant cohorts but Latin American cohorts have not been well studied. Therefore, we decided to apply exome sequencing, at a 50X depth, in a cohort of Colombian - South American (admixed population) trios. Although most of exome sequencing studies focus on deleterious mutations such as non-synonymous, missense and frame shift mutations, there is still a big caveat: What about the not-so-silent effect of silence mutations might have? We not only focused on the discovery of harmful variants, but we also evaluated the possible effect of silent mutations, since it is known that synonymous mutations can actually be a cause of different diseases and syndromes altering mRNA stability or translation. The global outcomes guide to a larger range of mutations that are related to ASD.

1387F

The Rare *TREM2* Variant R47H Has a Modest Effect on Alzheimer's Disease Risk. *A.R. Parrado[1], K. Mullin[1], W.K. Yip[2], B. Hooli[1], T. Liu[3], C. Lange[2], L. Bertram[3], R.E. Tanzi[1].* 1) Neurology, Mass General Inst Neurodegenerative Disease, Boston, MA, USA; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 3) Dept. Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Recently, two independent research groups (Jonsson et al. 2013 and Guerreiro et al. 2013) published results reporting a rare missense variant rs75932628 (R47H) in exon 2 of the gene encoding triggering receptor expressed on myeloid cells 2 (*TREM2*) significantly increase the risk for Alzheimer's disease (AD) with an effect size comparable to that of the *APOE* ε4 allele. Subsequently, Pottier et al. (2013) and Benitez et al. (2013) published independent short communications confirming the association between the minor T-allele at rs75932628 and increased risk for AD. Here we attempt to replicate the association between rs75932628 and AD risk by directly genotyping rs75932628 in two independent Caucasian family cohorts, the National Institute of Mental Health Alzheimer Disease Genetics Initiative Family Study and the National Institute on Aging Genetics Initiative for Late Onset Alzheimer's Disease (NIA-LOAD) Family Study, consisting of 927 families (with 1777 affecteds and 1235 unaffecteds) and in the Caucasian NIA-LOAD case-control sample composed of 378 cases and 686 controls. Additionally, we imputed genotypes in three independent Caucasian case-control cohorts (GenADA, TGEN2, and ADNI), containing 1906 cases and 1503 controls. Meta-analysis of the two family-based and the four case-control cohorts yielded a P-value of 0.04 providing additional independent support for the association between the T-allele at rs75932628 and increased AD risk, albeit suggesting a much lower effect size (OR = 1.48, 95% CI 0.94-2.35). In conclusion, our results confirm the association between SNP rs75932628 and AD risk, however, the risk effect is substantially smaller than in the two original reports.

1388W

Sequencing of coding region of the CNTNAP2 in individuals with Autism Spectrum Disorders. *K.P. Pena Gonzalez, L.M. Munera Salazar, O.A. Moreno Ramos, M.C. Lattig, Liga Colombiana de Autismo.* Universidad de los Andes, Bogota, Colombia.

Autism Spectrum Disorders (ASD) are a group of developmental disorders that involve three specific developmental areas: disrupted communication, difficulties in social interaction and repetitive/stereotyped behavior. Even though its etiology is still unknown, evidence suggests a strong genetic component evidenced by twin studies that demonstrate approximately 80% concordance. However, only about 10-20% of non-syndromic ASD cases have been explained by deleterious mutations, but each genetic effect explains only up to 1% of the ASD cases. Contactin Associated Protein-Like 2 (CNTNAP2) [MIM: 604569], encodes a protein from the Neurexin family with various roles in the Central Nervous System. The aim of this study was to search for deleterious variations (no-synonymous, low frequency, rare or private) in the CNTNAP2 gene in ten individuals with ASD by sequencing the coding region of this gen using Sanger technique. We found 25 variations in the CNTNAP2 gene, but we focused on a particular haplotype (rs61232377, rs77025884, and rs9648691) that was found in 4 individuals containing the intron 22/exon 23 region. According to bioinformatics analysis this haplotype suggests a formation of a splicing branch site on exon 23. This could be possibly involved in the formation of new isoforms of the gene, and therefore might be related to ASD.

1389T

Exploring the genetic landscape of Parkinson's disease - an exome study in Sardinia, a Mediterranean genetic isolate. *M. Quadri[1], X. Yang[2], G. Cossu[3], S. Olgiati[1], V. Saddi[4], G. Breedveld[1], L. Ouyang[2], J. Hu[2], N. Xu[2], J. Graafland[1], V. Ricchi[3], D. Murgia[3], L. Correia Guedes[5], S. Tesei[6], M.J. Marti[7], P. Tarantino[8], S. Asselta[9], F. Valldeoriola[7], M. Gagliardi[8], G. Pezzoli[6], M. Ezquerra[7], A. Quattrone[10,11], J.J. Ferreira[5], G. Annesi[8], S. Goldwurm[6], E. Tolosa[7], B. Oostra[1], M. Melis[3], J. Wang[2,12,13,14], V. Bonifati[1].* 1) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 2) BGI-Shenzhen, Shenzhen, China; 3) Neurology Service and Stroke Unit, General Hospital S. Michele AOB "G. Brotzu", Cagliari, Italy; 4) Neurology Division, S. Francesco Hospital, ASL No. 3 Nuoro, Italy; 5) Clinical Pharmacology Unit, Instituto de Medicina Molecular, University of Lisbon, Portugal; 6) Parkinson Institute, Istituti Clinici di Perfezionamento, Milan, Italy; 7) Neurology Service, Hospital Clínic of Barcelona, Barcelona, Spain; 8) Institute of Neurological Sciences, National Research Council, Cosenza, Italy; 9) Department of Medical Biotechnologies and Translational Medicine, University of Milan, Milan, Italy; 10) Institute of Neurology, University Magna Græcia, Catanzaro, Italy; 11) Neuroimaging Research Unit, Institute of Neurological Sciences, National Research Council, Germaneto, Italy; 12) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 13) King Abdulaziz University, Jeddah, Saudi Arabia; 14) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen,Copenhagen, Denmark.

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease, with a prevalence of >1% over age 65. The etiologic landscape of PD is complex. Rare, highly-penetrant mutations in different genes and common risk factors of small size-effects in several loci have been identified in patients with PD. However, additional disease determinants remain to be identified in most patients. Genetically isolated populations offer advantages for dissecting the genetic architecture of complex disorders. Here we report the results of our genetic study on the etiology of PD in Sardinia. We performed exome sequencing in 100 unrelated Sardinian PD patients, using the Agilent SureSelect 50Mb kit and Illumina HiSeq2000 sequencing, at an average coverage of ~30X. We first removed all synonymous, intronic, and intergenic variants. From the remaining variants, we selected the novel SNPs shared by at least five unrelated PD patients, and absent in dbSNP129 and 1000Genomes databases. This approach yielded a total of 4,587 SNPs, that were then genotyped in 500 independent Sardinian individuals (242 PD and 258 controls) using a NimbleGen SeqCap EZ custom platform. Association of each variant with disease status was then tested using Fisher's exact test implemented in PLINK/SEQ. Out of 186 variants with p-value <0.5 and odd-ratio >3, thirty-two were confirmed by Sanger sequencing. These 32 SNPs were then genotyped by TaqMan assays in an independent sample of 2,731 PD patients and 2,673 age- and ethnic-matched controls from Italy, Spain and Portugal. No variants surpassed the required level of significance according to Bonferroni correction for 4,587 tests (p value $< 1.09 \times 10^{-5}$). The lowest p value was 1,16 $\times 10^{-3}$ for a variant with final OR 2,9. However, a catalogue of interesting coding variants present only in PD patients was identified, that might point to novel genetic determinants of PD with moderate/strong effect size, and represents an important resource for replication in follow-up studies. Our study suggests that, with regard to the inspected exome target region, the genetic bases of PD are highly heterogeneous.

1390F

Genetic epidemiological study on migraine and episodic ataxia in the Japanese population. *Y. Takahashi, H. Ishiura, J. Goto, S. Tsuji.* Dept Neurology, Grad Sch Med, Univ Tokyo, Tokyo, Japan.

[Background] Migraine is an episodic disorder characterized by severe throbbing headache attacks, which is classified as migraine with aura (MA) and without aura (MO). Genetic background of migraine has been largely unknown, except for familial hemiplegic migraine (FHM). Causative genes for FHM include *CACNA1A*, which is also causative for episodic ataxia (EA) type 2, *ATP1A2* and *SCN1A*. Recently, mutations in *PRRT2*, initially identified as a causative gene for paroxysmal kinesigenic dyskinesia, have been reported in FHM patients. [Purpose] To investigate the genetic epidemiology of migraine and episodic ataxia in the Japanese population. [Subjects and Methods] DNA samples from 11 FHM, 2 sporadic hemiplegic migraine (SHM), 9 MA, 13 MO and 7 EA patients and 260 healthy individuals were included in this study. Mutational analysis for *CACNA1A* and *ATP1A2* were conducted employing a custom-designed resequencing microarray TKYMGN01. Those samples in which mutations in these genes were excluded were further subjected to mutational analysis of *SCN1A* and *PRRT2* employing a direct nucleotide sequence analysis. Novel nonsynonymous variants identified were further screened in 260 controls. [Results] Three of 11 FHM patients (27.3%) harbored causative mutations, including a previously reported mutation p.T666M in *CACNA1A* in 2 patients and a novel mutation p.H916L in *ATP1A2* in a patient. In addition, a known causative mutation c. 649dupC in *PRRT2* was identified in a SHM patient. No mutations were found in patients with MA, MO or EA. Common clinical features of FHM patients with the mutations included relatively long-duration of attacks, disturbance of consciousness, fever, and cerebral edema during hemiplegic attacks. [Discussion and Conclusion] This study indicated that a substantial number of patients with hemiplegic migraine in the Japanese population were accounted for by mutations in known causative genes including *CACNA1A*, *ATP1A2* and *PRRT2*. In addition, this study have raised the possibility that p.T666M in *CACNA1A* might be a relatively frequent mutation in FHM in the Japanese population, as supported by the finding that the two of three previous reports with mutations in *CACNA1A* in the Japanese population represented the p.T666M mutation. Genetic epidemiology of MA, MO or EA, in contrast, have not been accounted for by these known causative genes for FHM, warranting further investigation for genes associated with these common types of migraine.

1391W

Whole transcriptome analysis by next generation sequencing in autism spectrum disorders. *C. Zusi[1], P. Prandini[1], G. Malerba[1], L. Xumerle[1], R. Galavotti[1], A. Pasquali[1], C. Patuzzo[1], V. Mijatovic[1], R. Ciccone[2], M. Fichera[3], MC. Bonaglia[4], E. Trabetti[1], PF. Pignatti[1].* 1) University of Verona, Verona, Verona, Italy; 2) University of Pavia, Pavia, Italy; 3) Scientific Institute Oasi Maria Santissima, Troina, Italy; 4) Scientific Institute E. Medea, Bosisio Parini, Italy.

Autism Spectrum Disorders (ASDs) represent a group of childhood neurodevelopmental and neuropsychiatric disorders characterized by deficits in verbal communication, impairment of social interaction, and restricted and repetitive patterns of interests and behaviours. Evidences indicate that ASDs have strong genetic bases. This study is part of a Telethon project involving several Italian research groups; it aims to analyze differences of gene expression level between 27 ASD subjects and 23 control individuals. Nineteen (19) of the 27 ASD subjects are characterized for CNVs potentially involved in the onset of autism and 8 subjects have a deletion in the 22q13.3qter region. EBV transformed lymphoblastoid cell lines have been established for all subjects and transcriptomes have been analyzed through Next Generation Sequencing technology (RNA Sequencing). Differential expression analysis, performed on the whole cohort and the subgroup with the 22qter deletion, identified 295 and 448 differentially expressed genes respectively, at the nominal p-value <0.05. Gene set enrichment analysis (GSEA) revealed that autoimmune disorders and antigen processing and presentation terms are the most enriched ones. Subgroup's GSEA highlights the involvement of axon guidance term, confirming that LCLs could exhibit biomarkers relevant to autism. Further analysis was carried out to identify genes strongly dysregulated in ASD subjects, finding a total of 68 genes. Three dysregulated genes, that cluster within a CNV on chromosome 16p13.1, were identified in one subject. The remaining 65 dysregulated genes did not map on any considered CNV. Additional analysis on differentially expressed genes will help understanding the genetic bases of ASD pathophysiology and unravelling potential new pathways involved in the disease.

1392T

Investigating the association between OPA1-AS1 gene variants and autosomal dominant optic atrophy. *I. Nakata, K. Linkroum, W. Abdrabou, M. Janessian, E. Gaier, D.M. Cestari, S. Lessell, J. Rizzo, J.L. Wiggs.* The Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Boston, MA.

Purpose: Autosomal-dominant optic atrophy (DOA) is one of the most common inherited optic neuropathies causing progressive bilateral degeneration of the optic nerve and ultimately irreversible blindness. Mutations in *OPA1*, encoding a dynamin-like mitochondrial GTPase indispensable for mitochondrial network structure and morphology, are identified to be responsible for DOA in approximately 50% of cases. Genetic abnormalities responsible for the remaining 50% of cases are not yet known. *OPA1* antisense RNA 1 (*OPA1-AS1*) is a long non-coding mRNA located between intron 5 and 7 of the *OPA1* gene that may function to regulate the expression of *OPA1*. To investigate the association between *OPA1-AS1* variants and DOA, we performed a sequencing-based association study. Methods: The study was approved by the Massachusetts Eye and Ear Infirmary Institutional Review Board and adheres to the tenets set forth in the Declaration of Helsinki. Fifty-nine DOA patients from 53 independent families were sequenced in this study. Disease causing *OPA1* mutations were previously detected in 17 of these 59 patients. Genomic DNA was sequenced using primers designed to amplify all 3 *OPA1-AS1* exons. PCR products were sequenced according to standard protocols. Genotypes of known SNPs in the general population were retrieved from the HapMap CEU data sets (n = 174). Results: Four *OPA1-AS1* SNPs were observed in cases: rs9291059, rs34307082, rs3772393, and rs9832709. Of the 59 DOA patients, 2 patients were heterozygous for rs9291059 [minor allele frequency (MAF) = 0.018] and one patient was heterozygous for rs34307082 (MAF = 0.008). Significant linkage disequilibrium was observed between common SNPs rs3772393 and rs9832709 (D' = 1.0 and r^2 = 1.0). We found that the MAF of *OPA1-AS1* rs3772393 was significantly higher in the DOA cases than that in general population controls (0.53 versus 0.42, P = 0.0400, odds ratio = 1.77, 95% confidential interval = 0.84-3.76). Of the 16 patients homozygous for the rare variant of rs3772393, 7 (43.8%) have a disease-causing *OPA1* mutation, although none of these were located in the *OPA1-AS1* exons and are not in linkage disequilibrium with rs3772393. Novel variants in *OPA1-AS1* were not identified by sequencing. Conclusions: These data suggest that *OPA1-AS1* variants may associate with developing DOA and that additional study of the potential role of *OPA1-AS1* in DOA is warranted.

1393F

Exome Sequencing on Samples with Bipolar Disorder: A Preliminary Survey. *T. Zhang[1], F.J. McMahon[2], D.T. Chen[2], J.C. Wang[1], J.P. Rice[1].* 1) Dept. of Psychiatry, Washington University in St. Louis, St. Louis, MO; 2) Genetic Basis of Mood and Anxiety Disorders Unit, National Institute of Mental Health, Rockville, MD.

Bipolar disorder is a psychiatric diagnosis for a mood disorder with lifetime prevalence of about 1%. Previous genetic studies have linked many chromosomal regions and candidate genes appearing to influence the risk of bipolar disorder. In this research, we aim to identify the potential causal variants for bipolar disorder via performing exome sequencing on 6 related case samples from a four-generation family. The sequencing data were released for 195,949 variants, and 20,094 variants (10.25%) were shared among all these six related cases. Genotype data annotation was performed by Annovar and by implementing several standard filtering criteria we have reduced the number of targeted variants to 9,657 variants. To narrow the susceptible regions, we have genotyped 733,202 common SNPs for all these six cases and performed linkage analysis. A preliminary low-density linkage analysis has identified three suggestive peak region in chromosomes 4, 8 and 10 (with LOD score greater than 2.0). We are now performing high-density/ haplotype-based linkage analysis to further narrow down the candidate regions and identify the potential susceptible variants for bipolar disorder in this family.

1394W

Exploration of brain-specific somatic mutations with massively parallel sequencing data derived from human postmortem brain. *M. Nishioka[1,2], M. Bundo[1], J. Ueda[3], S. Murayama[4], K. Kasai[2], T. Kato[3], K. Iwamoto[1].* 1) Department of Molecular Psychiatry, The University of Tokyo, Tokyo, Japan; 2) Department of Neuropsychiatry, The University of Tokyo, Tokyo, Japan; 3) Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Saitama, Japan; 4) Department of Neuropathology, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan.

Accumulating evidence has challenged the conventional assumption that somatic cells in one individual, except for rare cases, have the identical genome. Several groups have reported somatic single nucleotide variations (SNVs) and structural variations (SVs) in normal cells other than oncogenic cells in humans and model animals. Especially, brain cells are supposed to go through characteristic genomic alterations such as retrotransposition and chromosomal aneuploidy during development. We hypothesized that such somatic genomic alterations in brain cells contribute to biological diversity in cognitive and mental traits, and that some somatic mutations are involved in the pathophysiology of neuropsychiatric disorders. We explored brain-specific somatic SNVs and SVs with exome sequencing data derived from human postmortem brain and liver with no neuropsychiatric disorders as a preliminary study. The samples are fresh-frozen postmortem frontal cortex and liver derived from 75-year-old Japanese male. We obtained neuronal genomic DNA by anti-NeuN antibody-based FACS technique. The exome sequencing of neuronal and liver genomic DNA was performed with Illumina GAIIx. The total reads were approximately 100 million reads per each sample, with more than ×100 coverage. We analyzed somatic mutation events using the generated read data with SNV/SV detection software such as RetroSeq, Delly, and SVDetect. We found that sequencing data derived from neuronal nuclei indicated much higher rate of retrotransposition and LTR-related genomic alterations. Although such genomic alterations might be results of sequencing or informatics error, some would be results of somatic mutations occupying very low fraction of sample tissues. We discuss the physiological and pathophysiological meanings of brain-specific somatic mutations and the validity of detection methods, mainly focusing on the results of transposable elements.

1395T

Identification of New Genes and Pathways for Rare Infantile Forms of Spinal Muscular Atrophy and Arthrogryposis. *L. Baumbach-Reardon[1], J.M. Hunter[1], M.E. Ahearn[1], J. Kiefer[2], C. Balak[1], G. Lambert[1], J. Getz[1], D. Duggan[1], B. Wirth[6], W. Tembe[3], C. Legendre[3], W. Liang[5], A. Kurdoglu[4], J. Corneveaux[4], M. Huentelman[4], D. Craig[4], J. Carpten[1], J. Hall[7].* 1) Cancer Genomics, Translational Genomics Res Inst, Phoenix, AZ; 2) Knowledge Mining, Translational Genomics Res Inst, Phoenix, AZ; 3) Center for Bioinformatics, Translational Genomics Res Inst, Phoenix, AZ; 4) Neurogenomics, Translational Genomics Res Inst, Phoenix, AZ; 5) Collaborative Sequencing Center, Translational Genomics Res Inst, Phoenix, AZ; 6) Institute of Human Genetics, University Hospital of Cologne, Germany; 7) University of British Columbia, Vancouver, BC, Canada.

Since discovery of the SMN gene in 1992, many rare clinical variants of SMA have been reported, including a subset associated with AMC. The genetic basis of these disorders remains unknown and continues to present important diagnostic challenges in the newborn period. In 2008, we uncovered the genetic cause of one of these rare lethal, infantile forms {X-linked SMA (SMAX2, OMIM 3018300)} through identification of the first human mutations in UBA1. We continue to evaluate families and affected male cases (AM) with XL-SMA-like phenotypes with the goal of discovering the genetic bases of these diseases. For the past year, our approach has been targeted sequencing of UBA1, followed by whole exome sequencing of UBA1-mutation negative cases. We have sequenced all UBA1 coding regions in all affected individuals in our cohort using a custom designed assay on Ion Torrent. As expected, UBA1 sequencing revealed common variants in each individual, but surprisingly no new disease associated variants were identified. We performed whole exome sequencing of 37 samples (15 families including 15 AM) on Illumina HiSeq 2000/2500. Alignment and extensive variant analysis was performed using custom in-house pipelines. Selected variants of interest were validated by qPCR Taqman assays. Exome sequencing revealed probable pathogenic mutations in many cases; we highlight results from three of these. First, a family with two affected boys revealed novel compound heterozygous deleterious CHRND mutations (OMIM 100720) that cause a form of lethal congenital myasthenic syndrome. We validated that these mutations co-segregate with disease in this family. Second, a novel start loss M1V mutation in SCML2 (OMIM 300208) was detected in the proband and mother of a family with an X-linked history of fetal and neonatal deaths. SCML2 encodes a component of the polycomb transcriptional repressor complex, and has not been previously associated with human any disease. Third, in a singleton affected boy, we identified a rare homozygous mutation in the RNA exosome component, EXOSC3, that is known to cause Pontocerebellar hypoplasia, type 1B. In retrospect, his phenotype is consistent with this diagnosis. This study demonstrates the importance of exome sequencing in identification of causal mutations and clarifiying previously undiagnosed cases of neonatal hypotonia ±AMC, and opens the door to further understanding the underlying genetic complexities of these often phenotypically-related disorders.

1396F

Genetics of spinocerebellar ataxias in Portuguese families: screening for SCA15, SCA28 and SCA36. *J.R. Loureiro[1], A.I. Seixas[1], J.L. Loureiro[1,2], A. Carracedo[3,4], M.J. Sobrido[3], P. Coutinho[1], J. Sequeiros[1,5], I. Silveira[1].* 1) IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Portugal; 2) Serviço de Neurologia, Centro Hospitalar entre Douro e Vouga, Portugal; 3) Fundación Pública Galega de Medicina Xenómica-SERGAS, Santiago de Compostela, Spain; 4) Genomic Medicine Group, School of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain; 5) ICBAS, Universidade do Porto, Portugal.

The spinocerebellar ataxias are rare neurodegenerative diseases clinically and genetically very heterogeneous. To this moment, 32 autosomal dominant SCAs have been genetically identified and 24 causative genes implicated in these pathologies. Trinucleotide repeat expansions are the cause of nine SCAs (SCA1, 2, 3, 6, 7, 8, 12, 17 and DRPLA). Intronic pentanucleotide repeats, an ATTCT expanded tract or an insertion, are also causative of these pathologies (SCA10 and 31). Besides repeats, classical mutations are the origin of at least 12 of these diseases (SCA5, 11, 13, 14, 15, 18, 19/22, 23, 26, 27, 28, 35). More recently, intronic GGCCTG hexanucleotide expansions have been found in SCA36. Mutation screening in known SCA genes allowed us characterize approximately 200 Portuguese families. Machado-Joseph disease/SCA3 is the most frequent (52%) followed by DRPLA (5%) and SCA2 (2%), whereas the remaining showed very low frequencies or were not found. Thus, more than 100 Portuguese SCA families remain without a molecular diagnosis. Here we present the results of mutation screenings for SCA15, SCA28 and the recently described SCA36 in these Portuguese families. All previously reported SCA15 patients have deletions of several exons of the ITPR1gene; SCA28 is caused by missense mutations or small in/dels in the AFG3L2 gene; GGCCTG hexanucleotide expansions, ranging from 650-2500 repeats, in intron 1 of NOP56, cause SCA36, the most common type of SCA in Galicia, the Spanish region on the border with Northern Portugal. We carried out quantitative real-time PCR, direct sequencing of exons of interest or repeated-primed PCR to screen for ITPR1 genomic deletions, AFG3L2 mutations and NOP56 expansions, respectively, and failed to find pathogenic alterations. In conclusion, SCA15, SCA28 and SCA36 are very rare among Portuguese SCA families and most probably a considerable number of SCA genotypes remain to be identified. Funding: FCT (Project PTDC/SAU-GMG/098305/2008) and COMPETE with co-funding by FEDER; Instituto de Salud Carlos III-FIS PI12/00742.

1397W

Repeat interruptions in spinocerebellar ataxia type 10 expansions are strongly associated with epileptic seizures. *K.N. McFarland[1], J. Liu[1], I. Landrian[1], S. Raskin[2], M. Moscovich[1,3], E.M. Gatto[4,5], H.A.G. Teive[3], A. Ochoa[6], A. Rasmussen[7], T. Ashizawa[1].* 1) Dept of Neurology and The McKnight Brain Insitute, University of Florida, Gainesville, FL; 2) Core for Advanced Molecular Investigation, Graduate Program in Health Sciences, Center for Biological and Health Sciences, Pontifical Catholic, University of Paraná, Curitiba, Paraná, Brazil; 3) Movement Disorders Unit, Neurology Service, Hospital de Clínicas, Federal University of Paraná, Centro, Curitiba, PR 80060-150, Brazil; 4) Departamento de Neurología, Sanatorio de la Trinidad Mitre, Buenos Aires, Argentina; 5) Instituto de Neurosciencias Buenos Aires, INEBA, 1428 Buenos Aires, Argentina; 6) Department of Neurogenetics, Instituto Nacional de Neurologia y Neurocirugia Manuel Velasco Suarez, Mexico City, DF, Mexico; 7) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK.

Spinocerebellar ataxia type 10 (SCA10), an autosomal dominant neurodegenerative disorder, is the result of a non-coding, pentanucleotide repeat expansion within intron 9 of the Ataxin 10 gene. SCA10 patients present with pure cerebellar ataxia; yet, some families also have a high incidence of epilepsy. SCA10 expansions containing penta- and heptanucleotide interruption motifs, termed 'ATCCT interruptions,' experience large contractions during germline transmission, particularly in paternal lineages. At the same time, these alleles confer an earlier age at onset which contradicts traditional rules of genetic anticipation in repeat expansions. Previously, ATCCT interruptions have been associated with a higher prevalence of epileptic seizures in one Mexican-American SCA10 family. In a large cohort of SCA10 families, we analyzed whether ATCCT interruptions confers a greater risk for developing seizures in these families. Notably, we find that the presence of repeat interruptions within the SCA10 expansion confers a 6.80-fold increase in the risk of an SCA10 patient developing epilepsy and a 12.92-fold increase in having a positive family history of epilepsy. We conclude that the presence of repeat interruptions in SCA10 repeat expansion indicates a significant risk for the epilepsy phenotype and should be considered during genetic counseling.

1398T

Androgen receptor gene polymorphism is associated with impulsivity in women with alcoholism. *A.M. Manzardo, D. Mettman, E.C. Penick, A.B. Poje, M.G. Butler.* Psychiatry and Behavioral Sciences, Kansas University Medical Center, Kansas City, KS.

The androgen receptor (AR) gene, located on the X chromosome, contains a common polymorphism involving CAG repeats in exon 1 which impacts disease and could contribute to the unequal gender ratio in alcoholism. CAG repeats in the AR gene influence androgen sensitivity and are known to correlate with impulsivity in males. CAG repeat lengths over 35 produce a progressive neurodegenerative disorder (Kennedy disease) in men. We report the first preliminary study examining the association between the number of CAG repeats and measures of impulsivity in females with chronic alcoholism. Thirty-five women and 85 men with chronic alcoholism were previously recruited for a nutritional clinical trial. Twenty-six well-characterized females (19 African-American and 7 Caucasian) and 55 males (41 African-American and 14 Caucasian) with alcoholism agreed to participate for genetic testing. Genomic DNA was isolated from peripheral blood and CAG repeats determined by analyzing PCR amplified products using the polymorphic AR gene assay. The CAG repeat length for males and average CAG repeat length (CAGave) of both X chromosomes for females was correlated with raw scores from the Barratt Impulsivity Scale, version 11 and the Alcoholism Severity Scale. CAGave repeat lengths for Caucasian females [mean (SD) =18.1 (1.5)] were significantly longer than seen in African-American females [mean (SD)= 16.5(1.65); F=5.3, $p<0.03$], but CAG repeat lengths in men did not differ by race [mean(SD)=15.4(2.8)]. The average number of CAG repeats were significantly positively correlated ($p<0.05$) with impulsivity scores in females but not in males. The strongest relationship was observed for Caucasian females who showed a significant correlation for 4 first order factors (Motor Impulsiveness, Self-Control, Cognitive Complexity and Perseverance) and 2 second order factors (Motor Impulsiveness and Non planning Impulsiveness). Women with an average CAG repeat length of $\geq$18 representing the upper quartile of the repeat range showed significantly greater mean raw impulsivity scores. The CAG repeat length appeared to have less effect in African-American compared with Caucasian women possibly due to a shorter average repeat length. The results suggest that increasing CAG repeat length in females may contribute to psychopathology in alcoholism. Replicative studies with more females with alcoholism of both races are warranted.

1399F

Analysis of KDM5C defects associated with ARX Epilepsy-related mutants and evaluation of rescue strategies. *L. Poeta[1], A. Padula[1], F. Fusco[1], C. Shoubridge[2], G. Manganelli[1], S. Filosa[1], P. Collombat[3], G. Friocourt[4], M. Passafaro[5], K. Helin[6], L. Altucci[1,7], S. Gustincich[8], J. Gecz[2], M.V. Ursini[1], M.G. Miano[1].* 1) Institute of Genetics and Biophysics 'Adriano Buzzati Traverso', CNR, Naples, Italy; 2) Department of Paediatrics, University of Adelaide, South Australia, Australia; 3) Inserm U1091 Diabetes Genetics Team, Nice, France; 4) Inserm U1078 Laboratoire de Génétique Moléculaire et de Génétique Epidémiologique, Brest, France; 5) Institute of Neuroscience, CNR, Milan, Italy; 6) Centre for Epigenetics, University of Copenhagen, Copenhagen, Denmark; 7) Second University of Naples, Naples, Italy; 8) SISSA, Area of Neuroscience, Trieste, Italy.

Malignant Epilepsy linked to Aristaless-related homeobox (*ARX*) mutations presents severe pharmaco-resistant paediatric seizure. ARX is a crucial transcription factor regulating proliferation and migration of GABA-interneurons. We have recently identified a crucial epileptogenesis path, linking functionally ARX to another XLID/Epilepsy gene, Lysine-specific demethylase 5C (*KDM5C*). It encodes an H3K4me2/3 demethylase, which functioning is mediated by interactions with REST/NSRF, a master epigenetic hub critical for neuronal differentiation. We found that ARX Epilepsy-related mutations, which fall in PolyAlanine tracts or in the HD domain, cause a spectrum of functional damages of the ARX-KDM5C interaction, which severity depends on the type of alteration. In *Arx* KO embryonic brain and ES-oriented GABAergic neurons, a defective ARX-KDM5C-H3K4me3 path has been found in association with a mis-regulation of KDM5C Epilepsy-disease targets. *Arx* shRNA injections have been performed in mature neurons and endogenous ARX-KDM5C levels, spine density and morphology have been analyzed. We tested in vitro correction of KDM5C-H3K4me3 defects by exploiting three approaches: transcription factor targeting, gene knock-up and epigenetic modifications. To upregulate the endogenous KDM5C/KDM5C content, in a locus-specific manner, we tested *KDM5C* transcriptional induction by PHF8/ZNF711 stimulation and KDM5C translational increase by SINEUP method, aiming to balance the spectrum of KDM5C defects associated with the hypomorphic ARX PolyAlanine mutants. To achieve KDM5C-H3K4me3 path correction, we also screened a number of compounds targeting chromatin enzymes. We used, as cell disease model, neuronally-differentiated *Arx* KO/*Kdm5c*-depleted ES cells, which show GABAergic abnormalities in association with a global increase of H3K4me3 signal. A strong compensation of *Kdm5C*/KDM5C downregulation has been obtained at crucial time-point of neuronal commitment. Although many other ARX targets could have important roles in the XLID/epilepsy phenotype, we believe that restoring or upregulating the expression of KDM5C gene or protein, ideally through endogenous physiological mechanisms, should accelerate the identification of alternative therapy to cure *ARX* Epilepsy phenotypes and many other neuropathologies with malignant seizure.

1400W

Validation of a robust PCR-only assay for quantifying *FMR1* CGG repeats on clinical samples. *J.K. Moore[1], M.J. Basehore[2], G. Filippov[1], K. Adler[1], M.J. Friez[2], M. Schermer[1].* 1) Molecular Diagnostics, PerkinElmer, Inc, Waltham, MA; 2) Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, SC.

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability and is caused by an expansion of a CGG repeat in the 5' end of the *FMR1* gene. An accurate measurement of the repeat size is part of the classification and diagnosis of FXS and other Fragile X-associated disorders. To date, sizing of *FMR1* trinucleotide repeats in the clinical laboratory has required the use of capillary sequencer instruments to size PCR fragments, or the manual preparation and labor intensive measurement of the number of CGG repeats by Southern blot. Our goal is to validate and establish a simple and robust PCR-only assay for quantification of CGG repeat alleles. **Methods**: A blinded set of 200 archived clinical DNA samples were analyzed by PCR amplification using a FragilEase™ kit (PerkinElmer). The repeat number was calculated from the fragment size measured on a low-cost capillary electrophoresis 2100 Bioanalyzer instrument (Agilent Technologies). The repeat numbers of those samples were interpreted by the use of standards with known repeats in each run. The sample set included both male and female individuals with normal (n=167), intermediate (n=10), premutation (n=11), and full mutation alleles (n=12). Calculated repeat sizes were compared to genotypes originally obtained by the current gold standard of PCR with fragment sizing on a capillary sequencer followed by Southern analysis of no-result PCR samples or single allele females. **Results**: All DNA samples were amplified successfully. The twelve full-mutation samples were successfully amplified with the largest allele size measured at over 653 repeats. The allele classifications from the PCR-only assay were concordant with those obtained with the reference method except for one intermediate sample that was classified as normal by the FragilEase™ assay. The interassay imprecision of the repeat size measurement was within 1-2 CGG repeats for a majority of the samples. The turnaround time is within 6 hours and can handle 10 samples in a single run. **Conclusion**: This new PCR-only method is capable of classifying all different FXS allele types correctly. Our PCR-only method allows Fragile X testing to be performed in a broader spectrum of clinical laboratories and is a fast, robust method for diagnosis and potentially for population screening of the Fragile X syndrome.

1401T

AGG interruptions affect the risk of having a child with fragile X syndrome: a follow-up study. *C. Yrigollen[1], B. Durbin-Johnson[2], R. Hagerman[3, 4], L. Zhou[5], E. Berry-Kravis[5], F. Tassone[1, 3].* 1) Biochemistry and Molecular Medicine, University of California Davis, Sacramento, CA; 2) Public Health Sciences, University of California Davis, Davis, CA; 3) MIND Institute, University of California Davis, Sacramento, CA; 4) Pediatrics, University of California Davis, Sacramento, CA; 5) Neurological Science, RUSH University, Chicago, Il.

AGG interruptions in the Fragile X Mental Retardation 1 *(FMR1)* CGG trinucleotide repeat increase the stability of this repetitive locus during transmission. We have previously reported the presence of AGG interruptions within an *FMR1* premutation allele reduces the risk of the allele expanding to a full mutation during maternal transmission particularly in the 60-90 CGG repeat range. It has also been shown that the instability of alleles of 45-69 CGG repeats is less frequent and of a lesser magnitude when AGGs are present. The number of AGGs present inversely correlates with frequency of size instability, frequency of expansion to a full mutation, and magnitude of change in allele size during instability. Our follow-up study includes a larger clinical sample size comprised of 267 mothers previously reported, and 158 newly acquired mothers. The participants of this study were recruited through clinical research studies. Both this study, and that previously published, demonstrate the risk of having a child with fragile X syndrome (FXS) inversely correlates with the number of AGGs that are present the maternal allele particularly in the 60 - 90 CGG repeat size range. Thus, we have refined our previously published risk model using 425 mothers and 584 transmissions. We compared the distribution of AGG interruption patterns in premutation mothers from this clinical cohort to premutation mothers of newborns identified through our pilot study of newborn screening (NBS) for FXS. The number of transmissions is insufficient to determine if the risk model changes depending on observations from samples recruited through clinical settings and those in an unbiased general population collection. We have also determined the AGG patterns within the premutation alleles of 47 newborns identified during our pilot study. Preliminary data indicates no significant differences in the percent of premutation alleles containing 0, 1 and 2 AGGs in subjects seen clinically compared to subjects in the NBS cohort, for any given size range. Although the number of newborns is quite low it appears that less alleles with zero AGGs, are present in the unselected NBS sample relative to alleles from full mutation families. These results are expected as the population alleles from the NBS cohort, unlike those in the clinical cohort, are not biased toward alleles that have expanded to a full mutation in the family already. Larger sample size is needed to fully evaluate these relationships.

1402F

Applying NGS to familial ALS cohorts to identify novel genes. *K.L. Williams[1], J.A. Fifita[1], G.A. Nicholson[1,2], I.P. Blair[1].* 1) Australian School of Advanced Medicine, Macquarie University, Macquarie University, NSW, Australia; 2) Molecular Medicine Laboratory, Concord Hospital, New South Wales, Australia.

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive neurodegenerative disorder. Familial ALS accounts for ~10% of ALS cases with the remainder being sporadic. ALS is genetically heterogeneous, with mutations in known genes explaining 60% of familial cases in Australia. We aim to combine genetic linkage analysis, exome sequencing, high-throughput genotyping and bioinformatic strategies to identify novel causative ALS genes in Australian ALS families. The 66 ALS families used in this study are categorised into Types 1, 2 and 3 based on sample availability. All families are negative for mutations in *SOD1*, *TARDBP*, *FUS* and the hexanucleotide repeat expansion in *C9ORF72*. Exome capture and sequencing (Illumina TruSeq and HiSeq2000) was performed on 79 individuals from these families. Type 1 families (n=3, sample n=9) contain DNA samples from multiple family members including affected, unaffected, obligate carriers, and at-risk individuals. To identify potential ALS loci an 8cM genome-wide microsatellite linkage scan was performed on 70 individuals from these families. Resultant linkage data was coupled with filtered exome data to identify 1, 2 and 4 variants respectively from the Type 1 families. These variants are currently being validated in extended patient and control cohorts, patient tissues, and by *in vitro* and *in vivo* functional studies. Type 2 families (n=3, sample n=10) have DNA samples from either two affected individuals plus a 'married-in' control or >2 affected individuals. Following exome sequencing and filtering, there were 18, 28 and 37 remaining potential causative variants across the Type 2 families. To reduce the number of variants, we are performing high-throughput iPLEX genotyping (Sequenom MassARRAY) in 700 age-matched population controls. Each potential candidate gene was also screened through exome data from family Type 3 index cases (n=60) to search for additional variants in these genes. Scripts are being developed to identify related individuals among family Type 3 index cases. If related individuals are identified they will be combined to create a Type 2 family and reduce the number of potential causative variants. We are implementing new strategies utilising NGS to identify novel ALS genes across three different ALS family types. Identifying novel ALS genes will give insights into the biological basis of motor neuron degeneration, allow development of new disease models and provide new targets for therapeutic development.

1403W

Founder Mutation for Huntington Disease in Caucasus Jews. *H.N. Baris[1, 2], O. Melamed[1], C. Bram[3], N. Magal[1], E. Pras[2, 4], H. Reznik-Wolf[4], Z.U. Borochowitz[5], B. Davidov[1], R. Mor-Cohen[2, 6], D.M. Behar[1,7].* 1) The Raphael Recanati Genetic Institute, Rabin Medical Center, Beilinson Hospital, Petah Tikva, Israel; 2) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Van Leer Institute, Jerusalem and University of Florida, Gainesville, FL, USA; 4) The Danek Gartner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer, Israel; 5) The Simon Winter Institute for Human Genetics, Bnai-Zion Medical Center, Rappaport Faculty of Medicine and Research Institute, Technion-Israeli Institute of Technology, Haifa, Israel; 6) 6Amalia Biron Research Institute of Thrombosis and Hemostasis, Sheba Medical Center, Tel-Hashomer, Israel; 7) Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa, Israel.

Huntington disease (HD), an autosomal dominant disorder involving HTT, is characterized by a triad of chorea, psychiatric illness and cognitive decline. Diagnosis and age of onset depend on the degree of expansion of the trinucleotide CAG repeat within the gene. The prevalence of HD is known for Europeans but has not been studied in the Israeli population. Between 2006 and 2011 we diagnosed in our adult genetics clinic 11 HD patients from 10 different families. Nine of the 10 probands were Caucasus Jews (CJ) (Azerbaijani), and one was Ashkenazi Jewish. We performed haplotype analysis to look for evidence of a founder mutation, and found that of the nine CJ, eight shared the same haplotype that was compatible with the A1 haplogroup. We calculated the coalescence age of the mutation to be between 80 and 150 years. Ninety percent of our HD patients are CJ, although the CJ comprise only 1.4% of the Israeli population. Our findings suggest a higher prevalence of HD among CJ compared to the general Israeli population and are consistent with a recent founder mutation in this group. This should raise the index of suspicion for HD in CJ, thereby allowing for earlier diagnosis so that prenatal testing can be offered when desired, and treatment once it is available.

1404T

TDP43 interact with Pur α and rescue the neurodegeneration caused by Expanded Hexanucleotide GGGGCC Repeat. *J. Li[1,2], M. Poidevin[1], W. Thomas[1,3,4], P. Jin[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 2) Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, People's Republic of China; 3) Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA; 4) Department of Veterans Affairs Medical Center, Atlanta, GA, USA.

Recently, an expansion of GGGGCC repeats in the first intron of C9orf72 was found to be a common cause of both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Previously, we found that Pur α interact with GGGGCC hexanucleotide repeat in vitro and in vivo in a sequence-specific fashion that is conserved between mammals and Drosophila. Here we have extended this work to investigate the role of TAR DNA-binding protein (TDP-43), the most common pathological inclusion in ALS/FTD, in riboGGGGCC (rGGGCC) repeat toxicity. To do this, we used Neuro2a and Drosophila models of the expanded hexanucleotide repeat to show that the overexpression of the TDP43 is sufficient to rescue neurodegeneration. Next, we delineated the interaction between Pur α and TDP43. We show that Pur α interacts with TDP43 in an RNA independent manner that is dependent on the N-terminus of TDP43. Finally, we performed a Drosophila RNAi screen using TDP43 interacting protein to identify additional genetic interactors of Pur α / TDP43 complex. Our results suggest that the expanded rGGGGCC repeat may interact with TDP43 indirectly through Pur α, causing the sequestration of TDP43 and others genetic interacting proteins from their normal function. This mechanism could potentially play a role in the pathogenesis of ALS/FTD illness.

1405F

A predictive model for the *MAOA* promoter VNTR and its application in public GWAS datasets. *T. Wang[1], A. Lu[2], V. Rao[1], S. deJong[1], R.M. Cantor[1,2], R.A. Ophoff[1,2,3].* 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, California; 2) Department of Human Genetics, University of California Los Angeles, Los Angeles, California; 3) Department of Psychiatry, UMC Utrecht, The Netherlands.

Monoamine oxidase (MAO) A and B are adjacent genes on chromosome X that encode the main enzymes for neurotransmitter turnover. Many drugs used as treatment for clinical depression and anxiety inhibit the MAO enzymes, this suggests that they have an important role in psychiatric traits. Both *MAOA* and *MAOB* share a common promoter where a 30bp variable number tandem repeat (VNTR) was identified 1.2kb upstream the transcription start site of the *MAOA*. These alleles exist in the 2, 3, 3.5, 4, and 5 repeats (R), where 3R and 4R are the most common within the population. 2R, 3R, and 5R were shown to possess low activation of *MAOA*, whereas 3.5R and 4R were shown to have much higher levels of activation in in vitro studies. However, the reported in vitro promoter high/low activity has poor correlation with in vivo studies examining *MAOA* or *MAOB* expression in post-mortem brain. Furthermore, these in vivo studies are underpowered to detect significant associations. We developed a predictive tool for the *MAOA* promoter VNTR based on an advanced machine learning method, multicategory vertex discriminating analysis (VDA). We PCR genotyped the VNTR in 400 individuals of European descent, including 30 HapMap trios where there was SNP genotype information within 2MB of the *MAOA* promoter region. 300 individuals were used as a training dataset to build a VDA model usingonly overlapping SNPs in all datasets and searched by 10-fold cross validation to optimize the accuracy rate for predicting MAOA promoter VNTR. A test dataset of 87 individuals was used to validate the VDA model. This algorithm highlighted 5 SNPs, along with gender, that correctly classified the VNTR in 97% of our training dataset, and 96.7% in our validation dataset. With the predictor tool we are now able to study the *MAOA* promoter VNTR in available GWAS data sets of neuropsychiatric traits such as major depressive disorder, bipolar disorder and schizophrenia. The predictor tool enables us to establish whether some alleles are indeed high and low expressing (as reported in previous in vitro studies) in large gene expression data sets of human brain. Lastly, analysis is underway to test whether specific *MAOA* VNTR alleles affect monoamine neurotransmitter turnover by examining cerebrospinal fluid (CSF) monoamine metabolite levels in a large cohort of healthy controls.

1406W

MSH3 Polymorphisms and Protein Levels Affect CAG Repeat Instability in Huntington's Disease Mice. *S. Tome[1], K. Manley[3], J. Simard[2, 5], G.W. Clark[6, 7], M.S. Slean[2, 5], M. Swami[4], P.F. Shelbourne[4], E.R.M. Tillier[6, 7], D.G. Monckton[4], A. Messer[3], C.E. Pearson[2, 5].* 1) inserm, Paris, France; 2) Genetics and Genome Biology, The Hospital for Sick Children, TMDT Building 101 College St., 15th Floor, Room 15-312 East Tower, Toronto, ON, M5G 1L7; 3) Wadsworth Center, New York State Dept. of Health, & Department of Biomedical Sciences, University at Albany, SUNY, Albany, NY 12208, USA; 4) Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK; 5) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 6) Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada; 7) Campbell Family Institute for Cancer Research, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada.

Expansions of trinucleotide CAG/CTG repeats in somatic tissues are thought to contribute to ongoing disease progression through an affected individual with Huntington's disease or myotonic dystrophy. Broad ranges of repeat instability arise between individuals with expanded repeats, suggesting the existence of modifiers of repeat instability. Mice with expanded CAG/CTG repeats show variable levels of instability depending upon mouse strain. However, to date the genetic modifiers underlying these differences have not been identified. We show that in liver and striatum, the R6/1 Huntington's disease (HD) (CAG)~100 transgene, when present in a congenic C57BL/6J (B6) background, incurred expansion-biased repeat mutations, whereas the repeat was stable in a congenic BALB/cByJ (CBy) background. Reciprocal congenic mice revealed the *Msh3* gene as the determinant for the differences in repeat instability. Expansion-bias was observed in congenic mice homozygous for the B6 Msh3 gene on a CBy background, while the CAG tract was stabilized in congenics homozygous for the CBy Msh3 gene on a B6 background. The CAG stabilization was as dramatic as genetic deficiency of Msh2. The B6 and CBy Msh3 genes had identical promoters but differed in coding regions, and showed strikingly different protein levels. B6 MSH3 variant protein is highly expressed and associated with CAG expansions, while the CBy MSH3 variant protein is expressed at barely detectable levels, associating with CAG stability. The DHFR protein, which is divergently transcribed from a promoter shared by the *Msh3* gene, did not show varied levels between mouse strains. Thus, naturally occurring MSH3 protein polymorphisms are modifiers of CAG repeat instability, likely through variable MSH3 protein stability. Since evidence supports that somatic CAG instability is a modifier and predictor of disease, our data are consistent with the hypothesis that variable levels of CAG instability associated with polymorphisms of DNA repair genes may have prognostic implications for various repeat-associated diseases.

1407T

***Mlh1* and *Mlh3* modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches.** *R. Mouro Pinto[1], E. Dragileva[1], A. Kirby[2,3], A. Lloret[1], E. Lopez[1], J. St. Claire[1], G.B. Panigrahi[4,5], C. Hou[6], K. Holloway[7], T. Gillis[1], J.R. Guide[1], P.E. Cohen[7], G-M. Li[6], C.E. Pearson[4,5], M.J. Daly[2,3], V.C. Wheeler[1].* 1) Center for Human Genetic Research, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, United States of America; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America; 4) Program of Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 6) Graduate Center for Toxicology and Markey Cancer Center, University of Kentucky College of Medicine, Lexington, Kentucky, United States of America; 7) Department of Biomedical Sciences, Cornell University, Ithaca, New York, United States of America.

The Huntington's disease gene (*HTT*) CAG repeat mutation undergoes somatic expansion that correlates with pathogenesis. Modifiers of somatic expansion may therefore provide routes for therapies targeting the underlying mutation. Huntington's disease Hdh^{Q111} mice exhibit higher levels of somatic *HTT* CAG expansion on a C57BL/6 genetic background (B6.Hdh^{Q111}) than on a 129 background (129.Hdh^{Q111}). Linkage mapping in (B6x129).Hdh^{Q111} F2 intercross animals identified a single quantitative trait locus underlying the strain-specific difference in expansion in the striatum, implicating mismatch repair (MMR) gene *Mlh1* as the most likely candidate modifier. Crossing B6.Hdh^{Q111} mice onto an *Mlh1* null background demonstrated that *Mlh1* is essential for somatic CAG expansions and that it is an enhancer of the *HTT* CAG pathogenic process in striatal neurons. Hdh^{Q111} somatic expansion was also abolished in mice deficient in the *Mlh3* gene, implicating MutLγ (MLH1/MLH3) complex as a key driver of somatic expansion. Strikingly, the *Mlh1* and *Mlh3* genes encoding MMR effector proteins were as critical to somatic expansion as *Msh2* and *Msh3* genes that encode DNA mismatch binding proteins. The *Mlh1* locus is highly polymorphic between B6 and 129 strains. While we were unable to detect any difference in base-base mismatch or short slipped-repeat repair activity between B6 and 129 MLH1 variants, repair efficiency was MLH1 dose-dependent. MLH1 mRNA and protein levels were significantly decreased in 129 mice compared to B6 mice, consistent with a dose-sensitive MLH1-dependent DNA repair mechanism underlying the somatic expansion difference between these strains. Together, these data identify *Mlh1* and *Mlh3* as novel critical genetic modifiers of *HTT* CAG instability, point to *Mlh1* genetic variation as the likely source of the instability difference in B6 and 129 strains and suggest that MLH1 protein levels play an important role in driving of the efficiency of somatic expansions.

1408F

Small Molecule Ligand Distorts RNA G-Quadruplex Structure of the Disease-Associated r(GGGGCC)n Repeat of the *C9orf72* Gene and Blocks Interaction of RNA-Binding Proteins. *C.E. Pearson[1, 2], B. Zamiri[3], K. Reddy[1, 3], R. Macgregor[3].* 1) Genetics & Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Program of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A1; 3) Graduate Department of Pharmaceutical Sciences, Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Ontario, Canada M5S 3M2.

Certain DNA and RNA sequences can form G-quadruplexes, which can affect genetic instability, promoter activity, RNA splicing, RNA stability, and neurite mRNA localization. Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) can be caused by expansion of a (GGGGCC)n repeat in the *C9orf72* gene. Mutant r(GGGGCC)n-containing transcripts aggregate in nuclear foci possibly sequestering repeat-binding proteins like hnRNPA1 - suggesting a toxic-RNA pathogenesis, as occurs in myotonic dystrophy (DM1) and fragile X associated tremor ataxia (FXTAS). Furthermore, the *C9orf72* repeat RNA and the FXTAS repeat RNA were recently demonstrated to undergo the non-canonical repeat associated non-AUG translation (RAN-translation) into pathologic peptide repeats in patient brains, a process that is thought to depend upon RNA structure. RAN-translation was initially observed in tissues from DM1 and spinocerebellar ataxia type 8 patients. We previously demonstrated that the *C9orf72* r(GGGGCC)n RNA forms repeat tract length-dependent G-quadruplex structures that can be bound by the splicing factor ASF/SF2. Here we show that a small molecule ligand, which can bind some G-quadruplex forming sequences, can bind and distort the G-quadruplex of the r(GGGGCC)8 RNA, and this ablates interaction of either hnRNPA1 or ASF/SF2. These findings provide proof-of-concept that nucleic acid binding small molecules can distort the secondary structure of the *C9orf72* repeat, which may beneficially disrupt protein interactions, which may ablate either protein-sequestration and/or RAN-translation into potentially toxic dipeptides. Disruption of secondary structure of the *C9orf72* RNA repeats may be a viable therapeutic avenue for various diseases, as well as a means to test the role of RNA structure upon RAN-translation.

1409W

Haplotype analyses in 15 Brazilian and Peruvian families with Spinocerebellar Ataxia type 10 (SCA10). *M.L. Saraiva-Pereira[1,2,3,13,14], T.C. Gheno[1,2,3], G.V. Furtado[1,2,3,14], J.A.M. Saute[3], K.C. Donis[3], A.M.V. Fontanari[1,3], V.E. Emmel[1,2,3], J.L. Pedroso[4], O. Barsottini[4], C. Godeiro[5], H. van der Linden[6], E.T. Pereira[7], W. Marques-Junior[8], R.M. Castilhos[3,14], I. Alonso[9], J. Sequeiros[9], M. Cornejo-Olivas[10], P. Mazzetti[10], V.L. Torman[11], L.B. Jardim[1,2,3,12,14], Rede Neurogenetica.* 1) Lab. de Identificacao Genetica-Centro de Pesquisas, Hospital de Clinicas de Porto Alegre, Porto Alegre, RS, Brazil; 2) Programa de Pós-Graduação em Genética e Biologia Molecular, UFRGS, Porto Alegre, Brazil; 3) Serviço de Genética Médica, HCPA, Porto Alegre, Brazil; 4) Universidade Federal de São Paulo, São Paulo, Brazil; 5) Universidade Federal do Rio Grande do Norte, Natal, Brazil; 6) Centro de Reabilitação Dr. Henrique Santillo, Goiânia, Brazil; 7) Universidade Federal de Santa Catarina, Florianópolis, Brazil; 8) Universidade de São Paulo, Ribeirão Preto, Brazil; 9) UnIGENe, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal; 10) Centro de Investigación en Neurogenética, Instituto Nacional de Ciencias Neurologicas, Lima, Peru; 11) Departamento de Estatística, UFRGS, Porto Alegre, Brazil; 12) Departamento de Medicina Interna, UFRGS, Porto Alegre, Brazil; 13) Departamento de Bioquímica, UFRGS, Porto Alegre, Brazil; 14) Instituto de Genética Médica Populacional, INAGEMP.

Spinocerebellar ataxia type 10 (SCA10) is a neurodegenerative disease characterized by cerebellar ataxia and seizures, due to expansion of a pentanucleotide repeat (ATTCT) in the ATXN10 gene. To date, SCA10 has only been reported in patients from Latin-American countries. The numerous families of Amerindian descent with SCA10 mutations and the absence of this variant in European countries suggest a founder effect for this mutation. The objective of this study was to identify a previously described haplotype, shared by 3 SCA10 Brazilian families, in additional patients from Brazil and new cases from Peru. Repeat primed-PCR was performed to detect expanded SCA10 alleles in Brazilian and Peruvian individuals with a SCA of unknown cause. Haplotypes were constructed, based on polymorphic markers within and outside the gene. We have identified 23 confirmed SCA10 cases: 3 families have been described previously (Alonso et al, 2006; Almeida et al, 2009), while 12 families were newly identified, within a large cohort of patients with ataxia. From those, 3 families were from Peru, while the remaining 9 were from Brazil. The same polymorphic markers and haplotypes were typed in 100 individuals (200 alleles) from the general local population. A different haplotype distribution was found in SCA10 patients and the control group. A common haplotype, 19CGGC14, was found in 11/13 of Brazilian and in 1/3 of Peruvian families. Haplotype 19CGGC16 was found in one Peruvian and one Brazilian families; the remaining two families showed other haplotypes. Individuals with SCA10 mutations have been previously described in Brazil, whereas this is the first report of SCA10 in Peru. We confirmed the presence of a consistently recurrent intragenic haplotype, suggesting a common ancestry for most, if not all patients. Supported by: CNPq, FIPE-HCPA, FAPERGS, INAGEMP, RIBERMOV.

1410T

Detailed audiological evaluation of a patient with xeroderma pigmentosum with neural degeneration. *D. Mercer[1], F. Tsien[2], A. Hurley[1].* 1) Department of Communication Disorders, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Department of Genetics, Stanley S. Scott Cancer Center, Louisiana State University Health Sciences Center, New Orleans, LA.

Xeroderma pigmentosum (XP) is a rare autosomal recessive condition characterized by defects in DNA excision repair, leading to extreme sensitivity to ultraviolet light damage and increased susceptibility to melanomas and carcinomas. Approximately 25% of XP patients also exhibit progressive neural degeneration which includes mental deterioration, cortical thinning, and sensorineural hearing loss. We describe a patient with XP complementation subtype D. The patient initially presented in our clinic at 7 years of age after failing a hearing screening. At that time, he was found to have a unilateral mild high-frequency sensorineural hearing loss with a configuration suggestive of noise exposure. He also demonstrated central auditory processing deficits. He was re-evaluated at age 11 years, and his hearing loss had progressed to a bilateral moderate sensorineural hearing loss. At age 13 years, the patient acquired type 1 diabetes. Genetic confirmation of XP was obtained at 18 years of age. XP testing revealed compound heterozygous missense mutations in complementation group D, one of the XP subtypes associated with neural degeneration and sensorineural hearing loss. Serial audiograms and electrophysiological testing have been obtained on this patient over a course of 11 years. His hearing loss has gradually progressed to a bilateral precipitous sloping sensorineural hearing loss that is mild in the low frequencies and severe-to-profound in the high frequencies. Auditory evoked potentials showed deteriorating waveform morphology of the auditory brainstem response and late evoked potentials (P300). The change in P300 recordings may be due to cortical thinning, which has been reported in XP with neural degeneration. In addition to neural degeneration and peripheral neuropathy, the patient has been treated for claw-toe deformity. These findings will further characterize the phenotype of XP type D.

1411F

Characterizing the aging human brain transcriptome. *N. Pochet[1,2,3], D. Borges-Rivera[3], B. Haas[3], S. Rajagopal[1,2,3], J. Xu[1,2,3], C. McCabe[1,2,3], O. Gevaert[4], G. Srivastava[1,2,3], A. Regev[3], J. Schneider[5], D. Bennett[5], P. De Jager[1,2,3].* 1) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital; 2) Harvard Medical School; 3) Broad Institute of MIT and Harvard; 4) Cancer Center for Systems Biology, Department of Radiology, Stanford University; 5) Rush Alzheimer's Disease Center, Rush University Medical Center.

Aging in humans is often accompanied by cognitive decline which, in many cases, leads to dementia and a syndromic diagnosis of Alzheimer's disease (AD). Our goal is to characterize the aging human brain transcriptome in order to gain mechanistic and functional insights into the neurobiological processes implicated in aging-related cognitive decline and Alzheimer's disease.

We sequenced RNA from the dorsolateral prefrontal cortex (DLPFC) of 550 individuals from two prospective cohort studies of aging that include brain donation at the time of death: the religious order study and the memory and aging project. The subjects are all non-demented at the start of the study and, at the time of death, differ in the extent of aging-related cognitive decline, ranging from cognitively non-impaired individuals to mildly impaired individuals, and to individuals clinically diagnosed with Alzheimer's disease at the time of death. In addition to the transcriptome data, we have genotype data on each individual and DNA methylation data from the same region of the brain in each subject, which inform our RNA analyses.

We applied *ab initio* and *de novo* transcriptome assembly approaches to systematically study the aging human brain transcriptome for differentially expressed genes and isoforms, novel transcripts and (*trans*-)splice isoforms, as well as mutations and RNA editing events in expressed genes. We selected differentially expressed genes and isoforms, by relying on the known annotated genes and isoforms, and we tested these for enrichment in known categories. We identified novel transcripts and (*trans*-)splice isoforms, by comparison to known annotated genes and measures of support at the read, read-pair and expression levels. We detected mutations and RNA editing events, by comparing the transcriptome to the reference genome and genotyping data. Finally, we enhanced the Module Networks algorithm to associate the (epi)genetic changes we found with changes in mRNA levels and (*trans*-)splicing variants, and to compare clinical characteristics between individuals stratified by the combination of aberration and expression/splicing changes.

1412W

Mucolipidosis type IV: progressive gliosis, synaptic dysfunction and cognitive deficits in Mcoln1 knock-out mouse. *Y. Grishchuk[1], S. Sri[2], W. Ma[1], N. Rudinskiy[3], M. Cottle[1], K. Stember[1], E. Sapp[3], A. Muzikansky[4], M. Difiglia[3], R. Betensky[4], B. Hyman[3], R. Kelleher[1], J. Cooper[2], S. Slaugenhaupt[1].* 1) Center for Human Genetics Research, Massachusetts General Hospital/Harvard Medical School, Boston, MA; 2) The Institute of Psychiatry, King's College London, London, UK; 3) MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; 4) Harvard School of Public Health, Harvard Medical School, Boston, Massachusetts, USA.

Mucolipidosis type IV (MLIV) is an autosomal-recessive neurodegenerative disease causing severe cognitive impairment, motor decline and progressive loss of vision. It is caused by mutations in MCOLN1 gene, which encodes a lysosomal/late endosomal non-selective cation channel TRPML1. The mechanism of how loss of TRPML1 leads to severe CNS pathology remains currently unknown and there is no therapy. Therefore, we are performing neuropathological examinations of an MLIV mouse model (Mcoln1 knock-out mouse), starting from very early stages and following disease progression. Using different experimental approaches such us behavioral testing, in vivo imaging, electrophysiology, histochemistry and electron microscopy, we describe structural and functional abnormalities associated with Mcoln1 loss in the mouse brain. Our results revealed that the first signs of motor and cognitive decline are seen at 2 months of age. These early behavioral phenotypes were accompanied by activation of astrocytes and microglia, as well as severe dysmyelination. Estimates of neuron numbers in the regions of the brain with the most astrocytosis and microglyosis revealed no neuronal loss at 2 and 3 months of age, indicating that activation of glia precedes neuronal death. Evaluation of resting Ca2+ activity in the 3 month-old Mcoln1-/- brains in vivo showed no difference between knockouts and littermate controls, implying normal calcium homeostasis at this age. Evaluation of the synaptic function in the Schaeffer collateral pathway revealed elevated paired-pulse facilitation (PPF) and long-term potentiation (LTP) in Mcoln1-/- hippocampi. Changes in LTP were accompanied with impaired learning and memory in the fear conditioning test. Overall, observed changes in synaptic transmission were not a result of neuronal loss in either the CA3 or CA1 fields of the hippocampus, as determined by stereologic counts. However, in the CA1 stratum radiatum we observed changes in synaptic morphology in excitatory synapses, reduced thickness of myelin sheaths and accumulation of abundant aggregates of inclusion bodies. Altogether, here we report for the first time synaptic dysfunction and cognitive deficits associated with TRPML1 loss which could explain the clinical manifestations seen in MLIV patients; and the early involvement of glial cells in MLIV disease pathophysiology, suggesting a new avenue towards the development of therapies for this devastating disease.

1413T

Identifying novel candidate genes for neurological diseases. *K. Summers.* Roslin Inst, Univ Edinburgh, Roslin, Midlothian, United Kingdom.

Over 80% of human genes are expressed in brain. Microdeletion syndromes (where groups of contiguous genes are deleted) almost always involve some form of intellectual disability and suggest that at least one in ten genes is critical for brain development. Within the mammalian genome there are a large number of genes that are not fully annotated and whose function is unknown. To identify such novel genes which are specific to the nervous system, the tool Biolayout *Express*^{3D} was employed to cluster sets of genes with shared expression patterns across mouse neurological cell types. This approach is model-free, since clusters are generated independent of known pathways or functions, and thus has the potential to find new pathway members. Published gene expression microarray data using the mouse MOE-430 platform (Affymetrix) were drawn from GEO-Datasets (NCBI). Tissues and cell types analysed included a range of brain regions, spinal cord, glial cells and non-neuronal cells. Biolayout *Express*^{3D} produced 187 coexpression clusters of more than 10 nodes (microarray probes). There were 16 clusters of genes (1027 probes representing 833 genes in total) expressed highly across the nervous system. Of these, 195 probes had no annotation and 151 had minimal annotation. Thus there was little information for one third of genes showing expression in brain and other neurological tissues. These genes included putative synaptic proteins, transcription factors of unknown target and other genes with no homology to any known protein domain. Knowledge of the other genes in the same cluster provides a functional annotation and identifies novel candidates for diseases affecting this function. Cluster 3 contained *Eno2*, the gene for neuron specific enolase, indicating that this is a neuron-specific cluster, and genes for a number of known synaptic proteins. There were also 96 genes of minimal annotation. For example, this cluster contained *Dbc1* (Deleted in Bladder Cancer 1; OMIM602865), thought to be a tumour suppressor, and now implicated in neuronal/synaptic function. Another gene in this cluster is *Susd4* (Sushi-Domain Containing 4; no OMIM entry), which was discovered by genome sequencing and is deleted in cases of autism and Fryns syndrome. Further analysis using knock-out mouse models, recombinant protein analysis and RNA interference will validate the role of the novel genes as candidates for diseases of the brain and nervous system.

1414F

The genetic architecture of schizophrenia: How do CNVs and polygenic scores contribute to disease risk? *S.E. Bergen[1], J.W. Smoller[2], J. Sebat[3], S. Purcell[4], B. Neale[5], K.S. Kendler[6], ISC, Swedish Schizophrenia Consortium, Psychiatric Genomics Consortium CNV and Schizophrenia groups.* 1) Medical Epidemiology and Biosataistics, Karolinska Institute, Stockholm, Sweden; 2) Psychiatric and Neurodevelopmental Unit, Massachusetts General Hospital, Boston, MA; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA; 4) Mount Sinai School of Medicine, New York, NY; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 6) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.

Both rare CNVs and common SNPs contribute to the genetic risk for schizophrenia. Several specific CNV regions and an increased burden of large deletion CNVs have demonstrated associations with schizophrenia. A significantly higher polygenic risk score in subjects with schizophrenia has also been established, confirming that many common SNPs confer risk for this disorder as well. The relationships between these rare and common genetic risk factors have not been thoroughly investigated, and we sought to address the following questions: 1) Do cases with CNVs have lower polygenic risk scores compared to cases without CNVs? 2) Do cases with CNVs have higher polygenic risk scores compared to controls with CNVs? 3) Do controls with CNVs have lower polygenic risk scores than controls without CNVs?

We investigated the polygenic risk score differences within and between case and control groups by CNV carrier status using the Swedish Schizophrenia Consortium (N=4646) as the discovery sample to score the International Schizophrenia Consortium (ISC) (N=4921) subjects. Analyses will be extended to CNV and GWAS data from the Psychiatric Genomics Consortium. CNV carriers were defined by the two classes of CNVs conferring the greatest disease risks: 1) having one of 12 specific CNVs previously associated with schizophrenia or 2) carrying any large CNV deletion greater than 500kb.

Within schizophrenia cases, CNV carriers did not demonstrate significantly lower risk scores than non-carriers. Cases with either class of CNV membership had higher polygenic scores compared to control subjects carrying CNVs. Control subjects with specific associated CNVs had lower polygenic scores than other control subjects, but controls with and without large deletions had similar scores. These initial results are partly inconsistent with an additive model of CNV and polygenic risk. The presence of an associated CNV alone is not sufficient to result in schizophrenia, but also requires a context of increased risk from common variants.

1415W

Copy Number Variants Implicated in Autism Spectrum Disorder Among Individuals of Hispanic Ethnicity. *N. Dueker[1], A.J. Griswold[1], H.N. Cukier[1], J. Jaworski[1], Y.N. Park[1], S. Slifer[1], I. Konidari[1], P.L. Whitehead[1], M. Schmidt[1], J.R. Gilbert[1,2], J.L. Haines[3], M.L. Cuccaro[1,2], E.R. Martin[1,2], M.A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Vanderbilt University School of Medicine, Vanderbilt University, Nashville, TN.

Autism spectrum disorder (ASD) is a developmental disorder characterized by social, communicative and behavioral impairments leading to average yearly medical costs that are 4.1-6.2 times higher for those with an ASD compared to those without. ASD is a highly heritable disorder and numerous copy number variants (CNVs) have been implicated with risk for developing ASD among individuals of European ancestry. However, it is unknown whether CNVs contribute to genetic risk of ASD among those of Hispanic ethnicity. Though Hispanic individuals comprise 16.7% of the current US population and are the largest minority in the US, investigation into genetic risk factors in this population is lacking. Therefore, we performed CNV analyses to identify loci associated with ASD in a large sample of Hispanic individuals. Participants for this study were drawn from a family-based study of ASD that included 291 ASD families and 118 pediatric controls that were genotyped using either the HumanOmni2.5-4v1_c or Human1M-Duo2_B platforms. All participants were self-reported, or had at least one parent of self-reported, Hispanic ethnicity. CNVs were detected using PennCNV and burden analyses were performed in PLINK v.1.0.7 using 284 ASD probands and 110 pediatric controls passing our quality control criteria. Analyses were restricted to CNVs spanning at least 5 SNPs and having length >30 kb. A total of 1,563 deletions and 1,342 duplications meeting our criteria were detected in our ASD probands, as well as 1,302 deletions and 794 duplications in our controls. The average size of deletions was larger in ASD probands compared to controls (ASD probands μ=92.2 kb, controls μ=79.1 kb, Burden p-value=0.04). However, the number of deletions, average size of duplications and number of duplications did not significantly differ between the two groups (Burden p=1). These burden results are similar to previously published data on CNVs in European Ancestry populations. We also detected CNVs in well-established ASD CNV regions in our probands, including a 500kb deletion at 16p11.2, and identified 25 case-unique CNVs in ASD candidate genes including Neurexin 1 (*NRXN1*). Deletions within *NRXN1* have been previously implicated in ASD and we identified 3 probands with deletions overlapping *NRXN1* in our study. These results provide preliminary evidence that CNVs, particularly deletions, may contribute to ASD risk among individuals of Hispanic ethnicity.

1416T

High resolution copy number variation analysis in monozygotic twins discordant for schizophrenia and delusional disorder. *F. Nishimura[1,2], A. Yoshikawa[1], T. Kato[2], T. Sasaki[3], K. Kasai[1], C. Kakiuchi[1,2].* 1) Department of Neuropsychiatry, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 2) Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Wako-shi, Saitama, Japan; 3) Office for Mental Health Support and Graduate School of Education, University of Tokyo, Tokyo, Japan.

Introduction: Schizophrenia is one of the major mental disorders with approximately 1% prevalence rate. It is the most debilitating psychiatric disorder, inflicting much burden upon the patient, the family, and the society. The family, twin and adoption studies have established that the genetics play a significant role in the etiology of schizophrenia. Disease concordance rate of schizophrenia within monozygotic twin is about 50% and on the other hand, disease concordance rate of schizophrenia within dizygotic twin is about 15%, suggesting that the genetic determinant and non-genetic factors play an important role. Copy number variation have been attracted a lot of attention lately and possibly plays a role in the etiology of schizophrenia. In this study, we examined the copy number variation differences within each twin that may potentially contributed to the etiology of schizophrenia. **Method:** We performed array comparative genomic hybridization on genomic DNA derived from peripheral blood obtained from two pairs of monozygotic twins discordant for schizophrenia and one pair discordant for delusional disorder using SurePrint G3 Human CGH 1x1M microarray from Agilent Techologies. All hybridizations were performed in duplicate, with the Cy3 and Cy5 dyes being swapped between the twin members of each pair to eliminate a possible dye-specific bias. Analyses of the data were carried out using the manufacturers' software, called Agilent Genomic Workbench. The objective of this study was clearly explained and written informed consent was obtained from all subjects. The study was approved by the Ethical Committee of the Faculty of Medicine, the University of Tokyo. **Results:** We found no differences in copy number variants between the three pairs of monozygotic twin. **Conclusion**: We have undertaken the high resolution copy number variation analysis in monozygotic twins discordant for schizophrenia and delusional disorder and identified no copy number variation differences within each monozygotic twin. CNV may not play a role in discordance of the development of psychosis in this present three monozygotic twin pairs.

1417F

Characterization of age-related expression of the interleukin-1beta gene in immature zebrafish brain after pentylenetetrazole evoked-seizures. *P. Barbalho, D. Nakata, C. Maurer-Morelli.* Department of Medical Genetics, State University of Campinas (UNICAMP)), Campinas, Sao Paulo, Brazil.

Introduction: Clinical and experimental evidence have demonstrated that seizures activate the inflammatory response increasing the release of interleukin-1 beta and that this activation is age-regulated in the developing brain. In this study, we aimed to investigate the interleukin-1 beta (il1b) transcript response after seizures in immature zebrafish brain. Material and Methods: Zebrafish were maintained according to standard procedures and all experiments were approved by the Animal Ethics Committee/UNICAMP. Seven and 15 days post-fertilization (dpf) zebrafish larvae were separated in Seizure (SG) and Control (CG) groups. Animals from SG group were individually exposed to Pentylenetetrazol (PTZ) 15mM for 20min. Animals from CG were exposed to same handling condition, but in normal bath water. A pool of 20 heads was used to compose a single larvae sample (n=5 each group). One hour after PTZ exposure animals were crioanesthetized and their heads were immediately isolated, frozen in N2(L) and total RNA extracted by Trizol (Invitrogen). Reverse transcriptase quantitative-PCR amplifications were carried out in triplicates with efla as endogenous control (TaqMan®, Applied Biosystems). The relative quantification (RQ) was calculated by the equation RQ=2 -ΔΔCT and the statistical analysis was performed by Mann-Whitney test. Results: The mRNA profile of the il1b gene showed an up regulation in SG when compared with the CG in both 7 dpf [(CG1h 1.44 ± 0.34; SG1h 2.46 ± 0.17 (p= 0.02)] and 15 dpf [(CG1h 1.86 ± 0.27; SG1h 3.63 ± 0.68 (p= 0.02)]. The i1lb transcript levels were 1.5 times more pronounced in 15dpf than in 7dpf larvae. Conclusion: This is the first study investigating age-related response of il1b gene after seizure in immature zebrafish brain. Our results showed that seizures induced in immature zebrafish brain promote an age-specific cytokine expression. This study contributes for a characterization of zebrafish as an experimental model of epilepsies. Support: FAPESP, CNPq.

1418W

A corrective gene silencing by RNA interference to control over-expressed *SNCA*. *M. Takahashi[1], M. Suzuki[2], N. Fujikake[2], M. Murata[3], K. Wada[2], Y. Nagai[2], H. Hohjoh[1].* 1) Dept. of Molecular Pharmacology, National Inst. of Neurosci., NCNP, Kodaira, Tokyo, Japan; 2) Dept. of Degenerative Neurological Diseases, National Inst. of Neurosci., NCNP, Kodaira, Tokyo, Japan; 3) National center Hospital, NCNP, Kodaira, Tokyo, Japan.

An over-expression of the wild-type *alpha synuclein* (*SNCA*) gene has the potential for developing Parkinson's disease (PD). For reduction of the over-expressed *SNCA*, the gene silencing against the wild-type *SNCA* by RNA interference (RNAi) may be a possible strategy for treatment for such an *SNCA*-over-expressed PD. However, a conventional RNAi that thoroughly silences *SNCA* may be unsuitable for such a treatment, and previous studies suggested that the wild-type *SNCA* plays an important role in neuronal functions. In this study, we attempted to establish a new RNAi technique to control the level of the over-expressed *SNCA* to its normal level. Various siRNAs were designed and subjected to an *in vitro* assay to examine their RNAi activities. From the screening, siRNAs that conferred approximately a half of inhibition of the expression of wild-type *SNCA* were selected as potential candidates for the particular RNAi. To further assess the effects of such an RNAi treatment *in vivo*, we focused on *Drosophila* model of PD, and then we examined the siRNAs in cultured *Drosophila* S2 cells and slightly modified the siRNA sequences such that they could confer an appropriate RNAi activity in *Drosophila* cells. As for establishment of transgenic RNAi flies, the selected siRNA was converted into a shRNA-expression construct (plasmid) that is controlled by an *elav*-GAL4 driver for pan-neuronal expression. Using the resultant shRNA expression plasmid, we are currently generating transgenic flies, which express *SNCA* and the shRNA in neurons. We would like to present and discuss the data of the fly models as well as the selected siRNAs for controlling the over-expressed *SNCA* in the meeting.

1419T

Partitioning the Heritability of Tourette Syndrome and Obsessive Compulsive Disorder Reveals Differences In Genetic Architecture. *L. Davis on behalf of the Tourette Syndrome Association International Consortium for Genetics (TSAICG).* Section Genetic Medicine, The University of Chicago, Chicago, IL.

Despite the completion of the first genome-wide association studies in Tourette Syndrome (TS) and obsessive compulsive disorder (OCD), very little is known about the genetic architecture of these two phenotypcially related early-onset neuropsychiatric disorders. The direct estimation of heritability from genome-wide common variant data as implemented in the program Genome-wide Complex Trait Analysis (GCTA) has provided a means to quantify and partition heritability attributable to all interrogated variants or to specific variant subsets of interest. We conducted multiple partitioning analyses to identify genomic elements that concentrate TS and OCD heritability. We partitioned by chromosome, MAF bin, and by annotation of variants that regulate gene expression in the brain. In addition, we assessed heritability for early onset and adult onset OCD and finally tested for genetic overlap between TS and OCD. After extensive quality control, we proceeded with analysis on a final data set of 617 TS cases and 4,116 TS controls genotyped on 393,387 SNPs, as well as 1,061 OCD cases and 4,236 OCD controls genotyped on 373,846 SNPs. Our analysis yielded a heritability point estimate of 0.58 (se = 0.09, p=5.64e-12) for TS, and 0.37 (se = 0.07, p=1.5e-07) for OCD. Among other notable results, we found that SNPs with a minor allele frequency of less than 5% accounted for 21% of the TS heritability and 0% of the OCD heritability. We discovered a disproportionately large contribution to OCD heritability originating from chromosome 15. Additionally, results showed that parietal eQTLs accounted for significantly more TS heritability (p=5.36e-46) and OCD heritability (p=2.80e-16), and cerebellum eQTLs accounted for significantly more TS heritability (p=4.02e-15) and OCD heritability (p=1.37e-14) than expected based on the number of SNPs tested under a uniform distribution model. Finally, we found a bivariate genetic correlation of 0.41 (SE=0.15) between TS and OCD. These findings suggests that 1) very little, if any, heritability is truly missing (i.e., unassayed) from TS and OCD GWAS studies of common variation and 2) shared environment does not result in excessive bias in twin and family studies of TS and OCD. The results further indicate that while there is some genetic overlap between these two phenotypically-related neuropsychiatric disorders, the two disorders have distinct genetic architectures. We present genomic characterization of both phenotypes.

1420F

Association of *DLG4* haplotype with increased risk of schizophrenia. *S. Balan[1], K. Yamada[1], E. Hattori[1], Y. Iwayama[1], T. Toyota[1], T. Ohnishi[1], M. Maekawa[1], M. Toyoshima[1], Y. Iwata[2], K. Suzuki[2], M. Kikuchi[3], T. Yoshikawa[1].* 1) Laboratory for Molecular Psychiatry, RIKEN- Brain Science Institute, Saitama, Japan; 2) Department of Psychiatry and Neurology, Hamamatsu University School of Medicine, Shizuoka, Japan; 3) Department of Psychiatry and Neurobiology, Kanazawa University Graduate School of Medicine, Kanazawa, Japan.

The post-synaptic density (PSD) of glutamatergic synapses are characterized by the expression of an array of proteins critical for maintaining synaptic dynamics. Alteration of protein expression levels in this matrix is a marked phenomenon of neuropsychiatric disorders including schizophrenia, where cognitive functions are impaired. To query the genetic predisposition of the genes expressed in the PSD with schizophrenia, a family-based association analysis of genetic variants in PSD genes viz; *DLG4, DLG1, PICK1* and *MDM2*, was performed, in Japanese pedigree samples (124 pedigrees, n = 376 subjects). A significant association of the variant rs17203281 in the *DLG4* gene was observed, with preferential transmission of the C allele (p = 0.02), although significance disappeared after correction for multiple testing. However replication analysis of this variant, yielded no association in a Chinese schizophrenia cohort (293 pedigrees, n = 1163 subjects) or in a Japanese case-control sample (n = 4182 subjects). Further the *DLG4* expression levels between postmortem brain samples from schizophrenia patients showed no significant changes from controls. Interestingly, a five marker haplotype in *DLG4*, involving rs2242449, rs17203281, rs390200, rs222853 and rs222837, was enriched in a population specific manner, where the sequences A-C-C-C-A and G-C-C-C-A accumulated in Japanese (p = 0.0009) and Chinese (p = 0.0007) schizophrenia pedigree samples, respectively. None of the variants in other examined candidate genes showed any significant association in these samples. The present study highlights a putative role for *DLG4* in schizophrenia pathogenesis, evidenced by haplotype association, and warrants further dense screening for variants within these haplotypes.

1421W

MGluR gene network alterations confer risk for Autism Spectrum Disorder in 22q11.2 Deletion Syndrome. *T.L. Wenger[1, 2, 3], C. Kao[2], D.M. McDonald-McGinn[1], A. Bailey[1], R.T. Schultz[3], B.S. Emanuel[1], B.E. Morrow[4], E.H. Zackai[1], H. Hakonarson[2].* 1) Division of Human Genetics and Molecular Biology, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Center for Autism Research, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Division of Translational Genetics, Albert Einstein College of Medicine, Bronx, NY.

Children with 22q11.2 Deletion Syndrome (22q11.2 DS) are at elevated risk for Autism Spectrum Disorder (ASD), with 20% of individuals receiving a diagnosis. The medical and psychiatric manifestations of 22q11.2 DS are quite variable between individuals, despite most having involvement of the same 3 Mb region encompassing ~45 genes. Genetic modifiers for the risk of ASD in 22q11.2 DS as well as all other forms of syndromic ASD are poorly understood. We previously showed that copy number variants in the metabotropic glutamate receptor (mGluR) network are associated with an elevated risk of syndromic ASD and are uncommon in typically developing children. We hypothesized that alterations in the mGluR network outside of the deleted region may provide a ‘second hit’, conferring greater risk for development of ASD. The purpose of this study was to determine whether alterations in the mGluR network outside of 22q11.2 are associated with an increased risk of ASD in children with 22q11.2 DS. METHODS: DNA from 75 children with 22q11.2 DS were selected, including children with clinical diagnosis of ASD (n=25) and without ASD (n=50). High density microarray and custom gene chip were used to determine whether children had copy number variants in genes in the mGluR network, including those in the canonical pathway and the mGluR interactome. RESULTS: Among children with 22q11.2 DS, 20% of those with ASD had an alteration in an mGluR network gene, compared with 2% (1/50) of those without ASD. (p= 0.014) DISCUSSION: Approximately 20% of cases of ASD occur in the context of another genetic syndrome. We previously demonstrated alterations in the mGluR network are enriched in children with many forms of syndromic ASD. All individuals with classic 22q11.2 DS are missing one copy of RANBP1, an mGluR network gene located within the deleted region. Decreased expression of RANBP1 is also found following prenatal exposure to thalidomide and valproic acid, which are both associated with increased risk of ASD. While all individuals have one missing copy of RANBP1, the rate of ASD in 22q11.2 DS is only 20%, similar to the rate in younger siblings of children with ASD. This suggests that additional factors contribute to risk of ASD in 22q11.2. Our data suggest that CNV alterations in the mGluR network may be one important factor in determining which children with 22q11.2 DS will develop ASD.

1422T

Co-expression networks implicate human mid-fetal deep cortical projection neurons in the pathogenesis of autism. *A.J. Willsey[1], S.J. Sanders[1], M. Li[2,3], S. Dong[1,4], A.T. Tebbenkamp[2,3], R.A. Muhle[1,3,5], S.K. Reilly[1,3], L. Lin[6], S. Fertuzinhos[2,3], J.A. Miller[7], M.T. Murtha[8], C. Bichsel[2,3], W. Niu[1,3,5], J. Cotney[1,3], A.G. Ercan-Sencicek[8], J. Gockley[1], W. Han[2,3], X. He[9], L. Klei[10], J. Lei[11], L. Liu[11], C. Lu[11], X. Xu[2,3], E.S. Lein[7], L. Wei[4], J.P. Noonan[1,3], K. Roeder[9,11], B. Devlin[10], N. Sestan[2,3], M.W. State[1,5,8,12,13].* 1) Department of Genetics, Yale University School of Medicine, New Haven, CT; 2) Department of Neurobiology, Yale University School of Medicine, New Haven, CT; 3) Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, CT; 4) Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing, China; 5) Child Study Center, Yale University School of Medicine, New Haven, CT; 6) Department of Computational Biology and Bioinformatics, Yale University, New Haven, CT; 7) Allen Institute for Brain Science, Seattle, WA; 8) Program on Neurogenetics, Yale University School of Medicine, New Haven, CT; 9) Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 10) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 11) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 12) Department of Psychiatry, Yale University School of Medicine, New Haven, CT; 13) Department of Psychiatry, University of California, San Francisco, San Francisco, CA.

Autism spectrum disorder (ASD) is a complex developmental syndrome of unknown etiology. De novo loss of function mutations, detected by exome sequencing, have recently led to the identification of nine high-confidence ASD genes (hcASD). Working on the hypothesis that these pleiotropic genes must disrupt shared biological processes to result in a common phenotype, we have attempted to identify brain regions and developmental periods at which they converge. We have constructed co-expression networks around these hcASD ‘seed’ genes, leveraging a rich expression dataset encompassing multiple human brain regions across the lifespan. By assessing enrichment of a second, independent set of ASD genes within spatio-temporally defined subsets of this data, we demonstrate convergence in mid fetal layer 5/6 cortical projection neurons. The approach provides critical information on when, where and in what cell types specific mutations in specific genes may be most productively modeled to further clarify ASD biology.

1423F

Genome-wide methylome and transcriptome analyses reveal novel epigenetic regulation patterns in schizophrenia and bipolar disorder. *C. Xu[1,2,3], Y. Xiao[2], Y. Li[2], C. Camarillo[3], Y. Ping[2], J. Xu[2], T. Arana[1], H. Zhao[2], P. Thompson[4], C. Xu[2], Z. Zhao[2], H. Chen[2], C. Mao[8], Y. Zhang[2], B. Su[2], M. Escamilla[1,3], A. Ontiveros[5], H. Nicolini[6], A. Jerez[7], X. LI[2].* 1) Departments of Psychiatry, Paul L. Foster School of Medicine, Texas Tech University Health Science Center, El Paso, TX, USA; 2) College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, the People’s Republic of China; 3) Center of Excellence in Neuroscience, Paul L. Foster School of Medicine, Texas Tech University Health Science Center, El Paso, TX, USA; 4) Southwest Brain Bank, Department of Psychiatry, UTHSCSA; 5) Instituto de Informacion de Investigacion en Salud Mental, Monterrey, Mexico; 6) Medical and Family Research Group, Carracci S.C., Mexico City, Mexico; 7) Centro Internacional de Trastornos Afectivos y de la Conducta Adictiva-CITACA, Guatemala; 8) University of Toronto, Toronto, ON, Canada.

Schizophrenia (SC) and bipolar disorder (BP) are two complex disorders where the inheritance pattern is likely to be complicated by epigenetic factors yet to be elucidated. Using next-generation sequencing, we quantitatively compared the DNA methylation landscapes of SC and BP to normal controls using blood samples (N=10) and two brain regions, Brodmann area (BA)9 and BA24 (N=36), from the Latino population, the largest minority group in the U.S. We first identified the substantial differentially methylated regions (DMR) in SC and BP. The DMRs identified from the peripheral blood samples preferentially occurred at gene regions with specific features, including promoters, 3'-UTRs, and 5'-UTRs. In particular, we identified distinct patterns of promoter aberrant methylation around transcription start sites, where aberrant methylation occurred not only on CpG-islands (CGIs), but also on flanking regions and CGI sparse promoters. To our surprise, different brain regions showed completely distinct distributions of DMRs across the genomes. In the BA9 region of both SC and BP, we observed widespread hypomethylation, in comparison to normal controls, preferentially targeting the terminal ends of chromosomes. In contrast, in the BA24 region, both SC and BP displayed extensive gain of methylation. Notably, in the two brain regions of SC and BP, only a few DMRs overlapped with promoters, whereas, a greater proportion occurred across introns and intergenic regions. Pathway analysis highlighted that many pathways relevant to psychiatric disorders were associated with DMRs located at introns. Transcriptomic analysis revealed consistent dysfunctional pathways with those determined by DMRs. Furthermore, DMRs in the same brain regions of SC and BP could distinguish BP and/or SC from normal controls while differentially expressed genes could not. The results from the brain regions support a major role of brain region-dependent aberrant DNA methylation in the pathogenesis of the disorders. Pathway analysis of genes with distinct promoter aberrant methylated patterns identified in the peripheral blood samples showed a significant enrichment of epigenetic changes in biological pathways directly relevant to psychotic disorders and neurodevelopment. This comprehensive methylome-map from peripheral blood and postmortem brain provide the precise genomic locations that undergo methylation changes and further our understanding of epigenetic regulation in the etiology of SC and BP.

1424W

Neuronal cellular models of 22q11.2 DS exhibit disruptions to the miRNA regulatory pathway and may increase the rate of schizophrenia in individuals with 22q11.2 DS. *W. Manley[1], M.R. Ababon[2], M.P. Moreau[3], P.G. Matteson[2], J.H. Millonig[2], L.M. Brzustowicz[1].* 1) Department of Genetics, Rutgers University, Piscataway, New Jersey 08854-8095, USA; 2) Center for Advanced Biotechnology and Medicine, Piscataway, NJ 08854, USA; 3) RUCDR, Rutgers University, Piscataway, New Jersey 08854-8095, USA.

Schizophrenia is a complex and poorly understood disease caused by the interplay of many environmental and genetic factors. The 22q11.2 deletion syndrome is a disorder that is caused by the microdeletion of part of chromosome 22 leading to a 25-fold greater chance of developing schizophrenia in affected individuals versus the general population. The missing 22q11.2 region contains the DGCR8 gene, which is required for the initial step of miRNA biogenesis. However, the 22q11.2 deletion itself is not sufficient to cause schizophrenia, since 75% of individuals with this deletion do not develop the disease. We hypothesize that the 22q11.2 deletion increases the risk of schizophrenia through alterations in miRNA regulatory networks via depletion of multiple miRNAs. These miRNAs may serve as a protective buffer against the accumulation of deleterious mutations at other schizophrenia risk loci. We have profiled miRNAs in multiple cell lines from individuals with 22q11.2 that are haploinsufficient for DGCR8. Additionally, we have also developed neuronal human cellular models of disruptions to the miRNA regulatory network caused as a result of DGCR8 deficiency. The neuroblastoma cell line SH-SY5Y was transfected via electroporation with GIPZ lentiviral shRNAmir constructs specific for DGCR8 (Open Biosystems). The GIPZ constructs were prepared using PureYield Plasmid MaxiPrep System (Promega) and sequenced (GeneWiz) prior to use. For selection purposes, all constructs contained genes for GFP and puromycin resistance. The levels of DGCR8 gene expression was quantified using the 7900HT Fast Real-Time PCR System (Applied Biosystems) along with Taqman Gene Expression Assays to ensure DGCR8 reduction. We have characterized disruptions to the miRNA regulatory network in these DGCR8 deficient cell lines using Taqman Array Human Microarray Cards Version 3.0. Here we will present the miRNAs that we have identified to be differentially expressed in DGCR8 deficient cells. We predict that these miRNAs found to be differentially expressed could be involved in the elevated risk of developing schizophrenia in individuals with 22q11.2 deletion syndrome versus the general population. We have also performed RNA-seq using Illumina TruSeq protocols to characterize differences at the RNA level between DGCR8 deficient cells and otherwise healthy control cells. Future studies will focus on determining the biological targets of these miRNAs using miRNA target prediction software.

1425T

A case of monozygotic female triplets diagnosed with autism: comprehensive genomic analyses. *Z. Talebizadeh, S. Soden, A. Gadashova.* Pediatrics, Children's Mercy Hospital, Kansas City, MO.

Twin study design is a classical approach commonly used in many genetic studies; in particular, it may enhance the power of conventional strategies for detecting genetic causes of complex disorders. Autism is a common early onset neurodevelopmental disorder belonging to a group of conditions known as autism spectrum disorders (ASDs). Although there is strong evidence for genetic involvement in susceptibility to ASD, the etiology of most cases remains unknown. We present a case of female triplets diagnosed with autism, and discuss results of their genomic data analyses, including genome-wide copy number variations (CNVs) and expression profiling. No pathogenic CNVs were detected in the triplets' peripheral blood samples. Genome-wide microRNA and exon array profiling were performed, and compared with three unaffected females as controls, to examine the potential abnormalities at the noncoding or alternative splicing levels. The exon arrays were analyzed at both exon and gene levels. One hundred thirty five genes showed differential exon expression in the triplets, compared with controls. By correlating expression levels of microRNAs with their predicted target mRNA(s), differential expression was detected in 51 genes (i.e., inversely correlated with microRNAs). Pathway analysis using the 135 potentially alternatively spliced genes and the 51 genes obtained from microRNAs and mRNAs correlation showed enrichment in biologically relevant pathways. All three female triplets had skewed patterns of X chromosome inactivation. The absence of pathogenic CNVs along with microRNAs and alternative splicing profiles and X inactivation skeweness highlights the potential role of epigenetic factors in this case of female triplets with autism. Family history showed evidences of a maternal familial diabetes and overweight, which is in agreement with the recent report of the association between maternal obesity and autism. This comprehensive case report demonstrates the value of utilizing rare triplet cases with autism to further delineate the underlying mechanism of this complex disorder. Despite growing interest in utilizing twins in genetic and epigenetic studies, triplets and higher-order multiples have been rarely included in these studies. Our study demonstrates that while triplets are less common than twins, they present a valuable and potentially stronger resource for understanding the underlying genetic and epigenetic factors in autism.

1426F

Does Presymptomatic testing changes Age at ONset : a prospective study in Huntington disease (PAON study). *A. Durr[1,2,3,4,5], S. Tezenas du Montcel[2], MF. Jutras[3], C. Jauffret[1,3], S. Benaich[1,3], A. Herson[1], J. Feingold[1], M. Gargiulo[1].* 1) APHP Genetic Department, Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Paris France Paris, France; 2) AP-HP, Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Department of Biostatistics and Medical Informatics, Paris, France; 3) ICM (Brain and Spine Institute), Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Paris France; 4) Inserm, UMR_S975, CRICM, F-75013, Paris, France , Groupe Hospitalier Pitié-Salpêtrière Charles-Foix, Paris France; 5) UPMC Univ Paris 06, UMR_S975, F-75013, Paris, France.

In PAON, we studied prospectively the impact of presymptomatic testing (PT) on the onset of symptoms in Huntington disease (HD). The goal of our study was to estimate a possible test-effect on age at onset in carriers of the mutation who choose to know their genetic status. Does knowledge about being a carrier of the pathological mutation changes age at onset or onset modalities? In Paris, at the Salpêtriere University Hospital - Genetic Department, 1634 at-risk persons for HD requested genetic testing since 1992 and entered multistep and multidisciplinary care and counseling. We went back to those who received a test result before 2009, in order that their age at examination was reaching the mean expected onset of 40 years. Among them, 463 were non carriers and 302 carriers of an abnormal expanded CAG repeat. Detailed follow up examination was available for 208 individuals (69%). We interviewed and examined additional 62 persons (PAON cohort), using a self-administered questionnaires (to explore their carrier condition) and the UHDRS evaluation score. We estimated age at onset with the Langbehn formula (personal communication of Doug Langbehn and Clin Genet. 2004 Apr;65(4):267–77). Among the PAON cohort, there were 35 affected with UHDRS >5 (mean 18.3 +/– 11.9) and 27 unaffected (mean 2.9 +/– 1.8). Surprisingly, age at onset observed was 6.0 +/– 8.7 years earlier than onset calculated. Among those followed more regularly, 60 were affected (mean 26.1 +/– 18.3) and had ages at onset 5.2 +/– 7.2 years earlier than calculated. This did not differ significantly from a retrospective cohort of patients who did not have PT (n=522) where age at onset was 3.6 +/- 9.6 years earlier (p= 0.13 and p= 0.12, respectively). In conclusion, onset observed in a prospective cohort was earlier than that estimated based on CAG repeat length and age. This could be due to 1) a greater self-observance and a rising burden after PT, but the comparison with onset in patients without PT was not significant; 2) the fact, that the Langbehn formula underestimates onset. Importantly, for carriers of the mutation, knowledge about the genetic status after PT does not allow better awareness of the onset. Affected carriers did not always acknowledge symptoms related to the disease (44%) and unaffected carriers did feel affected (18.5%). A subjective anticipation (self-observation) of the onset was observed for both, affected (96%) and unaffected carriers (70%).

1427W

Genome-wide expression changes in a higher state of consciousness. *M. Ravnik-Glavac[1], S. Hrasovec[1], J. Bon[2], J. Dreo[2], D. Glavac[1].* 1) Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia; 2) Department of Neurology, University Medical Centre, Ljubljana, Slovenia.

Higher states of consciousness in which the human mind can transcend the boundaries of logic and reason are envisioned as natural to the experience and potential growth of every human being. So far they have been mostly monitored by electrophysiological methods. In this study we were particularly interested in discovering the molecular transcriptional basis of higher states of consciousness. In addition to phenomenological reports of meditators who participated in this study the generated higher states of consciousness were also EEG recorded. We assessed the whole genome gene expression analysis of long-term meditators in four separate trials and detected significant differential gene expression in association with higher states of consciousness. The number of differently expressed genes as well as high proportion of genes themselves differed between meditators. Despite this, gene ontology enrichment analysis found significant biological and molecular processes shared among meditators' higher state of consciousness.

1428T

Differences in gene expression and DNA methylation in drug-naïve first-episode psychosis patients. *V.K. Ota[1], C.S. Noto[2], A. Gadelha[2], M.L. Santoro[1], B.B. Ortiz[2], R.S. Stilhano[3], E.S. Gouvea[4], P.N. Silva[1,2], C.G. Olmos[4], L.M.N. Spindola[1], E.H.S. Andrade[2], M.I. Melaragno[1], M.A.C. Smith[1], S.W. Han[3], Q. Cordeiro[4], R.A. Bressan[2], S.I. Belangero[1,2].* 1) Morphology and Genetics, UNIFESP, Sao Paulo, Brazil; 2) Psychiatry, UNIFESP, Sao Paulo, Brazil; 3) Interdisciplinary Center for Gene Therapy, UNIFESP, Sao Paulo, Brazil; 4) Psychiatry, ISCMSP, Sao Paulo, Brazil.

Schizophrenia is a severe mental health disorder with a high heritability. The study of gene expression levels in blood of patients in the beginning of the disease, such as first-episode of psychosis (FEP) may be useful to detect changes in gene expression despite treatment effects. In this study we aimed to analyze gene expression in whole blood, comparing: a) drug-naïve FEP patients and healthy subjects; and b) drug-naïve FEP patients before and after treatment with antipsychotic drugs. Also, we investigated if those differentially expressed genes were regulated by DNA methylation. All the patients (n = 38) were evaluated by a trained psychiatrist, twice (at admission and 8 weeks after beginning antipsychotic treatment), and the healthy controls (n = 38) were also evaluated to exclude any psychiatric disorder. Whole blood was collected from each participant during clinical assessments. Expression levels of 40 genes related to neurotransmission and neurodevelopment were quantified with a customized RT2 Profiler™ PCR Array, which is based on SYBR Green detection of cDNA amplification. For methylation analysis, bisulfite sequencing was performed. For data analysis, we compared $2^{-\Delta}$Ct or DNA methylation percentage values using t-test (FEP patients x healthy controls) or paired t-test (before x after treatment). Significant downregulation of *GCH1* gene was observed (Fold regulation (FR)= −1.34, p=0.007) comparing FEP and controls. Downregulation of *GABRR2* (FR = −1.33, p=0.01) and upregulation of *GCHFR* (FR = 1.18, p=0.008) were found in FEP after treatment. Also, a significant hypermethylation in *GCH1* was detected in FEP comparing to healthy controls (p=0.034). *GCH1* codes for GTP cyclohydrolase I, an enzyme involved in the synthesis of BH4, which is an essential cofactor for tyrosine, serotonin and L-Dopa. Moreover, its expression is regulated by GCHFR, which seemed to be upregulated after treatment with antipsychotic drugs. Also, GABRR2, which is a GABA receptor, seems to be dysregulated after treatment. Therefore, *GCH1* may have a role in the genesis of psychosis whereas *GABRR2* and *GCHFR* may be related to the treatment leading towards a better understanding of illness. Funding for this study was provided by FAPESP 2010/19176-3; 2010/08968-6 and 2011/50740-5.

1429F

Investigating the role of the Histidine Decarboxylase Gene in Tourette Syndrome etiology. *I. Karagiannidis[1], H. Potamianou[1], G. Heiman[2,3], L. Deng[2,3], J. Xing[2,3], N. Sun[2,3], C. Nasello[2,3], P. Sandor[4], C. Barr[4], P. Paschou[1].* 1) Molecular Biology & Genetics, Democritus University of Thrace, Alexandroupoli, Greece; 2) Department of Genetics, Rutgers University, Piscataway, NJ, USA; 3) Human Genetics Institute of New Jersey, Piscataway, NJ, USA; 4) 4-Toronto Western Research Institute, University Health Network, Toronto, Ontario, Canada.

Gilles de la Tourette Syndrome (TS) is a childhood onset neurodevelopmental disorder with complex genetic background that manifests as multiple motor and vocal tics and high comorbidity rates with other neurodevelopmental disorders. While previous studies suggest that genetic and environmental factors contribute to the etiology of the disorder, its pathophysiology remains largely unknown. Recent studies report the possible implication of the histamine decarboxylase (HDC) gene, which is the key enzyme in histamine production, suggesting histaminergic dysfunction may play a role in the etiology of TS (Ercan-Sencicek et al., NEJM 2010; Fernandez TV et al., Biol Psychiatry 2012). We recently investigated variation across the HDC gene for association with TS by interrogating 12 tagging SNPs (tSNP) across the region in a large sample of 520nuclear families originating from seven European populations as well as Canada. Strong over-transmission of alleles at two SNPs (rs854150 and rs1894236) in the joint dataset was observed, as well as statistically significant associated haplotypes (Karagiannidis et al., J Med Genetics 2013).Using a subset of these samples, we sequenced the complete HDC gene in 384 individuals with TS using NextGen high-throughput sequencing technology and validation by Sanger sequencing. Our results may help elucidate the role and relative contribution of this pathway in TS etiology and will facilitate the study of the HDC gene in different populations of European ancestry.

1430W

Identification of target genes for the language-related FOXP2 and its isoform. *T. Ohta[1], N. Niikawa[1], T. Oikawa[2].* 1) Res Inst, Personal Hlth Sci, Hlth Sci Univ Hokkaido, Tobetsu, Japan; 2) Dept Comm Dis, Sch Psychol Sci, Hlth Sci Univ Hokkaido, Sapporo, Japan.

FOXP2 is the first key gene involved in the development of human languages. FOXP2 protein is a highly conserved transcription factor in any animals. Only two amino acids are different between human and chimpanzee FOXP2 proteins, and this difference was supposed to be associated with neuron development for language skill between the two species. Previous study demonstrated that the *FOXP2* gene was expressed not only in neuronal cells but also in multiple tissues. The aim of the present study is to identify genes targeted by FOXP2. We first generated various transgenes (Tg) in hek293-derived cells, using the flip-in recombinase system for the gene integration at the identical chromosome region and the Tet-ON/OFF systems to minimize transfection artifacts. The Tg included human *FOXP2* (Hum), a human *FOXP2* isoform (Iso), chimpanzee *Foxp2* (Chimp), or a negative control (CAT). This Iso was encoded by an alternative transcript with the complete open reading frame for a human FOXP2 with extra 25 amino acids, and was confirmed to be expressed in multiple human adult tissues, such as heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas, but not in any mouse tissues. We analyzed expression of the *FOXP2* from the transgenes (*FOXP2*-Tg) in Tet-ON or in Tet-OFF cells. The expression of *FOXP2*-Tg in each Tet-ON cells was higher than in Tet-OFF cells, and *FOXP2*-Tg in Tet-ON cells were 5-10 times more highly expressed than that in CAT. Using these cell lines, we performed microarray screening to pick up altered expression pattern among these 8 groups. Then, we confirmed the expression pattern again in GOTO cells, which is inducible into glial cells and neuronal cells, with stable and random integration of each Tg, using real time PCR. Consequently, results of the microarray analysis and real time PCR suggested several candidate target genes: *CRYAB*, GABA receptor genes, *UBE3A*, *GFRA* and *EDNRB*. The *CRYAB* gene was upregulated in all Tet-ON cells, and the gene was more highly expressed in Hum and Chimp Tet-ON than the CAT Tet-ON cells. Similar expression patterns were confirmed with Tg in GOTO cells. Interestingly, genes for some GABA receptors and the *UBE3A* gene were observed to be downregulated, while glial cell related genes, *GFRA* and *EDNRB* were upregulated in hek293 derived cells and in GOTO cells containing Hum Tg and Iso Tg. These affected genes could contribute to the human language development.

1431W

Reference-free quality assessment and complexity estimation of next-generation sequencing data. *S.Y. Anvar[1,2], L. Khachatryan[1], M. Vermaat[1,3], M. van Galen[1], I. Pulyakhina[1], Y. Ariyurek[2], K. Kraaijeveld[2,4], P.A.C. 't Hoen[1], J.T. den Dunnen[1,2], P. de Knijff[1], J.F.J. Laros[1,2,3].* 1) Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, Netherlands; 3) Netherlands Bioinformatics Center, Nijmegen, Netherlands; 4) University of Applied Sciences, Leiden, Netherlands.

Current methodologies for determining the quality and complexity of next-generation sequencing (NGS) data heavily rely on alignment-based quality measures. However, reference genomes are not always available and genomic information about complex regions of the genome such as large duplications or structural rearrangements are likely to be overlooked by alignment-dependent approaches. To address these limitations, we have developed a new methodology (kMer) that is independent of a reference sequence in determining the quality of NGS data and allows for pairwise comparison across a series of samples. kMer provides various tools to systematically and dynamically assess the NGS data after generating k-mer profiles. In addition, kMer provides a framework to estimate the complexity of the sequence library as well as estimating if the sample is sequenced in sufficient depth. We have applied kMer on four sets of NGS data that consist of 59 targeted resequencing, 43 whole-exomes, 49 full-genomes, and 665 RNA-Seq samples. In each set, kMer could precisely detect and separate samples based on various technical variations that were introduced during sample prep or sequencing such as high duplication rate, low capture performance, differing capture protocols, and high amount of library chimaeras. Notably, some of the aforementioned artefacts could only be characterised after vigorous quality assessments and have otherwise been missed by standard NGS QC tools. These often hidden artefacts undermine the potential application of NGS in clinical diagnostics as they may result in obscured downstream analysis. In addition, we show that kMer allows for estimating the complexity of NGS data, a vital property of NGS in diverse studies ranging from de novo assembly applications to detection of a shift in abundance in a series of microbiomes. In particular, we show that the complexity of NGS data is reflected on the distribution of k-mer profiles, with the resolution of separating differing exome capture kits, as full-genome, whole-exome, and RNA-Seq NGS data exhibit distinctive profiles with differing complexity. The various techniques that constitute this new paradigm allow for robust evaluation of the quality and complexity of NGS data and reliable and unsupervised view of genomic information.

1432T

RNA-seq reveals increased expression of other Osteogenesis Imperfecta (OI)-causing genes in diaphyseal bone from the *Col1a2*$^{p.G610C}$ mouse model of OI. *U.M. Ayturk[1, 2], C.M. Jacobsen[3, 4], A.G. Robling[5], M.L. Warman[1, 2, 6].* 1) Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA; 2) Department of Genetics, Harvard Medical School, Boston, MA; 3) Division of Endocrinology, Boston Children's Hospital, Boston, MA; 4) Division of Genetics, Boston Children's Hospital, Boston, MA; 5) Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN; 6) Howard Hughes Medical Institue, Boston, MA.

Osteogenesis Imperfecta (OI) is a disease characterized by increased skeletal fragility. Most patients with OI have defects in the synthesis of type I collagen caused by mutations in either *COL1A1* or *COL1A2*, which encode the α1(I) and α2(I). collagen chains. We developed an RNA-sequencing (RNA-seq) method for mouse diaphyseal bone and used this method to identify differences in gene expression between wild-type mice and mice with a p.G610C missense mutation in *Col1a*2 that models human type IV OI. Tibia diaphyseal bones from 12-week-old male wild-type and *Col1a2*$^{p.G610C/+}$ mice (3 mice/genotype, 2 tibiae/mouse), were used to create individual bar coded libraries. We obtained ~14 million RNA-seq reads per library. We then compared the RNA-seq data between OI and WT mice. We identified 21 genes whose expression was significantly increased by more than 2-fold in diaphyseal bone from the *Col1a2*$^{p.G610C/+}$ mice compared to wild-type mice. Interestingly, 5 of these genes (*Serpinf1, Serpinh1, Smpd3, Bmp1* and *Ifitm5*) have also been found to cause OI when mutated. Eight other differentially expressed genes (*Bglap, Bglap2, Id3, Ptprs, Cdkn1a, Creb3l1, Itga10* and *Col13a1*) have skeletal phenotypes when mutated in mice. The other differentially expressed genes (*Rcn3, Vasn, Nupr1, Olfml3, Metrnl, Cacnb3* and *Slc22a17*) have not yet been studied with respect to their roles in the skeleton. These data suggest that RNA-seq in mouse models of OI can identify differentially expressed genes that are also likely to have important roles in the development and/or maintenance of the skeleton.

1433F

SoftSearch: a tool to identify genomic structural variations from paired-end Illumina reads. *S. Baheti, S.N. Hart, V. Sarangi, R. Moore, J.D. Bhavsar, JP. Kocher.* Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN.

Genomic structural variation (SV) represents a significant, yet poorly understood contribution to an individual's genetic makeup. Advanced next-generation sequencing technologies are widely used to discover such variations, but there is no single SV detection algorithm that is considered a community standard. In response to this challenge, we developed a memory-efficient algorithm called SoftSearch, for discovering genomic rearrangements in Illumina paired-end next-generation sequencing data. SoftSearch simultaneously utilizes multiple strategies for detecting SV including split-read, discordant read-pair, and unmated pairs. Split-reads and read pairs that are co-localized and supporting the same directionality are used to refine the breakpoints and classify them according to their type, e.g. large insertions-deletions, inversions, duplications and balanced or unbalanced inter-chromosomal translocations. SoftSearch outputs predicted structural variants in universally accepted VCF format. After extensive comparisons we attempted to recall known structural variants that have previously been validated in the HapMap NA12878 sample. SoftSearch recalled the most true positive results (n=660), 4 times more variants than BreakDancer (n= 165). DELLY identified a comparable number of variants (n=654) variants, while CREST had the fewest (n=505). CREST and DELLY took the most amount of time to complete with more than 230 and 194 hours, respectively. The memory usage was also highest with these tools (5.2, and 7.1GB, respectively). BreakDancer was faster and used less memory requiring 48 hours and 1.5GB memory. Strikingly, SoftSearch completed the analysis in less than 20 hours using only 0.7GB memory. Only a small number of soft-masked bases from split reads and a few discordant read-pairs are necessary to identify an SV, which on their own would not be sufficient to make an SV call. Key features of SoftSearch are 1) not requiring secondary (or exhaustive primary) alignment, 2) portability into established sequencing workflows, and 3) is applicable to any DNA-sequencing experiment (e.g. whole genome, exome, custom capture, etc.). We show that SoftSearch can identify more true SVs with less evidence, including clinically relevant SVs in the BRCA2 gene missed by other SV tools while offering significantly improved overall performance.

1434W

Thousand Genomes And HLA Typing By NGS: Hidden Treasures In Public Short Read Data. *A. Berces[1], E. Major[1], K. Rigo[1], S. Juhos[1], T. Hague[1], P. Gourraud[2].* 1) Omixon Biocomputing, Budapest, Hungary; 2) University of California, San Francisco, US.

One of the important goals of the 1000 Genomes project was to find common mutation in diverse populations with the help of next generation sequencing. Earlier, the HapMap project with similar goals made it possible not only to map frequent mutations but there are already publications about sequencing based HLA typing using NGS. We are presenting a sufficiently fast algorithm using 1KG Illumina data to obtain HLA types for HLA-A, B, C, DRB1 and DQB1 genes. For validation the results of Sanger capillary sequencing based HLA typing was used for over thousand Coriell samples. According to our results it is possible to determine the correct HLA types from public 1KG whole-exome Illumina data with 90% or higher concordance if proper quality check measures are applied. We are also presenting the possible causes of mistypings. The method opens perspectives for typing other systems like MICA, MICB and KIR.

1435T

A consolidated genotype quality control reporting pipeline. *J. Boston, Y. Bradford, J. Haines, W. Bush.* Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Quality control is a critical part of any genetic analysis and must be performed in a reproducible, well-documented, and interactive way. To improve the throughput and reliability of this QC process, we developed a Clojure-based pipeline for generating quality control reports and plots in a systematic fashion. From a command-line interface, analysts submit data files to a centralized server and specify options for QC procedures. These include typical operations such as sex concordance, genotype call rate, sample efficiency, allele frequency distributions, Hardy-Weinberg tests, identity by descent calculations for relatedness, sample concordance checks, and tests for Mendelian inconsistencies. QC jobs are submitted to a scheduled computing cluster through portable batch system (PBS) scripts. Once completed, results from all operations are compiled into a single report of relevant information, complete with appropriate plots, eliminating the possibility of data transcription (through cut and paste operations) or other data processing errors. Because genotype QC is an iterative process, changes to the filtering process can be made after examining patterns in the original report, simply by reissuing the PBS script with new parameter settings. As such, the entire QC process is much more standardized and less prone to error. This consolidated workflow also improves and facilitates communication between analysts, study designers, and technicians about QC-related issues through a standardized report with multiple visualizations. While the pipeline was implemented for processing of genome-wide association data, the application is also applicable to next-generation sequence data.

1436F

Variome-based identification of Crohn's disease predisposition. *Y. Bromberg.* Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ., USA.

Crohn's disease (CD) is an auto-inflammatory disorder and one of the main subtypes of inflammatory bowel disease. CD affects over half a million people in North America alone. Tens of thousands of new CD cases are diagnosed every year, mostly in young adults in their twenties and early thirties. Human and animal models of CD illustrate strong familial patterns of disease predisposition, suggesting interplay of genetic and/or environmental causes. Recent GWA studies highlight key CD pathogenic pathways, but the discovered susceptibility loci account for less than 25% of disease heritability. Rather than the result of a single very severe malfunction, most Crohn's manifestations are thus likely due to 'unlucky' summations of many variants of individually small functional effects, as is the case with many other complex diseases. We have started building VariAD (Variome Analysis of Disease) - a novel computational tool for annotating an individual's CD disease predisposition on the basis of a combination of genome variants (individual variome). Disease associated variants often act by disrupting the sequence and the molecular functionality of the affected gene-product. Our method assumes that a genome-specific set of functionally significant coding single nucleotide polymorphisms (cSNPs) in CD genes is sufficient to describe an individual's CD susceptibility. We collected an expanded set of CD genes by mining scientific literature for phenotypic descriptors of CD. We annotate the 'broken-ness' of extracted disease genes based on their corresponding variants, taking into account the specific locus genotypes. The genome-specific combination of affected disease-genes drives the prototype CD prognosis engine. This research represents a unique way of looking at the variation in a single human genome, informed by molecular mechanisms of disease and achieves over 75% accuracy of prediction on the two data sets tested (122 case/control individuals). VariAD will motivate new testable hypothesis regarding the biological mechanisms of CD and provide a means for earlier prognosis, diagnosis and development of better treatments.

1437W

Clinical exome quality assurance through comprehensive coverage analysis improves the utility of exome sequencing. *C.J. Buhay[1], Q. Wang[1], M. Wang[1], Y. Han[1], H. Dinh[1], H. Doddapaneni[1], Y. Yang[2], Y. Ding[2], M. Bainbridge[1], E. Boerwinkle[1], J. Reid[1], D. Muzny[1], R. Gibbs[1].* 1) Baylor College of Medicine - Human Genome Sequencing Center, Houston, TX; 2) Baylor College of Medicine - Whole Genome Sequencing Laboratory, Houston, TX.

It is important for whole exome sequencing in a clinical setting to continually increase probability of diagnosis and improve clinical utility. Comprehensive coverage improvements of clinical exomes occur in three phases: annotation of targets of interest in the exome, automatic detection of inadequately covered exons, and development of methods to rescue clinically significant low coverage regions. To better characterize coverage across one popular whole exome design (VCRome2.1), targeted gene regions across 34 clinical exome samples were scrutinized. We focused on the performance of relevant diagnostic genes across the HGMD, Genetest, and COSMIC lists. Results show that >90%; of genes in the list have a base-depth coverage of 20X or better in our clinical samples. Leveraging the coverage tools developed in the process, we have implemented the Exome Coverage and Identification (ExCID) Report in our research pipeline. This workflow assesses exome coverage, annotates target regions with gene and exon coverage depth, as well as reports inadequately (<20X) covered exonic regions for every exome coming through the production pipeline. This tool allows investigators to efficiently assess performance across regions of interest. As part of a pilot project in the DNAnexus cloud platform, we are aggregating coverage plots from >5000 whole exome samples from our production pipeline. The implications of analyses at this scale are twofold: unsurpassed granularity of aggregate base coverage that leads to precise targeting and rescue of inadequately covered bases.

1438T

Tumorgraft DNASeq and RNASeq analysis. *J. Calley[1], R. Higgs[2], P. Ebert[1], T. Barber[1].* 1) Tailored Therapeutics, Lilly Research Labs, Indianapolis, IN; 2) Discovery Statistics, Lilly Research Labs, Indianapolis, IN.

Development of appropriate models for evaluation of novel oncology therapeutics remains a difficult challenge for drug development. While initial evaluation of potential therapeutics is largely carried out in monocultures of a particular cell line in vitro, xenograft/tumorgraft models implanted either subcutaneously or orthotopically are believed to provide a more realistic evaluation of the ability of the therapeutic to perform in vivo. Knowledge of the mutation status of the tumor tissue being studied is a critical component of the evaluation of efficacy in many models/histologies, particularly when the therapeutic is targeted toward a specific signaling molecule (i.e. EGFR). Determination of the mutation status of tumorgrafts cultured in mice presents unique challenges compared with either pure cell lines or primary tumors derived from humans due to mixture of the tumor DNA with mouse DNA from stromal tissue, as well as lack of a matching normal sample for 'subtraction' of germline variants. We have developed methods designed to mitigate the impact of mouse stromal DNA/RNA contamination, while potentially also allowing investigation of the unique function of the stroma in a given model.

1439F

Pathogenicity prediction of genomic variants using the gene-level frequency of variation in asymptomatic individuals. *C.A. Cassa[1], D.M. Jordan[2], M.S. Lebo[3], S.R. Sunyaev[1].* 1) Division of Genetics, Harvard Medical School/Brigham and Women's Hospital, Boston, MA; 2) Program in Biophysics, Harvard University, Cambridge, MA; 3) Lab for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, MA.

The clinical importance of novel genomic variation is often characterized using in silico techniques, which rely on the exclusion of common variation or inclusion of variants with deleterious effects predicted using evolutionary and functional considerations[1, 2]. We extend these techniques by considering the expected variation in each gene in asymptomatic individuals to predict the robustness of each gene to variation. Genes vary greatly in the number of missense and nonsense variants present in asymptomatic individuals, which provides important context about selective pressure and significance of new variants that are observed. Using data from asymptomatic individuals in the Exome Sequencing Project (N=6503), we calculate the expected numbers of heterozygous, homozygous rare, and compound heterozygous variants in each gene. We then used these data to rank-order genes, for different variant types (missense, nonsense) and minor allele frequency bins, across two populations. We then use these expected numbers of variants per gene to annotate a known list of variants that have pathogenicity classifications from a clinical genetics laboratory (Lab for Molecular Medicine). We use these variants to train a general pathogenicity classifier, using the random forest method with 10 trees and 5 attributes per tree (out of 24 total attributes)[3]. We trained and tested this classifier using 10-fold cross validation on missense variants that were classified as either pathogenic, likely pathogenic, likely benign or benign (N=3,373). The cross-validated accuracy of the classifier is over 90%, when predicting either the benign or pathogenic classes. This demonstrates that the gene-based expected variation has a high predictive value of pathogenicity. We have tested this method in additional independent datasets with similar results. We are attempting to improve these predictions using additional variant and gene-level metadata and other variant-based in silico predictive techniques, such as PolyPhen and SIFT. These findings indicate that the background variation in asymptomatic individuals has high predictive value for variant pathogenicity assessment. This technique may be used to prioritize and filter observed variants for further review by researchers or clinical geneticists. This is one of the most urgent needs in clinical genomic interpretation[4] as a number of clinical labs[5-7] and direct-to-consumer groups[8] are providing WGS interpretation.

1440W

Estimation of Isoform-specific and Allele-specific Expression from RNA-seq Data of Genetically Diverse Population. *K. Choi, N. Raghupathy, S.C. Munger, D.M. Gatti, G.A. Churchill.* Gary Churchill's group, The Jackson Laboratory, Bar Harbor, ME.

The Diversity Outbred (DO) mouse population is a new heterogeneous stock derived from the same eight inbred founder strains as the Collaborative Cross (CC) recombinant inbred strains. The DO mice have uniformly high levels of heterozygosity and genetic diversity, and thus provide a high-resolution mapping resource for identifying key genetic factors underlying complex traits and disease. Further, application of RNA-seq technology to the DO is adding an unprecedented resolution to our understanding of the transcriptome, generating valuable information about alternative splicing, allele-specific expression, RNA editing, noncoding RNAs, and strain-specific structural variation. Caution is warranted however, as the integrity of all downstream analyses depends on robust read alignment and quantitation methods.

Numerous RNA-seq analysis methods are publicly available, however they are not well-suited for aligning and quantifying gene expression in highly heterogeneous populations like the DO. Most existing tools rapidly align reads against a single reference genome or transcriptome, and then summarize the alignment results into expression abundance. But, as each DO is a unique animal with a large number of SNPs and indels, this generally-accepted strategy of aligning to a common reference genome is often error-prone. Quantitation of allele specific expression is particularly problematic as misalignment due to unaccounted strain variation is common and skew estimates in many cases.

We present a flexible statistical model for estimating haplotype-specific expression abundance in genetically diverse populations like F2, DO, or even more complex crosses. Our method incorporates both alignments to the custom founder strain genomes and external genotype probability into a hierarchical Bayesian framework. Our approach also takes into account the population-level expression abundance and allele frequency to improve quantitation at the level of individual sample. Our model offers a unique opportunity to study sequencing data from genetically diverse populations, delivering a balanced summary of expression abundance as well as allele specificity by combining the estimates of individual samples and the population as a whole.

1441T

Web-based tools to support the clinical genetics lab. *D.M. Church[1], L. Kalman[2], V. Ananiev[1], N. Bouk[1], C. Chen[1], A. Doubintchik[1], M. Halavi[1], M. Landrum[1], P. Meric[1], L. Phan[1], D. Shao[1], D. Slotta[1], J. Trow[1], M. Ward[1], D.R. Maglott[1].* 1) NIH/NLM/NCBI, Bethesda, MD; 2) CDC, Atlanta, GA.

Many clinical laboratories that once relied on Sanger based sequencing are migrating to new high-throughput and low cost sequencing technologies. While these new sequencing technologies promise improved efficiencies for genetic testing, the approach introduces new challenges for laboratories that must be addressed. The analytical validity of sequencing based approaches is of concern to clinical laboratories. It is clear that variant calling concordance rates vary between technologies and analytical pipelines. As laboratories develop and validate new tests, it is critical to provide resources that allow them to assess the false positive and the false negative rate of their tests. To facilitate this we have collected data from numerous clinical and research testing laboratories that have analyzed two publicly available human genomes, NA12878 and NA19240 based on alignment to GRCh37. In addition to collecting variant calls, we gathered details about the experimental protocols as well as the sequence evidence used to determine the variants. This data is available and can be compared to user data on a publicly available genomic browser (http://www.ncbi.nlm.nih.gov/variation/tools/get-rm). Another area of concern involves the interpretation of called variants. As laboratories move to exome and genome testing, the number of variants to be assessed can quickly become overwhelming. To address this need we have developed the Variation Reporter (http://www.ncbi.nlm.nih.gov/variation/tools/reporter). This tool allows users to upload their own variant calls and receive a report that integrates information known to databases maintained at NCBI such as dbSNP, dbVar and ClinVar. Lastly, the management of reference sequences can be a challenge because there are differences in the version of the reference assembly used by laboratories for analysis. Additionally, even when laboratories compute variation using the current reference assembly, they often wish to report variation on a RefSeqGene/LRG sequence or on a transcript or protein sequence. Using the NCBI Remap tool (http://www.ncbi.nlm.nih.gov/genome/tools/remap) users can map features from one coordinate system to another. We anticipate that this tool will be especially important with the upcoming reference assembly update (GRCh38).

1442F

Project Mercury: provenance and repeatability through encapsulation of experiments and experiment software. *N.A. Clarke, M. Pollard, I. Colgiu, J. Randall.* Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom.

It has been widely observed that, since at least the start of the century, the cost of genome sequencing has been decreasing following an exponential curve. Assuming this trend continues, the next ten years could see the cost of sequencing drop below the expected costs for storage and, potentially, for processing of the generated data. Even should we fall off the curve, we are already at the stage where access to storage and compute can form an impediment to carrying out experiments. Current operational and cost models which treat sequencing as expensive whilst largely ignoring storage and computation will prove inadequate to dealing with the increasing volume of data being produced. Independently, with more and more of an experiment taking place in software, there is a growing need to be able to accurately monitor and control the computation taking place. Where particular versions of software may have meaningful effects on an experiment's outcome, we would like to be able to capture the full processing stack: understanding how results are produced and being able to package a means of producing them for storage or publication. In contrast to Stein[1], who proposes cloud computing as the solution to the increasing volume of sequence data, we argue that this situation calls for a new operating model incorporating full knowledge of sequence, storage and compute costs. Alongside this, we will need to develop tools capable of using such a model to intelligently manage sequencing data. Such a system will require greater introspection of the computation involved in running analyses of genetic data, and as such we gain the ability to capture the steps with which to replicate our results. Project Mercury is a system under construction at the Sanger Institute to encapsulate our analysis pipelines in such a controlled and introspected environment. Through capturing various metrics (CPU hours, memory, IO, external resources) as well as recording the steps involved in an analysis we are able to quantify the cost of generating a dataset, and to transparently replicate the experiment at another time or place. This ability to replicate analyses allows our system to trade-off between storage, compute and sequencing. We discuss the design and implementation of Mercury, as well as our initial experiences using it in human genetics pipelines. [1] Lincoln D Stein: The case for cloud computing in genome informatics Genome Biology 2010, 11:207.

1443W

Refining quality control for detection of rare structural variants with SNP arrays: presenting the 'plumbCNV' R Package with an application to Type I Diabetes. *N.J. Cooper, J.A. Todd.* Medical Genetics, Diabetes and Inflammation Laboratory, Cambridge, UK.

Detection of rare copy number variants (CNVs) is increasingly seen as an important step towards understanding genetic influences on human disease. It has been shown that comprehensive quality control for Log-R-Ratio (LRR) intensity data is required to reduce the high false positive rate for rare CNV detection (Shtir et al, 2013). Such a pipeline should include examination of LRR distributions for problematic samples, followed by LRR-based principal component analysis (PCA) correction of batch effects. Once CNVs are called, additional tests should be applied to filter the set of CNVs detected. 'plumbCNV' is presented as an R package that automates these procedures. The code is open source, and scalable to deal efficiently with very large datasets exceeding system memory limits. Parameters are highly customisable and diagnostic plots and tables are produced throughout to allow the researcher to remain in control of the filters and thresholds being applied. This scripted pipeline provides a robust and necessary prerequisite for detection of rare CNVs using SNP arrays. Further assessment of CNV association with disease and phenotypes can therefore be readily and reliably conducted without using CGH or dedicated CNV arrays. The procedure is demonstrated with a comparison of rare CNV rates and characteristics between 6,292 Type I diabetes and 8,332 Healthy samples.

1444T

The Transition to Clinical NGS: How Well Do You Know Your Sequencing Pipeline? *D. Corsmeier[1], B. Kelly[1], P. White[1,2].* 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) The Ohio State University, Columbus, OH.

As next generation sequencing technologies are rapidly adopted in clinical genetics settings, higher expectations must be placed on the bioinformatics pipelines used to transform hundreds of gigabytes of raw read data into the few lines of meaningful genetic variation that is useful to the clinician. Reducing these data by several orders of magnitude is a complex process rife with the capacity for error and inadvertent deviation from a best practices approach. Notwithstanding these potential pitfalls, even if secondary analysis is performed correctly using accepted methods, parallelization and downsampling techniques can introduce nondeterminism in the resulting variant call set. This brings into question the utility of a given computational approach in a clinical setting.

Churchill, our fully automated pipeline for secondary analysis, uses one of the most popular software combinations: the Burrows-Wheeler Aligner (BWA) for short read sequence alignment together with the Genome Analysis ToolKit (GATK) for variant calling and genotyping. Using novel parallelization approaches, we have overcome the computational bottleneck created by the exponential growth in the generation of genomic sequencing data, reducing analysis time from weeks to hours. In developing an algorithm so innovative in performance and efficiency, the validation of results is a necessity and the final hurdle towards potential clinical utility.

Commited to the repeatability of test results based on discrete digitized genomic data, we investigated three pipelines that use the BWA/GATK combination for secondary analysis: Churchill, GATK Queue, and HugeSEQ. Somewhat surprisingly, we discovered that nondeterminism can be introduced at virtually every step in the analysis if configuration parameters are not carefully selected. Of the analysis approaches tested, only Churchill preserved determinism while adhering to best practices, regardless of the level of parallelization. Further, we demonstrate that Churchill's speed and precision do not come at the expense of quality of the output variant call set.

1445F

Overcoming genetic heterogeneity in rare Mendelian disease gene discovery: an improved network analysis approach. *N. Dand[1], F. Sprengel[2], V. Ahlers[2], M.A. Simpson[1], R. Schulz[1], R.J. Oakey[1], T. Schlitt[1,3,4].* 1) Department of Medical and Molecular Genetics, King's College London, United Kingdom; 2) Department of Computer Science, University of Applied Sciences and Arts Hannover, Germany; 3) Institute for Mathematical and Molecular Biomedicine, King's College London, United Kingdom; 4) Novartis Institutes for BioMedical Research (NIBR), Basel, Switzerland.

In the past few years a number of groups have undertaken exome-sequencing studies and successfully identified genes responsible for rare Mendelian diseases, typically by sequencing a small number of unrelated patients and identifying shared genetic variants after filtering out those less likely to be causal. However, this strategy is not always successful and in particular can be limited by genetic heterogeneity, where a variant in any of several genes is sufficient to cause the same disease (for example due to shared functionality). We recently presented BioGranat-IG, a software tool which addresses this problem using gene or protein networks. BioGranat-IG identifies parsimonious, highly connected components in the network, each containing variants (after filtering) for all (or most) patients. We hypothesised that these components correspond to likely disease-causing protein complexes and/or pathways. We demonstrated the validity of our method using simulated data for two rare Mendelian diseases known to have a genetically heterogeneous cause.

Since publication of this work, application to real disease data has highlighted three main areas where further improvement is possible: (i) large and highly polymorphic genes tend to be over-represented in proposed pathways; (ii) 'hub' genes, which are highly-connected in the network, also tend to be over-represented, and (iii) the correct filtering criteria for input gene lists are unclear. In the present work, we develop an entirely different algorithmic approach, using a score-based network prioritization method which removes the need for variant filtering, but retains the information from individual patients. Instead of simulating data, we use randomly rearranged exome sequence data generated at the KCL GSTT Biomedical Research Centre along with variant information obtained from the Online Mendelian Inheritance in Man (OMIM) database to generate realistic estimates of the power and type-I error for our new approach. We show significant improvement over the performance of the original BioGranat-IG tool, and propose that our new approach provides an efficient way to identify genetic mechanisms which, when disrupted, can result in disease.

1446W

SG-ADVISER tools: de-identification, identification and visual analytics of SG-ADVISER data. *G. Erikson, P. Pham, W. Shipman, A. Torkamani.* Scripps Translational Science Institute, La Jolla, CA.

Following variant annotation, a major task in genome interpretation is the identification and prioritization of relevant variants. Scripps Genome Annotation and Distributed Variant Interpretation Server (SG-ADVISER) is a web-server developed at The Scripps Translational Science Institute for the annotation of genetic variant data. To facilitate the interpretation of the SG-ADVISER output, we have developed two additional tools: SG-ADVISER UI and the SG-ADVISER Deidentification tool. The SG-ADVISER UI is a visualization tool that allows scientists with little or no programming experience to easily and quickly view, manipulate, sort, and filter the SG-ADVISER output file. A variety of custom and advanced filters allow filtering based on any combination of sample and annotation information. For example, it is possible to sort or filter on Coding Variants, Novel Variants, Cancer Genes or Chromosome Position etc. At any point during the process, the tool can provide summary statistics, and output the filtered results to a new file. The tool is written in Java and uses multi-threaded architecture and paginated view to make data analysis of large files (3GB+) possible on a desktop computer. To protect one's privacy we created SG-ADVISER Deidentificator, which is a tool that extracts the genotype(s) from the VCF files and implants clinically associated variants into the list of transmitted variants. For the clinically associated variants we used the Flagged SNPs 132 table from the UCSC Genome Browser. Identificator strips the clinically associated variants previously implanted and imports the genotype(s) back into SG-ADVISER annotated file. Both tools are cross-platform compatible and can be downloaded from the website: genomics.scripps.edu/ADVISER. This web portal can also be used for uploading de-identified data to the SG-ADVISER pipeline for annotation.

1447T

Novel set of bioinformatics tools for performing distributed down-stream data analysis in very large GWAS and NGS projects.
L. Eronen, T. Kanninen, P. Sevon. Biocomputing platforms Ltd, Helsinki, Finland.

The amount of data produced in large GWAS or NGS studies today can be measured in Terabytes. Complex data analyses, like imputing and down-stream analysis of imputed data as well as NGS data alignment and variation calling takes considerable amount of CPU time and may become a research bottleneck.

For many analysis tasks performance can be increased by data level parallelization: distributing data analyses by genomic regions, subjects and phenotypes. However, limited bandwidth between the database and external calculation resources, saturation of the disk system and task of dividing data set to smaller segments significantly limit the performance gains from distributed analysis. On the other hand, there's often a lot of manual work involved in setting up a massively parallel run.

BC|TOOLS is a software suite intended to reduce this overhead. Key components of the suite are a set of command line tools for submitting and managing distributed workflows, a workflow engine that can utilize different kinds of calculation resources (local servers, different calculation clusters or cloud), a set of file format converters that are applied automatically to make output of one analysis tool compatible with the next tool, and a set of split and merge tools for parallelizing the workload and merging results from subjobs. The system provides an API that enables easy integration of new tools.

To facilitate the massively parallel analysis of very large data sets we have developed a compressed virtual file system (BCFS), where massive data files are partitioned into tiles of manageable size, each covering only a subset of the genome, subjects and/or traits. All data files required during analyses are stored to BCFS, providing unified access to data from all calculation nodes. Physical storage may be distributed; e.g. imputation results can be stored as tiles on disks near the computation resources, while storing all clinical data on a local workstation.

In this presentation we evaluate feasibility and performance of BC|TOOLS by performing a distributed analysis workflow of a dataset with 100,000 subjects, consisting of imputation and association analysis of the imputed data in a cloud environment.

1448F

A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data.
H. Feng[1,2], K.N. Conneely[1], H. Wu[2]. 1) Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA 30322, U.S.A; 2) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, U.S.A.

DNA methylation is an important epigenetic modification that has essential roles in cellular processes including genome regulation, development, and disease, and is widely dysregulated in most types of cancer. Recent advances in sequencing technology have enabled the measurement of DNA methylation at single nucleotide resolution through methods such as whole-genome bisulfite sequencing and reduced representation bisulfite sequencing. In sequencing studies of DNA methylation, a key task is to identify methylation differences under distinct biological contexts, for example, in comparisons of tumor vs. normal tissue. A common challenge in these studies is that the number of biological replicates in each comparison group is limited due to the high costs of sequencing experiments. The small number of replicates leads to unstable estimation of within-group variation, which can reduce accuracy to detect differentially methylated loci (DML). The variance shrinkage method has been widely applied to alleviate this type of problem since the microarray days. However, the nucleotide resolution DNA methylation data are discrete and are typically modeled by distributions where the variances are dependent on means. This brings some technical difficulties to variance shrinkage. Here we propose a novel statistical method to detect DML when comparing two treatment groups in studies of nucleotide resolution DNA methylation data. The sequencing counts data are described by a lognormal-beta-binomial hierarchical model, which provides a basis for information sharing across different CpG sites. This model can be used to obtain empirical Bayes shrinkage estimates of the within-group variances, which are then plugged into a Wald test procedure for DML detection. We show through simulation that the proposed method provides greater power to detect true DML compared to existing methods, particularly when the number of replicates is very low. The proposed method has been implemented in freely available Bioconductor package DSS.

1449W

VAAST+ VSQR: making effective use of INDELs in disease-gene searches.
S. Flygare[1], B. Kennedy[1], C. Huff[2], M. Reese[3], L.B. Jorde[1], M. Yandell[1]. 1) Human Genetics, University of Utah, Salt Lake City, UT; 2) MD Anderson Cancer Center, University of Texas, Houston, Texas; 3) Omicia, Inc., Emeryville, California 94608, USA.

The Variant Annotation Analysis and Search Tool - VAAST - is a widely used tool for discovering disease-causing variants in personal genome data (Yandell et al., 2011). VAAST was used to discover a new human disease using next-generation sequencing data (Rope et al., 2011). Most recently, VAAST was used to identify the cause of Sturge-Weber Syndrome (Shirley et al., 2013). Despite its power, VAAST (and every search tool) is limited by the quality of its input data. Here we describe a new version of VAAST that addresses this problem. It works by employing VQSLOD scores provided by the GATK Variant Quality Score Recalibration procedure. We show that employing VQSLOD scores within VAAST results in a 30% increase in VAAST's power to identify disease-causing mutations, particularly disease-causing indels. VAAST is thus a powerful search tool that can effectively use indels in disease-gene searches. We have also developed a second quality measure that extends the VQSLOD approach to allow identification of what is termed 'induced false positives'. These induced false positives result from reads containing real, but rare, variants that drive the read's alignment to an incorrect region of the genome, producing falsely induced variants with misleading high VQSLOD scores. Preliminary results indicate that these scores used alone improve VAAST's power by 50%, compared to the 30% improvement seen with VQSLOD scores. Work is underway to integrate both quality scores into the VAAST algorithm, and even greater improvements are expected. As we will demonstrate, these new augmentations to VAAST make it an effective tool for identification of disease-causing variants using even the noisiest indel calls.

1450T

Sherlock: A Comprehensive Approach to Discovering Gene-Disease Associations in GWAS Using eQTL.
C. Fuller[1], X. He[2], H. Li[1]. 1) Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA; 2) Lane Center of Computational Biology, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

The genetic mapping of complex disease has traditionally relied on the identification of variations that perturb either the function or expression of nearby genes. A strong body of evidence suggests that many distal (i.e. trans-acting) loci influence gene expression, but such associations are typically ignored in the standard gene-disease association paradigm. Moreover, large numbers of potentially informative associations from genome wide association studies (GWAS) are discarded when filtering for only genome-wide levels of significance. We recently introduced a Bayesian method and online service (Sherlock) that addresses both of these issues by matching association patterns between GWAS and expression quantitative trait loci (eQTL) to implicate genes that cannot be identified by GWAS alone. Here, we present an alternate statistical test that uses the empirical distribution of GWAS p-values to identify disease genes from a panel of eQTL data sets. This new approach permits a robust comparison of association patterns without assumed priors and in the presence of the systematic inflation that is common in real-world GWAS. We discuss the results from a matrix of tests spanning a dozen diseases and numerous eQTL tissue types. We identify candidate disease genes that replicate across different input GWAS and eQTL data sets. Importantly, we note instances of correlated regulation among trans-acting loci: the risk alleles for the set of SNPs that implicate a given gene all push its expression in the same direction. In isolation, GWAS tend to generate large numbers of low-to-moderate effect loci, often poorly annotated and lacking any obvious functional consequence. Our method permits robust mapping of these loci to more functionally informative genes, providing a gain in both mechanistic insight and statistical power over the GWAS results alone. The approach is broadly applicable and could easily incorporate other molecular associations, such as metabolites, non-coding RNAs, and epigenetic modifications. It represents a straightforward means of mining both current and future data repositories (e.g. the NIH Genotype-Tissue Expression initiative) to inform research that is starved for insight amidst a sea of data.

1451F

Screening genome variants for disruption of regulatory activity with ZoomReG: capture, sequencing and computational identification of regulatory variants. *T. Gaasterland[1,2], L.E. Edsall[2,3], A.N. Dubinsky[4], R. Chappel[5], P. Ordoukhanian[5], S.R. Head[5], The NEIGHBOR Consortium.* 1) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 2) Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA; 3) Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, CA; 4) Department of Pediatrics, University of California San Diego, La Jolla, CA; 5) Next Generation Sequencing Core Facility, The Scripps Research Institute, La Jolla, CA.

Purpose: To detect, validate, and rank genome variants likely to disrupt regulatory activity, including microRNA control of protein expression, RNA-binding protein (RBP) control of alternative splicing, and DNA-binding protein (DBP) influences. Background: Genome-wide sequencing identifies many variants in untranslated regions (UTR), near intron/exon splice junctions, and near gene structures. Efficient computational screens and low false positive rates are imperative for complete interpretation of the impact in individuals or disease cohorts. Methods and Results: Genome variants with potential to alter regulatory activity are screened through three computational modules comprising the ZoomReG process. ZoomReG was applied to 420 exomes sequenced with protocols optimized for reliable near-target capture. Collectively, over 7M chromosome sites were observed as variant in the 420 exomes with 93 pct present in 1000 Genomes. Results showed ZoomReG returned known regulatory variants and ranked them reliably for degree of disruption. ZoomReG applies its ZoomMiR module to determine which UTR and coding region variants are contained within potential seed sites for microRNA::mRNA binding. ZoomReG uses the prior HOMER algorithm and a database of RBP motifs to determine whether variants overlap RBP motifs. It uses HOMER to evaluate whether SNPs overlap DNA-binding motifs and ENCODE DNase-sensitive regions. ZoomReG focuses on 50 bases upstream and downstream of each variant site and returns a disruption score that reflects whether binding centers on the site. E.g., a microRNA disruption prediction requires a SNP to appear in a satisfactory 'seed'. RBP binding is enhanced by tandem binding site motifs (e.g., YCAYCAY is stronger for Nova binding than YCAY). For RBP site disruption, SNPs in single motif matches have greater weight than in tandem matches. For a 3'UTR SNP shown in earlier work (NEIGHBOR GWAS of primary open angle glaucoma (POAG)) as associated with optic neuropathy, ZoomMiR in ZoomReG predicted a novel seed site disruption by the minor allele, validated through luciferase assay. Conclusion: ZoomReG provides critical screening functionality to identify genome variants that disrupt microRNA, RBP and DBP binding sites. Its ZoomMiR module revealed a molecular mechanism for a risk-associated SNP found in the NEIGHBOR POAG GWAS demonstrating this functionality is critical for complete genome sequence interpretation.

1452W

W4CSEQ -- a web server to process enzyme-based and sonication-based 4C-Seq data. *F. Gao[1,2], W. Lu[1], K. Wang[2].* 1) Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA 90089, USA; 2) Zilkha Neurogenetic Institute, Department of Psychiatry and Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA.

Circular chromosome conformation capture, when coupled with next-generation sequencing (4C-Seq), can be used to identify genome-wide interaction of a given locus with all of its interacting partners. Both restriction enzyme digestion and sonication methods can be used to fragment cross-linked chromatin for proximity-based ligation. We recently applied 4C-Seq to characterize the interactome of an enhancer element of Pou5f1 gene in pluripotent stem cells. We compared replicate consistency as well as similarity of 4C-Seq data generated by both enzyme-based and sonication-based fragmentation methods. In general, we found good correlation ($r>0.6$) for inter-chromosomal interactions identified by both methods when comparing biological replicate data. However, sonication generated less distal intra-chromosomal interactions compared to enzyme-based method. As note, data generated by both methods showed correlation with DNA early replication timing and enrichment of active histone marks in the enhancer interactomes. In addition, the interacting sites identified from sonication-based data are enriched with ChIP-Seq signals of transcription factors Oct4, Klf4, Esrrb, Tcfcp2i1 and Zfx that are key regulators for reprogramming and pluripotency. We found both methods are valuable tools for exploring long-range chromosomal interactions. To help bench scientists with minimal NGS experience to utilize 4C-Seq technique in the research, we have integrated our bioinformatics pipelines into a web-based server that can automatically process 4C-Seq raw data (FASTQ format) generated by a high-throughput Illumina sequencer, such as HiSeq 2000. This web server (w4cseq.usc.edu) automatically takes user specified input file(s), performs read mapping to the selected reference genome (currently only human and mouse genomes are included), calculates statistically significant interacting genomic regions, and generates relevant plots for illustration. For an input FASTQ file containing 10 million reads, the user is expected to receive the result in one day.

1453T

Multi-genome analysis improves individual genome annotation and powers disease elucidation. *G. Glusman, A.F.A. Smit, J.C. Roach, L. Rowen, D.E. Mauldin, R. Hubley, H. Li, H.C. Cox, A.B. Stittrich, S. Ament, J. Caballero, E. Bone, M. Brunkow, R. Gelinas, L. Hood.* Institute for Systems Biology, Seattle, WA.

Current genome annotation pipelines frequently yield misleading results due to sequencing technology-specific biases, errors in the reference genome and other causes. Here we introduce a suite of tools for 'multi-genome analysis' that improve the identification and interpretation of genomic variants in the context of disease. The Family Genomics group (familygenomics.systemsbiology.net) at the Institute for Systems Biology has undertaken multiple collaborative projects related to understanding the genetic basis of disease, with special emphasis on neurodegeneration. We currently have high quality whole-genome sequence (WGS) data from 1,000 individuals, produced by Complete Genomics, and funded by the University of Luxembourg (wwwen.uni.lu/lcsb). We analyze the data using custom workflows and the Ingenuity Variant Analysis platform (www.ingenuity.com/variants). Our collective WGS dataset serves as a superb resource for modeling systematic failures and biases in the technology. Use of multi-genome models improves our ability to analyze each individual genome, leading to fewer false positive and false negative findings. Simultaneous coverage analysis of several hundred genomes enables detailed normalization of the coverage profiles of individual genomes. This enables precise analysis of CNVs and the identification of large deletions that were previously undetectable. Some of the deletions we discovered explain the observed pattern of disease inheritance in the families we are studying. When analyzing personal genomes, certain genes frequently show up as mutated. We identified a set of >30,000 genomic segments significantly enriched in detrimental rare variants. Fine-grained annotation of these problematic regions leads to fewer spurious findings ~ a particular concern in clinical settings. We compiled genome-wide maps of regions prone to accumulating 'no call' failures and Mendelian and state consistency inheritance errors, to displaying extreme heterozygosity, and to being identified as 'identical by descent' among unrelated individuals. We make available several resources for improving the quality of personal genome analyses, individually and in the context of family pedigrees. The resulting improvements to sensitivity and specificity are crucial for achieving clinical-grade genome interpretation.

1454F

PhaseLift: a novel procedure to save time in imputing study genotypes.
M.M. Gorski[1,2], T.W. Winkler[1], K. Stark[1], M. Müller-Nurasyid[3,4,5], J.S. Ried[3], B.H. Weber[6], I.M. Heid[1,3]. 1) Department of Epidemiology and Preventive Medicine, Medical University Center Regensburg, Regensburg, Bayern, GermanyDepartment of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 2) Department of Internal Medicine II, University Medical Center Regensburg, Regensburg, Germany; 3) Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; 4) Department of Medicine I, University Hospital Groβhadern, Ludwig-Maximilians University, Munich, Germany; 5) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology and Chair of Genetic Epidemiology, Ludwig-Maximilians-University of Munich, Munich, Germany; 6) Institute of Human Genetics, University of Regensburg, Regensburg, Germany.

Genome-wide association (GWAs) studies usually apply imputation techniques to complement genome-wide SNP chip genotypes. The recent genotype imputation methods separate phasing from imputing. This allows for updating the imputation when a new reference panel is released without repeating the phasing step. When the study data is on an older genome build than the reference data, the current lift-over procedure is to harmonize on the genotype-level (pre-phasing lift-over) with consecutive re-phasing and re-imputing which does not fully utilize the advantage of the current two-step imputation method. We propose a novel lift-over procedure that phases genotypes on the original build and harmonizes the study haplotypes with the reference data (post-phasing lift-over; implemented in the software PhaseLift). This avoids re-phasing of study data when switching to a new reference panel. We use the KORA study (1,644 unrelated individuals) to compare imputed SNPs based on the 500K Affymetrix GWA array with additionally typed variants from the Illumina Cardio-Metabochip. To contrast the performance of both procedures, we imputed the GWA SNPs with both procedures and compare the estimated allele dosages with the additionally typed genotypes. We found that both approaches perform equally well with mean concordances of 93% and 93.4%. To see if one procedure yields higher imputation qualities (RSQs), we computed the difference of RSQs imputed by both procedures and found little difference (mean diff= 0.007, sd = 0.079). To investigate, whether the difference depends on minor allele frequency (MAF), we categorized the SNPs by low, medium and high MAF and again observed small differences (for MAF $\leq$ 0.05; 0.05 < MAF $\leq$ 0.2; MAF > 0.2: mean = 0.006, 0.008, 0.001; sd = 0.084, 0.051, 0.065). To examine, if the difference depends on RSQ, we compared SNPs imputed by both procedures, categorized by RSQ. We found that a comparable number of SNPs is poorly (RSQ $\leq$ 0.3: 78.75%, 78.78%), medium well (0.3 $\leq$ RSQ $\leq$ 0.8: 9.99%, 9.97%) and well imputed (RSQ > 0.8: 11.26%, 11.25%) in the pre- and post-phasing procedure, respectively. By using our new procedure you can save nearly 2 months of parallel computing on an 8 core-cluster, when re-imputing 3 times with the 1000G GIANT ALL reference panel. We demonstrate that our proposed post-phasing lift-over approach might encourage study partners to quickly accommodate updated reference builds to improve the information content of their data.

1455W

CIDRSeqSuite 4.0: A Toolbox for Next-Generation Sequencing Workflow Development. *S.M.L Griffith, D. Leary, J.D. Newcomer, K. Hetrick, M. Barnhart, K. Roberts, B. Marosy, B. Craig, D. Mohr, M. Zilka, A. Robinson, J. Goldstein, L. Watkins, K. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDRSeqSuite is a set of software tools that has been used to aid in next-generation sequence analysis at CIDR since 2009. A core feature of CIDRSeqSuite is a set of sequencing analysis workflows. These workflows have evolved to keep pace with changes to the predominant sequencing analysis tools used by the community, and similarly the choice of underlying technologies used in the implementation of these tools has evolved to keep pace with the larger number of analyses that must be done in parallel and in a fault-tolerant fashion. The initial workflow implemented for CIDRSeqSuite was designed to run on a single large server; samples were processed a few at a time in a single process. The current version of CIDRSeqSuite, however, focuses on individual tasks and the interdependencies among them. When an analysis is submitted to the CIDRSeqSuite server, all required tasks and their interdependencies are stored in a relational database. Such tasks include alignment, variant annotation, and even the demultiplexing of an individual tile from a flowcell lane. Tasks are submitted by the server process to an SGE-enabled cluster of over 25 nodes. As each task finishes, its status is stored in the database; any tasks whose dependencies are complete are then submitted themselves. Should a task fail to complete, it is immediately resubmitted to the compute cluster; if the number of allowable retries for a given task is exceeded, an email notification is sent to lab and informatics personnel to aid in troubleshooting. Notifications are also sent when analyses finish so that the data generated can be reviewed. Tasks can be submitted to the compute cluster via command line tools or a graphical user interface. Advantages to this distributed task-based approach include considerably faster demultiplexing: processing time is reduced from 6-8 hours to 30 minutes per flowcell. In addition to providing the apparatus for workflow development and execution, CIDRSeqSuite also encompasses several other tools including the generation of various QC reports, concordance reports (both between sets of sequencing data and between sequencing and genotyping data), and stand-alone paired-end and single-end demultiplexers.

1456T

Aneuploidy and normal cell contamination aware approach to detect copy number variations in cancer using next generation sequencing data. *R. Gupta, S. Katragadda, D. Vyavahare, K. Sandhu, V. Veeramachaneni, R. Hariharan.* Strand Life Sciences, 5th Floor, Kirloskar Business Park, Bellary Road, Hebbal Bangalore - 560024, Karnataka, India.

Background and Objectives: Recent growth in next generation sequencing (NGS) data has enabled us to detect copy number variation (CNV) at an unprecedented resolution. The objective of this study is two-fold: 1) Identify the CNV regions in the cancer genome and assign absolute copy number (CN); and 2) Compare CNV regions from different patients to identify regions that are commonly amplified or deleted, there by highlighting genes implicated in cancer. **Challenges:** Several technical and biological challenges inhibit the discovery of true segments and assignment of absolute CNs. In Particular, biological challenges include 1) Aneuploidy of cancer cells but many approaches assume diploid genome; 2) Contamination by normal and stromal cells compresses all signals towards CN state of 2; and 3) Heterogeneity in tumor cells i.e. there may be polyclonal tumors with in a tumor with each clone having different CNVs. **Methods:** Most of the approaches for detecting CNVs using NGS data are based on 1) read depth; 2) distance/orientation of read pairs; and 3) split reads. We used a method based on read depth and first compute the log-ratio of read depth in cancer and normal samples for fixed length windows, followed by Wavelet transformation of ratios to reduce the effect of random noise. An EM algorithm based probabilistic Gaussian mixture model is then built to model different CN states, and biological parameters of the sample, average ploidy and % normal cell contamination, are estimated. Finally, we used two segmentation approaches on the ploidy and contamination corrected log-ratio to obtain segments and corresponding CNs. First is naïve and heuristic approach, which quickly identifies gain/loss regions without quantifying the degree of gain or loss; and 2) popular CBS approach, which can distinguish different gain (or loss) regions. This CNV detection approach is integrated in Avadis NGS, which is our software tool for the processing and comprehensive end-to-end analysis of NGS data. **Experiments and Results:** We demonstrated the efficacy of the CNV detection approach on both simulated data and publicly available real sequencing data. For simulation set, we simulated log-ratio data to cover different scenarios by varying sample ploidy, % of normal cell contamination, number of CN states, % of data noise, etc. We also used publicly available sequencing data of cancer cell lines and tumor samples from NCBI SRA and construct CN profiles for multiple cancer types.

1457F

Assigning KIR Types From NGS Short Read Data. *T. Hague, E. Major, A. Berces, K. Rigo, S. Juhos.* Omixon, Omixon Biocomputing, Budapest, Hungary.

Killer-cell immunoglobulin-like receptors (KIRs) are one of the most polymorphic genes, and besides their sequence diversity there also a high homology among these receptors. This makes KIR typing exceptionally difficult and the picture is even more complex since not all KIR genes are present in an individual and one gene can be present in many copies. As next generation sequencing have spread to most of molecular genetic laboratories, there is a need for a method for KIR typing based on short reads. We are presenting an algorithm to determine the KIR type from NGS samples. The method is capable to predict the copy number of genes and we are also presenting quality check measures to estimate the goodness of typing.

1458W

Estimating exome genotyping accuracy by comparing to data from large scale sequencing projects. *V. Heinrich[1], T. Kamphans[2], J. Stange[3], D. Parkhomchuk[1], J. Hecht[4,5], T. Dickhaus[3], P.N. Robinson[1], P.M. Krawitz[1,5].* 1) Institute for Medical Genetics and Human Genetics, Charité Berlin, Germany; 2) Smart Algos, Berlin, Germany; 3) Department of Mathematics, Humboldt University Berlin, Germany; 4) BCRT, Berlin, Germany; 5) Max Planck Institute for Molecular Genetics, Berlin, Germany.

Next-generation sequencing (NGS) based methods, such as exome analysis, are currently being introduced as a tool for mutation detection into routine diagnostics. This increases the need for platform-independent methods of quality control. To date, quality scores have been used on a base- or genotype-level to indicate the reliability of single base or variant calls. However, the great variety of NGS platforms and analysis pipelines hinders the direct comparison of genotype-specific quality scores between platforms or for entire data sets and currently no criteria exist for assessing the overall quality of an exome. We present a genotype-weighted metric to compare all exome variants identified in a single sample together to an appropriate high-quality reference data set, with which we estimate the exome-wide genotyping accuracy based simply on the reported variants and without any further knowledge about the data generation. Our method represents a new way to evaluate the quality of entire whole-exome sequencing data in addition to current recommendations for sequencing depth and genotype likelihoods. The distance value of our metric corresponds to a quality parameter for an entire exome and allows comparing the quality of multiple exome datasets from the same or different NGS platforms. Based on simulated accuracy groups for variant calls we were able to assess the quality of an exome sample without detailed knowledge about the applied enrichment and sequencing technology or about the bioinformatics pipeline that was used to align the reads and call the genotypes. We envision that our approach to estimate the genotyping accuracy of exomes will facilitate the quality assessment of NGS data.

1459T

Highlander: variant filtering made easier. *R. Helaers, M. Vikkula.* Laboratory of Human Molecular Genetics, de Duve Institute (Université catholique de Louvain), Brussels, Belgium.

The field of human genetics is being revolutionized by exome and genome sequencing. A massive amount of data is being produced at ever-increasing rates. Targeted exome sequencing can be completed in a few days using NGS, allowing for new variant discovery in a matter of weeks. The technology generates considerable numbers of false positives, and the differentiation of sequencing errors from true mutations is not a straightforward task. Moreover, the identification of changes-of-interest from amongst tens of thousands of variants requires annotation drawn from various sources, as well as advanced filtering capabilities. We have developed Highlander, a Java software coupled to a MySQL database, in order to centralize all variant data and annotations from the lab, and to provide powerful filtering tools that are easily accessible to the biologist. Data can be generated by any NGS machine (such as Life Technologies' Solid or Ion Torrent, or Illumina's HiSeq) and most variant callers (such as Life Technologies' LifeScope or Broad Institute's GATK). Variant calls are annotated using DBNSFP and SnpEff, then imported into the database. The Highlander GUI easily allows for complex queries to this database, using shortcuts for certain standard criteria such as 'sample-specific variants', 'variants common to specific samples' or 'combined-heterozygous genes'. Users can then browse through query results using sorting, masking and highlighting of information.

1460F

Comparison of Unified Genotyper and SAMTools as variant callers across 84 exomes. *K. Hetrick II, H. Ling, E. Pugh, S. Griffith, B. Craig, B. Marosy, K. Doheny.* CIDR, Johns Hopkins Univ, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR has been utilizing samtools (v0.1.18) to perform variant calling on exomes (Barnhart, 1693/Poster, ASHG 2011). Calls were made and filtered per individual and then all unique calls within a study were back-genotyped for all samples from the samtools bcf output to create a multi-sample vcf for SNVs. We evaluated the Genome Analysis Toolkit 2 (GATK), to see if the significant advances in the GATK 2 framework (namely the addition of recalibrating indels with Base Call Quality Score Recalibration (BQSR) and creation of Reduced Reads Bams (RR) for analysis) improved variant calls. Briefly, 84 exomes (Agilent® SureSelect™ XT Human All Exon v4, Illumina® HiSeq™ 2000), were aligned with BWA-0.5.10 using the 1000 genomes phase 2 reference genome, duplicate molecules flagged with Picard (v1.74), Local Realignment, BQSR (GATK 2.1-5, 2.1-9) and RR (GATK 2.1-9) to create analysis BAM files. Multi-sample calling and filtering was done with Unified Genotyper (UG) and Variant Quality Score Recalibration (VQSR) (GATK 2.1-9). These calls were then compared to calls made from our existing samtools pipeline. The mean per sample on exon Ti/Tv for known SNVs (dbSNP 129) for both sets was ~3.2 and for novel snps, the Ti/Tv per sample mean was 2.8 for the UG+VQSR calls and 2.1 for the samtools calls. The mean per sample count of exon SNVs was 20,521 (GATK) and 20,628 (samtools). Across 9 trios in the sample set, the on exon per variant site mendel error rate was 3.15% (GATK), 2.79% (samtools). The variant concordance rate between a HapMap replicate was 99.17% (GATK), 98.20% (samtools). Across 3 siblings, the mean non-reference homozygote concordance, percent heterozygote concordance and sensitivity to heterozygote Illumina Exome Array genotypes (GATK,samtools) was 99.77%,99.37%; 99.97%,99.98%; 96.94%,98.10%, respectively. We have phased in a new workflow utilizing the GATK 2 framework which incorporates multi-sample calling/filtering with UG and VQSR and continue to evaluate all components to the NGS analysis pipeline to provide high-quality variants calls for NGS applications.

1461W

postMUT: A Statistical Tool for Combining Predictions of Missense Mutation Functionality using Capture-Recapture Methods. *S. Hicks[1], S.E. Plon[2], M. Kimmel[1,3].* 1) Department of Statistics, Rice University, Houston, TX; 2) Departments of Pediatrics and Human and Molecular Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Bioengineering, Rice University, Houston, TX.

Computational or *in silico* methods such as SIFT or PolyPhen-2 have been widely used to predict the impact of missense mutations on protein function, but these methods often report conflicting results. This leaves researchers without guidance in how to prioritize the mutations identified for further evaluation. Ad hoc combinations of *in silico* methods (e.g. at least two out of three) to prioritize missense mutations have been suggested, but these decision rules vary between groups and are not based on any statistical models. Other solutions such as 'consensus tools' have been developed based on a weighted average of scores from the individual methods to prioritize missense mutations, but these approaches do not account for the accuracy of each *in silico* method and are also not based on rigorous statistical principles. This work develops two statistical models referred to postMUT and postMUT (simple) which combine the conflicting functional predictions in a statistically rigorous manner. Both models estimate a unified posterior probability of functionality or pathogenicity for each missense mutation based on maximum likelihood estimates of the sensitivity and specificity of each individual *in silico* method in the absence of a gold standard by analyzing the subsets of data on which different algorithms agree or disagree. Our probabilistic approach requires no training set or calibration and is based on a mixture model with the weight parameter representing the overall proportion of deleterious mutations. We test the algorithms using missense mutations with known functional impact on protein function and show improved estimates of sensitivity and specificity using postMUT compared to postMUT (simple) which employs fewer parameters. We apply the postMUT models to estimate the overall proportion of deleterious mutations extracted from matched tumor/normal breast cancer genomes in categories such as normal cells only, tumor cells only, and normal and tumor cells. As the interpretation of missense mutations remains a difficult task, we show our statistical tool is scalable to the exome, provides a unified posterior probability of pathogenicity for individual variants to infer the functionality of missense mutations and can be incorporated in downstream analyses such as disease gene prioritization, ultimately inferring candidate genes. Supported by CPRIT grant RP101089, NCI grant CA155767, NCI T32 training grant CA096520 and NCN (Poland) grant 519579938.

1462T

Improved access to data sets via metadata-driven searches of experimental conditions at the new ENCODE Portal. *E.L. Hong[1], B.C. Hitz[1], E.T. Chan[1], D.T. Erickson[1], N.R. Podduturi[1], G. Roe[1], K. Rosenbloom[2], L.D. Rowe[1], C.A. Sloan[1], J.S. Strattan[1], G. Barber[2], G.A. Binkley[1], J. Garcia[2], D. Karolchik[2], .K. Learned[2], B. Lee[2], S. Miyasato[1], G. Moro[2], M. Simison[1], E. Weiler[2], W.J. Kent[2], J.M. Cherry[1].* 1) Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305; 2) Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA, 95064.

The Encyclopedia of DNA Elements (ENCODE) Project is a collaborative project to create a comprehensive catalog of functional elements in the human and mouse genomes. Now in its 8th year, the ENCODE Project has grown to include additional experimental techniques and genomic elements to survey. All experimental data (using more than 40 different experimental techniques to survey DNA-binding proteins, RNA-binding proteins, the transcriptional landscape, and chromatin structure in 400+ cell lines and tissues) and computational analyses of these data are submitted to the Data Coordination Center (DCC) for validation, tracking, storage, and distribution to community resources and the scientific community. Metadata describing important experimental conditions, such the biological samples, specific reagents, and protocols necessary to replicate the assay, have been expanded and are being submitted to a newly-formed DCC. As the volume of data increases, the identification and organization of data sets becomes challenging. Here, we describe the design principles of how metadata are organized and annotated at the ENCODE DCC in order to facilitate the identification and comparison of data sets generated by the ENCODE project. The organization of the metadata will allow flexible and powerful searches on the revamped ENCODE Portal, the public website of the ENCODE project, as well as support intuitive displays of biological samples, reagents, and protocols used for an experiment. Data from the ENCODE project can be accessed via the ENCODE portal (http://www.encodeproject.-org) and the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

1463F

Network analysis on established schizophrenia loci reveals significant genetic overlap with autoimmune and cardiometabolic disease. *L.M. Huckins, E. Zeggini.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The investigation of shared genetic factors underpinning diverse complex traits can aid the identification of common aetiopathology and give insights into shared biological pathways. We have developed a network-based method which searches for potential links between genes based on published GWAS co-citation. In order to investigate a set of association signals, we create a network in which each node represents a gene. Genes are selected on the basis of the position of the index associated variant, for example all genes within a recombination interval or genes residing either side of the signal of interest. We then create an edge between all pairs of genes which have previously been co-cited in a GWAS of any trait, or which have been cited in two separate GWAS for the same trait. We group genes into three categories; those which had been cited with a p-value lower than 10^-4, 10^-6 and 10^-8. This type of analysis is hypothesis-free, using only previous GWAS to define edges between genes. Our initial analysis focused on the database of GWAS associations reported in [1]. We tested the network analysis method on a set of 14 established schizophrenia (SCZ) loci. We computed clustering coefficients and compared these to the coefficients found when generating a million random gene sets of the same size. We found that the SCZ gene set was significantly enriched for genes which had been co-cited in GWAS of other traits at all three significance thresholds ($p=0.00015$ at the 10^{-4} association threshold). We observed overlap (4/14 genes) between GWAS of autoimmune diseases (Crohn's disease, rheumatoid arthritis, multiple sclerosis, amyotrophic lateral sclerosis, and type 1 diabetes) and SCZ. In addition, we also observed overlap between cardiometabolic-related disease GWAS loci (type 2 diabetes, glucose levels, coronary artery disease and hypertension) and SCZ (5/14 genes). Both of these findings reflect well-established epidemiological links; studies have shown a possible autoimmune etiology for schizophrenia and non-affective psychosis [2], as well as indicating a possible shared genetic basis for schizophrenia and T2D[3]. We are extending our analyses to the full NHGRI catalog of published GWAS. [1] AD Johnson, CJ O'Donnell (2009) BMC Med Genet 10:6 [2] WW Eaton et al. (2010) Bipolar. Disord. 12(6): 638-646 [3] P.I LIN, A.R. SHULDINER (2010) Schizophr res 123, no2-3, pp. 234-243.

1464W

SNP discovery in family sequencing datasets using Bayesian networks. *A. Indap, G. Marth.* Biology, Boston College, Chestnut Hill, MA.

Genome wide association studies are well powered to detect SNP variants that segregate at intermediate frequency with modest effect for complex traits. There has been increased interest in discovering rare alleles of large effect to see if these variants can explain the heritability of complex traits. Since rare alleles are potentially enriched in families with multiple affected individuals there has been increased interest in family based sequencing studies. Yet, most variant discovery methods do not explicitly model family relatedness when analyzing data. Here we describe a variant discovery method, Pgmsnp, which models the pedigree relationships and sequencing data as a Bayesian Network (BN). A BN represents the dependencies of the data as a directed acyclic graph. The application of a belief propagation algorithm to compute posterior genotype probabilities on a simulated trio dataset of 1 Mbp shows that per sample detection sensitivities (100-90%) and genotype discrepancies (0–15%) vary as a function of average read coverage (20x-5x), with the offspring sensitivity and accuracy performing the best. Further work is actively being pursued to improve the performance metrics of Pgmsnp and to apply it to different pedigree structures as well as empirical datasets.

1465T

Mapping Diagnostic Test Requisition Data to the Human Phenotype Ontology. *R. James, C. Shaw.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The clinical implementation of sequencing diagnostics encompassing coding regions (i.e., 'whole exome') has great potential to contribute to the diagnosis of human disease. The exome approach is made difficult because of the large number of variants observed in individual patients. Efficient use of available clinical information on patients may improve the curation and prioritization of observed variants. Unfortunately, phenotypic clinical inputs for genetic testing are rarely complete and often do not strictly adhere to a controlled vocabulary. We hypothesized that natural language processing applied to clinical notes, indication forms and/or elements of medical records can standardize and improve indication data. We used these techniques in conjunction with hierarchical semantic relationships of The Human Phenotype Ontology to develop methods for processing indication data. Improved similarity metrics were developed and systematically evaluated via their performance in Monte Carlo simulations of possible feature combinations that may occur in patients being sent for genome wide diagnostics. These methods can be applied to analyze the correspondence between the requisition content and the ultimate diagnosis obtained. We conclude that such approaches are useful to improve the speed and accuracy of exome analysis.

1466F

Multiple Testing Correction method for Linear Mixed Model. *J.W. JOO[1], E. Kostem[2], E. Kang[2], B. Han[3,4], E. Eskin[1,2,5].* 1) Bioinformatics phD program, University of California, Los Angeles, Los Angeles, CA, USA; 2) Computer Science, University of California, Los Angeles, Los Angeles, CA, USA; 3) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Department of Human Genetics, University of California, Los Angeles, CA, USA.

Multiple hypothesis testing is a major issue in genome-wide association studies which often analyze millions SNPs. There are several methods to estimate a per-marker p-value threshold in order to obtain a significant result given multiple markers. The permutation test is widely believed the gold standard for multiple testing correction for traditional association tests and accurately takes into account the correlation structure of the genome unlike the traditional Bonferroni correction. Since the permutation test is computationally very expensive, several alternative methods have been presented to speed up the permutation test. Recently, mixed model association has became a popular approach for genome wide association studies as it can correct for population structure. Population structure complicates association analysis by inducing spurious correlation between genotypes and phenotypes and may cause false positive associations. Unfortunately, permutation is not applicable for mixed models because if we either permute phenotypes or genotypes, the true correlation structure between individuals is eliminated and leads to an inflation of p-values. Theoretically, bootstrapping can be applied to mixed model, however, it is often computationally impractical for large datasets. In this paper, we propose an efficient and accurate multiple testing correction method for linear mixed model, slideLMM. The key idea behind slideLMM is that we utilize the kinship matrix to transform the correlation structure between the variants and then take advantage that the statistics will then follow a multi-variate normal distribution. Utilizing this approach, slideLMM corrects for the population structure to give an accurate per-marker threshold even the individuals have different degrees of relatedness. Applied to a the Hybrid Mouse Diversity Panel(HMDP) which is a mouse association study panel with significant amounts of population structure, our method shows better performance in both speed and accuracy compared to previous methods.

1467W

Filtering for compound heterozygous sequence variants in non-consanguineous pedigrees. *T. Kamphans[1], P. Sabri[1], A. Knaus[1], V. Heinrich[2], P. Krawitz[2], P. Robinson[2].* 1) Smart Algos, Berlin, Germany; 2) Medical Genetics, Charite, Berlin, Germany.

The identification of disease-causing mutations in next-generation sequencing (NGS) data requires efficient filtering techniques. In patients with rare recessive diseases compound heterozygosity of pathogenic mutations is the likeliest inheritance model if the parents are non-consanguineous. We developed a web-based compound heterozygous filter that is suited for data from NGS projects and that is easy to use for non-bioinformaticians. We analyzed the power of compound heterozygous mutation filtering by deriving background distributions for healthy individuals from different ethnicities and studied the effectiveness in trios as well as more complex pedigree structures. While usually more then 30 genes harbor potential compound heterozygotes in single exomes, this number can be markedly reduced with every additional member of the pedigree that is included in the analysis. In a real data set with exomes of four family members, two sisters affected by Mabry syndrome and their healthy parents, the disease-causing gene PIGO, which harbors the pathogenic compound heterozygous variants, could readily be identified. Compound heterozygous filtering is an efficient means to reduce the number of candidate mutations in studies aiming at identifying recessive disease genes in non-consanguineous families. A web-server is provided to make this filtering strategy available at www.gene-talk.de.

1468T

Big Data Management in the Era of Genomic Medicine. *J. Kaufman, S. Fuchs, D. Healey, A. Hopkins, L. Louie, D. Maltbie, C. Martin, M. Penley, M. Wong.* Annai Systems, Los Gatos, CA.

As sequencing technologies continue to evolve and the use of sequencing data makes its way from research into the clinic and hospital, the proliferation of data will continue to accelerate. With this trend and the application of this data to personalized medicine, new challenges in data storage, sharing, security, analysis and retrieval of information will arise. While many of these issues are only now starting to be addressed and anticipated, there is a considerable dearth of readily available solutions to these problems. The creation of data repositories capable of managing genomic information in a manner that enables streamlined access to data has emerged as a critical requirement as the application and use of such data progresses. One highly relevant example of a data repository solution that fulfills multiple needs for a variety of different users is The Cancer Genomics Hub (CGHub). CGHub is a genomic data repository built to support all three major NCI cancer genome sequencing programs: TCGA, TARGET, and the CGCI. CGHub was launched in 2012, hosted by UC Santa Cruz and with only TCGA data online, has more than 44,000 data files totaling more than 500 Terabytes with capacity to grow quickly to 5 Petabytes as additional datasets become available. CGHub is co-located with a biocompute farm that enables cancer researchers the ability to seamlessly access the data files for subsequent analysis using a variety of commercially and/or freely available tools. This repository was built with products and technologies developed at Annai Systems. In the CGHub example, data are stored in a vast public repository enabling widespread access to a large number of researchers and clinicians. There are also a growing number of smaller sized private repositories used to inform drug discovery, disease diagnosis and patient treatment. While some of the requirements of these repositories are quite similar to one another, there are a number of differences with respect to how the data are used, who will access it, and what type of regulatory and security considerations must be adhered to. Because of these differences, having a set of tools that can be used to provide flexible, scalable solutions that can address multiple use cases is of paramount importance. We will discuss how Annai Systems' portfolio of products and services can be used to create 'big data' repositories that can facilitate access to, and sharing of, sequence data and related meta-data.

1469F

Churchill: A Cloud-Enabled, Ultra-Fast Computational Approach for the Discovery of Human Genetic Variation. *B. Kelly[1], J. Fitch[1], D. Corsmeier[1], D. Newsom[1], P. White[1,2].* 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) The Ohio State University, Columbus, OH.

Next generation sequencing (NGS) has revolutionized genetic research, empowering dramatic increases in the discovery of new functional variants. The technology has been widely adopted by the research community and is now seeing rapid adoption clinically, driven by recognition of NGS's diagnostic utility and enhancements in quality and speed of data acquisition. Compounded by declining sequencing costs, this exponential growth in data generation has created a computational bottleneck. Current analysis approaches can take weeks to complete resulting in bioinformatics overheads that exceed raw sequencing costs and represent a significant limitation for those utilizing the technology. Churchill is a computational approach that overcomes these challenges, fully automating the analytical process required to take raw sequencing instrument output through the complex and computationally intensive processes of alignment, post-alignment processing, local realignment, recalibration and variant calling. Through implementation of novel parallelization techniques we have dramatically reduced the analysis time for whole human genome resequencing from weeks to hours, without the need for specialized analysis equipment or supercomputers. As increasing numbers of molecular diagnostic laboratories implement NGS in clinical settings, Churchill provides a solution to the data analysis challenges these laboratories will immediately face. Compared with alternative analysis pipelines, Churchill is simpler, faster, deterministic and capable of running on all popular Linux environments. Furthermore, Churchill optimizes utilization of available compute resources and scales in a near linear fashion, enabling complete human genome resequencing analysis in ten hours with a single server, three hours with our in-house cluster and under two hours using a larger HPC cluster. Churchill is cloud-compatible and we demonstrate the expansive degree of parallelization Churchill can achieve using Amazon's Elastic Cloud Compute (EC2) instances. Not only does this allow laboratories to potentially reduce analysis time by leveraging the cloud's ability to easily scale, but it also enables low-cost resequencing analysis without needing to invest in the required infrastructure to build their own high-performance computing cluster. Churchill eliminates the NGS bioinformatics overhead and is a prime candidate to overcome the bottleneck even faster sequencing will create.

1470W

Accessing more human genetic variation with short sequencing reads. *C. Kennedy, N. Chennagiri, B. Breton, M. Umbarger, P. Saunders, V. Greger, C. Micale, G. Porreca.* R&D, Good Start Genetics, Cambridge, MA.

Despite advancements in analysis methods, considerable human genetic variation remains inaccessible to short read sequencing. We combine targeted assembly and alignment of short reads with coverage- and base quality-enabled genotyping on the Amazon Elastic Compute Cloud to re-analyze exome data from the 1000 Genomes Project, benchmarking our results against Sanger sequences for a subset of samples and disease associated genes. Our results demonstrate high detection accuracy for variants, including insertions and deletions, up to 99.7% sensitivity with 1.7% false discovery. Applying our method broadly revealed significant population-dependent allelic diversity in regions previously considered inaccessible to variant detection. Specifically, microsatellite polymorphisms and pathogenic alleles in genes associated with Niemann-Pick disease, maple syrup urine disease, Usher syndrome, cystic fibrosis, and heritable amyotrophic lateral sclerosis. Scalable computational resources and highly accurate analyses, such as those described here, should accelerate clinical adoption of short read sequencing technologies by improving the power to detect variants.

1471T

Genomics in Clinical Research on the DNASTAR Cloud. *M. Keyser[1], J. Carville[1], T. Schwei[1], T. Durfee PhD[1], A. Pollack-Berti PhD[1], D. Nash[1], J. Stieren[1], S. Baldwin[1], R. Nelson PhD[1], K. Dullea[1], J. Schroeder[1], P. Pinkas PhD[1], G. Plunkett III PhD[1,2], F. Blattner PhD[1,2,3].* 1) DNASTAR, Inc., Madison, WI; 2) University of Wisconsin, Department of Genetics, Madison, Wisconsin, USA; 3) Scarab Genomics LLC, Madison, Wisconsin, USA.

DNASTAR offers an integrated suite of software for assembling and analyzing sequence data from all major next-generation sequencing platforms supporting key workflows on both a desktop computer as well as the DNASTAR Cloud. The cancer genomics workflow that integrates some of the most powerful functionality in the software includes assembling and analyzing multiple samples using one reference template; probabilistic identification of SNPs, small indels and genotype calls with known variants correlated to their dbSNP and COSMIC IDs and GERP reference data; review and filtering of SNPs from multiple samples within a single project; identification of structural variations; and, for large multi-sample projects with hundreds of individual data sets, tools for SNP quantitation, filtering, set comparison, clustering and indication of the gene disruption impact from called SNPs. In addition, DNASTAR offers multi-sample copy number variation reporting for further analysis. Interactive views within the software facilitate fast, comprehensive analysis, helping scientists move quickly from raw next-gen sequencing data to genetic and genomic impact, including gene ontology. By using innovative algorithms within the software, scientists can have all of the assembly and analysis capabilities available to them on either their desktop computer or the DNASTAR Cloud, supporting large data sets generated by any or all of the next-gen sequencing instruments and platforms.

1472F

A general framework for estimating the relative pathogenicity of human genetic variants. *M. Kircher[1], D.M. Witten[2], G.M. Cooper[3], J. Shendure[1].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Hudson Alpha Institute for Biotechnology, Huntsville, AL.

As genetic information is insufficient to unambiguously implicate many disease-causal variants, annotations that enrich for causal variation are essential. Current annotations tend to exploit a single information type (e.g. conservation) and/or are restricted in scope (e.g. to missense changes). A broadly applicable metric that objectively weights and integrates diverse information is needed. Here, we describe Combined Annotation Dependent Depletion (CADD), a framework that integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations. We implement CADD as a support vector machine, trained to use 63 annotations to differentiate 14.7 million variants derived on the human lineage from 14.7 million simulated variants. We pre-compute CADD-based scores (C-scores) for all 8.6 billion possible single nucleotide variants of the reference genome and enable scoring of short insertions/deletions. C-scores strongly correlate with allelic diversity, pathogenicity of both coding and non-coding variants, and experimentally measured regulatory effects, and also highly rank causal variants within individual genome sequences. Finally, C-scores of complex trait-associated variants from genome-wide association studies (GWAS) are significantly higher than matched controls and correlate with study sample size, likely reflecting the increased accuracy of larger GWAS. Thus, the ability of CADD to quantitatively prioritize functional, deleterious, and disease causal variants across a wide range of functional categories, effect sizes and genetic architectures is unmatched by any current annotation and will be widely useful for the identification of causal variation in both research and clinical settings.

1473W

Gene-set test for rare variants. *J. Lee, B. Min, S. Lee, J. Ka, T. Park, T2D-Consortium.* Seoul National University, Seoul, South Korea.

Recently, many methods have been developed for the association studies of rare variants. However, current rare variants studies have focused only on gene-level association with diseases. We investigate gene-set-level association in order to sum up the effect of rare variants on diseases from the same gene set. We first present a new unified quadratic test which is shown to be generally more powerful than or as powerful as other tests. Using this quadratic test, we then develop the gene-set-level association tests for rare variants. This gene-set test is capable of handling the different directions of effects. It can also perform association test for the data including both common and rare variants. Through simulation studies, the proposed test was shown to outperform other rare variants tests in terms of average power under the various scenarios. We applied the proposed gene-set test to discover the association between the lipid-related traits and 2.5 million variants from 10,000 exome sequencing data from five major ethnic groups (African American, East Asian, European, Hispanic and South Asian) in T2D-GENES consortium. Using 1452 canonical pathways from MSigDB (v3.1) database, we identified several known gene sets at the 5% significance level after Bonferroni adjustment, which are related to the lipid including asparagine(N)-linked glycosylation, lipoprotein metabolism, and HDL-mediated lipid transport pathways. Within these pathways, B4GALT2 and PLTP were identified as most significant genes.

1474T

Monitoring, analyzing, and exploring Ion Torrent™ NGS data with Torrent Suite™ Software. *D. León.* Ion Torrent by Life Technologies, South San Francisco, CA.

Torrent Suite™ Software is specifically designed to allow Ion PGM™ and Ion Proton™ Instrument users to monitor, analyze and explore Ion Torrent™ NGS data. This report illustrates how instrument users can plan a sequencing run that includes not only instrument settings, but it also enable users to preset configurations for data analysis, variant calling and exporting to other annotation software. The Torrent Suite™ Software is accessed through the Torrent Browser, the associated web-based interface for monitoring, planning, reviewing and managing sequence data in real time. As sequence data are being generated, users can monitor a sequencing run and the quality of their data. After a sequencing run is complete, the main data analysis step of the software includes base calling and mapping of the sequence data to a reference genome. Variant calling for homozygous/heterozygous SNPs and small/large indels is the next automated step in the data analysis process with interactive tables of identified variants may be visualized and evaluated using Broad Institute's Integrative Genomics Viewer and two validation assay search sites (TaqMan® SNP Genotyping Assays and PCR/Sanger Sequencing Primers) on Lifetechnologies.com. The architecture of the software also enables users to create plugin applications that perform custom analyses and can connect to third-party providers. The plugins that are included in the Torrent Suite™ Software installation include: Alignment,CoverageAnalysis,ERCC_Analysis,FastQCreator,FilterDuplicates,IonReporterUploader, RunRecognitION,SampleID,SFFCreator,TorrentSuiteCloud,and variantCaller. In addition to providing data quality information and summary analysis reports, Torrent Suite™ Software provides customers the option to further annotate and filter their list of variants. Specifically, they can choose to run a pre-set workflow with Ion Reporter™ software after a sequencing run is complete. This step in the analysis can be pre-configured in the Torrent Browser when planning a sequencing run. This optional annotation step provides biological associations for the variants of interest and allows customers to extend the data analysis capabilities of Torrent Suite™ Software. In summary, this report demonstrates, how with a single web interface, Torrent Suite™ Software enables Ion Torrent ™ instrument users the ability to start a sequencing run by setting up the sequencing parameters and the desired downstream analyses. (Research Use Only).

1475F

FERRET: a User Friendly Tool to Quickly Extract Data from the 1000 Genomes Project. *S. Limou, G. Nelson, P. An, C. Winkler.* Basic Science Program, SAIC, Frederick National Laboratory for Cancer Research, Frederick, MD.

By sequencing individual genomes from several reference populations, the 1000 Genomes (1KG) Project provides a valuable and near-comprehensive resource on human genetic variation for the scientific community. Raw and annotated data are regularly and rapidly released on an ftp server, and variants may be accessed directly through the 1KG browser. Even though the 1KG browser is well-designed, accessing data of interest can become a tedious process requiring both a good knowledge of the website architecture and many clicks. Data accessible from the 1KG browser are not updated as quickly as the data released into the ftp server. Further, the 'VCF to PED' tool available from the 1KG browser does not support indels nor provide variant frequencies. We developed Ferret, a user-friendly tool to quickly extract data from the latest release on the 1KG ftp server. Ferret is a Perl script parsing the 1KG vcf files to extract allelic frequency and genotype data for each variant (SNP and indel) located in the region of interest for the reference population of interest. Genotype data are recapitulated into map, ped and info files, which may be loaded in PLINK or HaploView for further exploration of linkage disequilibrium pattern, haplotypes, and eventually tagSNP design for customized genotyping arrays. The main advantages of Ferret are (1) the handling of indels, (2) the user-friendly interface, (3) the output format, and (4) calculation of allele frequencies. Ferret is thus a straightforward program, even for non-specialists who are not adept at the 1KG bioinformatics tools: this software permits easy manipulation and visualization of 1KG data with well-known pre-existing tools. Ferret is publicly available at: https://ccrod.cancer.gov/confluence/display/BCGC/BCGC+Software. You may contact the first author (during and after the meeting) at sophie.limou@nih.gov.

1476W

dbNSFP v2.0: A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations. *X. Liu, X. Jian, E. Boerwinkle.* Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX.

dbNSFP is a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants (nsSNVs) in the human genome. This database significantly facilitates the process of querying predictions and annotations from different databases/web-servers for large amounts of nsSNVs discovered in exome-sequencing studies. Here we report a recent major update of the database to version 2.0. We have rebuilt the SNV collection based on GENCODE 9 and currently the database includes 87,347,043 nsSNVs and 2,270,742 essential splice site SNVs (an 18% increase compared to dbNSFP v1.0). For each nsSNV dbNSFP v2.0 has added two prediction scores (MutationAssessor and FATHMM) and two conservation scores (GERP++ and SiPhy). The original five prediction and conservation scores in v1.0 (SIFT, Polyphen2, LRT, MutationTaster and PhyloP) have been updated. Rich functional annotations for SNVs and genes have also been added into the new version, including allele frequencies observed in the 1000 Genomes Project phase 1 data and the NHLBI Exome Sequencing Project, various gene IDs for different databases, gene functional description, domains/conserved sites, pathways, tissues/organs the gene expressed in, other genes the gene interacted with, estimated probability of haploinsufficiency/recessive-disease-causing, etc. A companion java program is provided for quick local query of SNVs/positions/genes with support of using a vcf file directly as input. dbNSFP v2.0 is freely available for download at http://sites.google.com/site/jpopgen/dbNSFP.

1477T

NCBI's Conserved Domain Database as a tool for the interpretation of human sequence variation. *F. Lu, C.J. Lanczycki, Z. Wang, G.H Marchler, M. Gwadz, S.H. Bryant, A. Marchler-Bauer.* National Center for Biotechnology Information, Bethesda, MD., USA.

NCBI's Conserved Domain Database (CDD) is a protein classification and annotation resource. Manually crafted multiple sequence alignments (MSAs) are the foundation for domain models that are annotated with functional descriptions, links to the literature, and functional sites based on 3D protein structure or experimental evidence. Functional sites often have highly conserved sequence and/or structure and are closely linked to particular structural or functional roles. We investigate the prevalence of single nucleotide polymorphisms (SNPs) with human diseases at such functional sites and demonstrate how pre-computed sequence annotation available from CDD may assist in understanding phenotypes caused by sequence variation. We employ the medically important nuclear receptor (NR) family as an example for this analysis. Nuclear receptors are ligand-modulated transcription factors that act in the nucleus to regulate target gene transcription through the interaction of cofactor proteins. Aberrant receptors result in a wide range of prevalent human diseases and disorders, such as cancer, diabetes, obesity. About 10% of the most prescribed drugs act through nuclear receptors, reflecting their medical importance. Nuclear receptors share a common structural organization with a central well conserved DNA binding domain (DBD), a variable N-terminal domain, a non-conserved hinge and a C-terminal ligand binding domain (LBD). We have classified LBD and DBD into approximately 50 unique subfamilies. The domain models representing these subfamilies were manually curated to add functional sites annotations, specifically ligand binding sites, DNA binding sites, co-activator/repressor binding sites and residues involved in oligomerization. Genes corresponding to the human members of the NR family and non-synonymous SNPs collected for these genes were extracted from the NCBI's Entrez database. The SNPs were mapped to the domain footprints and functional sites annotated by CDD. We report on the correlations between genetic variations that overlap the annotated functional sites and human diseases.

1478F

H3M2: a novel algorithm for the detection of Runs of Homozygosity from second generation sequencing data. *A. Magi[1], L. Tattini[1], P. Flavia[2], M. Benelli[3], M. Seri[4], G.F. Gensini[1], G. Romeo[4], T. Pippucci[2].* 1) Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy; 2) U.O. Medical genetics Polyclinic Sant'orsola malpighi, Bologna, Italy; 3) Diagnostic Genetic Unit, Careggi Hospital, Florence, Italy; 4) Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy.

Runs of Homozygosity (ROHs) are genomic stretches that appear in homozygous state, showing both alleles as identical in a diploid genome. ROHs are medically relevant. In rare autosomal recessive disorders (ARD), a homozygous highly penetrant mutation will probably reside in an unusually long homozygous haplotype originated from recent consanguinity. In complex disorders, recessive risk variants can be associated to shorter ROHs in inbred/outbred populations. To date, ROH discovery has been performed using microarray-based technologies. We introduce a novel computational approach based on heterogeneous hidden markov model, H3M2 (Homozygosity Heterogeneous Hidden Markov Model), for the identification of ROHs from HTS data. To develop this method, we used whole-exome sequencing (WES) data produced by the 1000 genomes project (1000 GP) consortium previously genotyped by the HapMap consortium. As a mesure of homozygosity/heterozygosity for each polymorphic position i, we used the B-allele frequency (BAF) that is defined as the ratio between B-allele counts (NB the number of reads that match with the allele with minor frequency at position i) and the total number of reads mapped to that position (N, the depth of coverage). Then, based on the population allelic frequency of all the polymorphic positions tested, we calculated for each detected homozygous region a so-called Regional Diplotype Score (RDS) as a measure of statistical significance. To test the ability of our algorithm to identify ROHs of different size and/or number of SNPs, we applied it to WES data of six individuals sequenced by the 1000GP Consortium. H3M2 was able to identify more than 90% of the ROHs detected with HapMap SNPs array data, outperforming state of the art methods. We also applied the algorithm to 100 1000GP individuals from different populations, showing that the total lengthof ROHs per individual increases with the distance of a population from East Africa, in agreement with previous observation. Finally, we tested the algorithm on 10 offspring to consanguineous families affected with ARD. In those subjects for whom the disease-mutation was known we found that the RDS of the surrounding region ranked among the highest subject's ones (>95%ile), demonstrating that the proposed algorithm is a valid tool also for disease-locus mapping. H3M2 is a highly performing algorithm aimed at the identification of differently sized ROHs from HTS data, in populations as well as in individuals.

1479W

PRIOR: a bioinformatics tool for prioritization of candidate genes. *V. Makarov, I. Ionita-Laza.* Mailman School of Public Health, Columbia University, New York, NY.

Increasing volumes of next-generation sequencing data and the growing costs of experimental validation of detected changes call for the development of bioinformatics tools that can help prioritize genes for further analysis. We have developed a computational pipeline (PRIOR) that combines two published bioinformatics tools (snpEff and AnnTools) with additional methods developed by our group. Variant annotation is performed with the snpEff and AnnTools and parsed with the AnnTools' parser. The effects with the highest impact are considered candidates for experimental validation, while all other effects are preserved for possible future studies. The pipeline accepts input data in the standard VCF format and annotates the variants to include only novel (not in the dbSNP), potentially disease-causative mutations in the coding regions, which include non-synonymous single nucleotide substitutions (SNP/SNV) and short insertions/deletions (INDEL). The tool makes predictions of the potential importance of each gene based on the overall mutation rate in a study cohort, and calculates the mutation rate for each gene adjusted, firstly, for its length and, secondly, for mutation rate in general population based on the rate of novel SNP/SNV and INDEL reported by NHLBI Exome Sequencing Project (ESP) in 6500 exomes. For studies that involve related individuals, further adjustment is made to add more weight to variants found in not related individuals compared with those found in the family members. Additional annotation for each gene includes function description, functional pathways, and known disease association as reported in the dbNSFP database. The application is freely available for public use; the package includes installation scripts and a set of helper tools.

1480T

Shortcut to Analysis of Genomic Data. *D. Maltbie, S. Fuchs, D. Healey, A. Hopkins, J. Kaufman, L. Louie, C. Martin, M. Penley, M. Wong.* Annai Systems, Los Gatos, CA.

Finding and accessing the ever-growing amount of genomic data and meta-data can be difficult for researchers and clinicians. Once a target data set is identified, extracting that data set from repositories and running analyses can tax or overwhelm available information technology resources such as compute power, storage and networks. By leveraging specialized software tools and an information technology infrastructure designed specifically for genomic data, commonly encountered challenges can be overcome, resulting in a shortcut to data analysis. As organizations struggle with increasing demands from researchers for access to high performance compute platforms, they often experience reduced throughput during peak times, coupled with hardware that is woefully underutilized during non-peak times. Network limitations and firewall issues, manifested as lengthy, error prone Internet downloads are common when migrating datasets from external systems to compute centers. At the compute centers, temporary storage can be inadequate for the ever-increasing amount of data that must be migrated in preparation for analysis, creating yet another obstacle to the utilization of genomic data. A shortcut is needed to overcome these ubiquitous problems in compute, storage and networks. One solution is setting up optimally architected data repositories co-located with high performance compute centers on highly efficient networks. Annai Systems' web based reQuest data portal plus GNOS web services is one such solution. Researchers can locate target datasets through customized searches in reQuest across multiple GNOS enabled repositories. The GTFuse access tool allows access to specific regions of data without downloading the whole genome sequences. For example, TCGA credentialed researchers can swiftly compile datasets and run analyses on Annai System's BioCompute Farm. This is a GNOS web based service that provides elastic compute and storage, and is co-located with the Cancer Genomics Hub, the largest cancer genome repository. Scientists using the Cancer Genome Atlas data via the reQuest portal and working in the BioCompute Farm with GTFuse are able to extract specific gene segments from hundreds of whole genome sequences in hours as opposed to the days or weeks it could otherwise take. We will discuss how scientists and clinicians are empowered by shortcuts to finding, accessing and analyzing genomic data through the use of the tools described above.

1481F

High Throughput Exome Coverage and Capture of Clinically Relevant Cardiac Genes. *D. Manase[1,3], L. D'Alessandro[2], A.K. Manickaraj[1,3], S. Mital[2].* 1) Heart Centre Biobank Registry , Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Cardiology, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada.

Background: High throughput whole-exome sequencing (WES) technologies have been used to identify variants in human disease. However, variant identification is limited by the overall coverage provided by WES. We sought to quantify the overall capture and read depth (RD) coverage of clinically relevant cardiac genes in WES data obtained using state of the art sequencing platforms. Methods: Data from 200 human exomes sequenced via Illumina HiSeq using Agilent SureSelect for capture were obtained in raw sequence format and aligned to the human genome version 19 using Burrows-Wheeler Aligner. Of these, 45 BAM files were analyzed using SAMTOOLS. Fifty clinically relevant cardiac genes were selected for analysis including 31 genes on the American College of Medical Genetics list for reporting of incidental findings and an additional 19 genes associated with congenital heart disease. Gene coordinates were obtained for all protein coding regions on Ensembl and filtered for unique chromosomal locations. Both the RD across all coding exons and the BED file coverage from the sequencing capture kit were compared with the reference genome to assess observed versus potential capture. Results: Preliminary results of 50 cardiac genes analyzed on 45 exomes sequenced at 50X coverage revealed that a RD coverage of 50X was seen only on 24% of all exonic regions. RD between 0 to 9 reads was the predominant coverage observed in these genes occurring at 41% of all exonic regions. Of note, none of the 50 genes had more than 70% of all isoform regions covered to a RD of 50X or greater, and 1 gene had 100% of its isoforms covered at less than 10X reads. APOB and MYH11 were the only genes with at least half of all regions covered ≥ 50X. ACTC1, CFC1, GATA4, KRAS, NRAS, PTPN11, SMAD3, SOS1, TBX5, TGFBR1, TGFBR2, and TMEM43 had the least coverage in RD with 50% of the regions covered by only 0-9 reads. Conclusion: Low capture and read depth may impact variant analysis and interpretation, including false negatives wherein potential heterozygous mutations may not present within the reads that are sequenced while true calls may be ignored through low read depth filtering or even interpreted as artifact or misaligned read. A study bias towards well-covered genes may result in greater attribution of disease causation to mutations in well-represented genes with more confident calls. Improvements in WES technology are needed before widespread clinical use of exome data.

1482W

A Cloud Computing Based Open Science Platform for Analysis, Presentation, and Collaboration in Bioinformatics. *G. Manglik, J. Yeager.* CliQr Technologies, Sunnyvale, CA.

As the cost of sequencing continues to drop, the amount of bioinformatics data researchers need to manipulate and store increases exponentially, creating an ever-increasing demand for the computational methods required to manipulate data, perform statistical analyses, create graphics, and share results. For the majority of researchers this involves selecting the most relevant packages and either purchasing proprietary software or finding and compiling an appropriate open source equivalent. Maintenance of any such setup can become burdensome and the software support role can grow to dominate the time spent on actual research. We have developed a new and better method for quickly importing software applications to any Cloud computing provider such as Amazon AWS, HP Cloud, Rackspace, or Microsoft Azure, and have used this to create an Open Science Platform for use by researchers. This platform takes on the burden of compiling and supporting the common applications. Using the CliQr Open Science Platform researchers can share data and results with collaborators in an easy and unprecedented manner. Using our web browser based platform the bioinformaticist can create and import applications in addition to the open source tools within the Open Science Platform. The bioinformatist also has full access to the underlying Cloud computing system if desired. The portal enables the bioinformatics researcher to bypass maintaining and compiling software and concentrate on their investigations. The Open Science Platform can significantly reduce the analysis timeframe and cost to researchers. Pricing for the open source software is based upon the compute and storage resources consumed, reducing the barrier to use significantly and encouraging experimentation. We will discuss the technical challenges in the creation of this platform, issues around Cloud computing and bioinformatics, and how computation and storage requirements drive choice of Cloud. Use of multiple applications taking data from its raw form through to final presentation will be demonstrated. The benchmarking capabilities of the Open Science Platform will be presented, and the rapid transition of applications between Clouds to allow for optimum performance or cost behavior.

1483T

Liftover of Short Reads from Next Generation Sequences Aligned to Non- Canonical References as a Step towards a Diploid Alignment.
T.C. Markello[1,2], T.R. Gall[1], A.J. Brandt[1], P.J. Pemberton[1], D.R. Simeonov[1], E.D. Flynn[1], W.P. Bone[1], A.E. Links[1], D.R. Adams[1,2], C.F. Boerkoel[1], W.A. Gahl[1,2,3]. 1) Med Gen Branch, NIH/NHGRI, Bethesda, MD; 2) Undiagnosed Diseases Program, NIH, Bethesda, MD; 3) Office of the Clinical Director, NHGRI, NIH, Bethesda, MD.

Using a canonical reference genome for aligning short reads to a template is less computationally intensive than *de novo* assembly, but the alignment may be inaccurate if the reference is too dissimilar to the genome from which the reads were derived. Previous work has shown that it is possible to modify a canonical reference to more closely match the unknown sequence by applying corrections derived from hybridization experiments or imputed from population based haplotypes. The biggest unsolved challenge is to relate the alignment performed onto a modified reference back to the canonical reference. This is required in order to annotate the discovered variants for their significance to cause disease. A chain file contains all the information required for mathematical transformation of coordinates between references and is generated when modifying a genome sequence. If modifications are restricted to changing nucleotides from one base to another, then there is a one to one correspondence between reference positions. When modifications generate length discrepancies between the two sequences by including copy number variations, then positions in the modified reference must be lifted over back onto the canonical reference in chain blocks. After short reads are aligned to a modified reference their record includes a Compact Idiopathic Gapped Alignment Report (CIGAR) string that can span two or more chain file blocks and require modification to account for length differences during the liftover process. The bioinformatics group of the Undiagnosed Diseases Program (UDP) has written and implemented an extension to existing bioinformatic suites that preforms the CIGAR string liftover function. We have extended this code to liftover the complex variant combinations arising from the correction of alignment references in a sequenced parent-parent-child trio. We present an analysis of the challenges associated with this lift-over process, including examples of local alignments that are improved by our current process. A successful liftover of short reads aligned to a modified reference, or to reads simultaneously aligned to two differentially modified parental references, has the potential to improve read mapping in regions previously found to be problematic in Next Generation Sequencing. This approach is particularly important when striving to generate a complete list of candidates for novel undiagnosed diseases using genome-scale sequencing.

1484F

Workflows for variation detection and filtering using NGS data from targeted resequencing, whole genome analysis, and RNA-Seq.
M. Matvienko[1], C. Boysen[1], A. Joecker[2], A. Joecker[2], S. Monsted[2], R. Forsberg[2]. 1) CLC bio, Cambridge, MA; 2) CLC bio, Aahus, Denmark.

CLC Genomics Workbench provides a graphical and user-friendly framework for creating, distributing, installing, and running workflows. A workflow consists of a series of tools where the output of one tool becomes the input to another tool. In this way, the user can set up a workflow to go through multiple steps of analysis. For example, the read trimming output is used for mapping to reference, the mapping file is used for variant detection, and the resulting variant track is filtered against a standard variation track. Once the workflow is set up, it can be installed on the Workbench or on a Server. The workflow files are distributable and can be sent to other CLC bio users. Workflows can be run in batch mode like many of the Workbench tools, and allow for the analysis of multiple samples using the same pipeline. Any or all of the workflow parameters can be locked, so standard procedures can be used in production or regulated environments. Workflows created in CLC Genomics Workbench can also be installed on CLC Genomics Server. He re we show the example workflows created in Workbench for variation analysis on NGS data from human whole genome resequencing, targeted resequencing, and RNA-Seq data.

1485W

Choice of transcripts and software has a large effect on variant annotation.
D. McCarthy[1,2], A. Kanapin[2], M. Rivas[2], K. Gaulton[2], P. Humburg[2]. 1) Department of Statistcs, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom.

Functional annotation of variants is a crucial step in the analysis of whole-genome sequencing data. In disease studies, functional annotation results can have a strong influence on the ultimate conclusion of the study. Incorrect or incomplete annotations can cause researchers both to overlook potentially disease-relevant DNA variants and to dilute truly interesting variants in a pool of false positives. Researchers are aware of these issues in general, but until now the extent of the dependency of final results on the choice of transcripts and software used for annotation has not been appreciated to the extent it should.

We quantify the extent of differences in annotation of 10 million variants from a whole-genome sequencing study. We compare results using the RefSeq and Ensembl transcripts sets as the basis for variant annotation with the annotation software Annovar, and find only 63% agreement in annotations for putative loss-of-function variants. The rate of matching annotations remains low for loss-of-function and nonsynonymous variants combined (81%) and all exonic variants (86%). Further, we compare the results from Annovar and VEP when using Ensembl transcripts and see matching annotations for only 85% of exonic variants, with particularly large differences in the annotation of splicing variants. Using these comparisons, we characterise the types of apparent errors made by Annovar and VEP and discuss their impact on the analysis of DNA variants in whole-genome sequencing studies.

Our results show that variant annotation is not yet solved. Choice of transcript set has a large effect on the ultimate variant annotations obtained in a whole-genome sequencing study. Choice of annotation software has a smaller, but nevertheless important, effect. The annotation step in the analysis of a genome sequencing study must therefore be considered carefully, and a conscious choice made as to which transcript set and software are used for annotation.

1486T

Joint GWAS Analysis Demonstrates Increasing Similarity among Disparate Diseases as Genomic Resolution Rises from SNP to Gene to Pathway.
M. McGeachie[1,2], G. Clemmer[1], J. Lasky-Su[1,2], H.H. Chang[2,3], B. Raby[1,2], S. Weiss[1,2]. 1) Channing Div Network Med, Dept of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Children's Hospital Informatics Program, Children's Hospital Boston, Boston, MA.

Network models of genetic diseases, polygenic modeling, clinical comorbidity and gene pleiotropy studies have lead to a broader conception of the genetic causes of complex human diseases: one including increased relationships and interdependencies between disparate disorders. To assess the value of combining existing GWAS of separate cohorts and separate complex human genetic diseases for discovering additional disease-relevant biology, we compared six different GWAS from the WTCCC (bipolar disorder, cardiovascular disease, Crohn's disease, rheumatoid arthritis, and type 1 and 2 diabetes). We developed a novel methodology called Joint GWAS Analysis that is based on the enrichment of top SNPs among two GWAS for different diseases. We used Joint GWAS Analysis to examine the similarity of diseases at the SNP level, gene level, and biological pathway level. We demonstrate that this new method identifies more disease-relevant biology than could be obtained from single GWAS alone by comparing Joint GWAS Analysis and traditional single GWAS to the SNPs and genes reported in the NHGRI catalog of published GWAS for each disease. After running DAVID enriched pathway cluster analysis on known NHGRI catalog genes for each disease, Joint GWAS Analysis identified significantly more pathway clusters than single GWAS, showing gains of 27% to 100% in bipolar disorder, cardiovascular disease, and type 1 and 2 diabetes. Such gains were not realized at the SNP and gene level. These results imply that complex human genetic disorders, such as those assessed by WTCCC GWAS, show broad similarity at the functional level in enriched biological pathways, even in cases of disorders not traditionally considered to have common genetic or etiological underpinnings. We also make concrete hypotheses regarding novel pathway associations for several of the complex disorders considered, based on the results of Joint GWAS Analysis. Furthermore, the success of Joint GWAS Analysis indicates that meta-analysis of GWAS need not be limited to GWAS of the same phenotype.

1487F

Clinicall - A variant analysis tool for next generation clinical sequence data. *S. McGee, T. Kolar, M.O. Dorschner, J.D. Smith, D.A. Nickerson.* Genome Sciences, University of Washington, Seattle, WA.

Next-Generation Sequencing (NGS) has revolutionized molecular diagnostics. Test complexity and data volume have increased exponentially, creating significant challenges for clinical laboratories. Each test requires the comprehensive review of primary and variant data in a number of databases. To facilitate analysis, interpretation and reporting, we have developed an integrated tool to view sequence alignments, quality metrics and variant annotation. BAM and VCF files are easily uploaded, and variants filtered by user-defined gene sets, and settings for variant type, allele frequency and presence or absence in specific databases and functional annotations and ranking. This platform summarizes data for multiple samples simultaneously and links to existing databases to make quantitative decisions regarding validity and categorization of the variants. This is done by utilizing an extensive set of exome sequences to report a statistical measure of confidence for each newly discovered rare variant based on the quality of all calls previously seen at that site in more than 5,000 exomes. Data can be exported to one of the application's summary templates to create comprehensive reports that include the reported variants and pertinent commentary that can be accessed. This tool provides an easy and efficient way to integrate a myriad of data into a single viewable format for clinical genetics.

1488W

Streaming Algorithms for de Bruijn Graph Statistics. *P. Melsted.* Faculty of Industrial Eng., Mechanical Eng. and Computer Science, University of Iceland, Reykjavik, Iceland.

We present a streaming algorithm which can find key statistics for de Bruijn graphs without constructing the de Bruijn graph. The algorithm requires one pass over sequence data and uses a constant amount of memory, independent of the size of the dataset. The streaming algorithm can report the statistics to an arbitrary precision at the cost of increased memory usage. The statistics reported include the total number of k-mers and number of repeated k-mers. These statistics can be used to determine the per-basepair sequencing error rate without mapping to a reference. Additionally we can estimate average contig length for the de Bruijn graph, both when all k-mers are included and when unique k-mers are filtered.

1489T

Detection of Hereditary Mutations in the Mitochondrial Genome for Clinical Diagnosis. *M. Middha, M. Klebig, E. Highsmith, E. Klee.* Mayo Clinic, Rochester, MN.

Mutations in the mitochondrial genome are known to be associated with various pathological conditions and disorders caused by defects in components of the respiratory chain, including neurological disorders, maternally inherited diabetes & deafness (MIDD), among others. To improve clinical diagnosis, we are developing a Next-Generation Sequencing (NGS) test to analyze the whole mitochondrial genome (mtDNA). The definition and characterization of the analytical approach used to interpret the NGS data is a critical component of the test development process. Here we report on two critical analytical components related to the mitochondrial test development: cross-platform validation and large deletion detection. Using a dual-sequencing strategy, we sequenced ~35 mtDNA samples on both Illumina MiSeq and Ion Torrent PGM instruments. MiSeq data was analyzed using the CLC Bio Server. For the Ion Torrent data, we evaluated two methods, using the Torrent Server plug-ins or using the CLC Bio Server, to identify optimal analytics. Our goal is to use one NGS platform as an orthoginal confirmation platform for variants detected on the other. Variants can occur selectively in any percentage of mitochondrial copies (heteroplasmy), which can contribute to high variability in disease severity. Given the high coverage of sequencing in our samples, we are able to observe as low as 5% heteroplasmy. To define our false negative rate, we also sequenced 12 Hapmap samples and compared the concordance of variant calls. Mitochondrial DNA deletion syndromes such as Kearns-Sayre Syndrome & Pearson's Syndrome are caused by large deletions. To assess large deletion detection limits, two samples with known large deletions were diluted with a control mtDNA sample, to simulate heteroplasmy of 50%, 20%, 10%, 5%. We used two methods to detect large deletions, an internally developed SV detection tool, SoftSearch, and a coverage based analysis method. SoftSearch is designed to detect break points using split reads and discordant read pairs. For normalized coverage plots, we plot percent variation comparing it to the control sample to detect deletions 1kb & more. As next generation sequencing is transforming diagnostic testing, it is vital to establish a well-defined approach for analysis and interpretation of a clinical test. We provide a comprehensive clinical evaluation for the mitochondrial clinical NGS test characterizing sensitivity, specificity and confirmation of variant detection.

1490F

Accelerating Whole Genome Sequencing Bioinformatics for Clinical Application. *S. Middha, S. Baheti, S. Hart, J.P. Kocher.* Mayo Clinic, Rochester, MN.

Whole Genome Sequencing (WGS) has the potential to transform diagnostic testing in very near future. However, it is computationally expensive to align millions of short reads to the whole genome. This could be prohibitive for routine use in a clinical setting where speed of analysis can impact patient outcome. Although the time consuming steps of WGS bioinformatics are highly computational, clinical applicability can be improved by prioritizing reports based on clinical decision-making. Current clinically relevant information is largely related to protein coding regions of the genome. Based on that observation, we have designed a sequence alignment and variant calling workflow focused on protein-coding exon-only regions before processing the data over whole genome. Our **2-step workflow** significantly decreases turn-around time making it pertinent for clinical screening and diagnostics from WGS data. Our workflow first selects clinically relevant reads by aligning the entire sequencing data to the coding exons, limiting the initial reference to less than 2 percent of the whole genome. The resulting significantly reduced set of reads is then aligned to the whole genome to correct alignment artifact produced by the use of a reduced exon-only reference sequence. The data is re-aligned and re-calibrated followed by variant calling. The process thus uses a smaller reference sequence for all the sequence reads and then uses a smaller set of sequence reads to align to the whole genome. By using this 2-step approach we are able to reduce the turn-around time for whole genome alignment and variant calling from **~80 hours to ~15 hours**, a gain of more than 80 percent. The results are highly concordant to a standardly aligned WGS analysis apropos both the aligned reads as well as called variants. We repeated the analysis and generated results on 3 samples from a well characterized HapMap trio. Variant calls were also validated against SNP array data. We present those results and metrics.

1491W

Locus Reference Genomic (LRG) record: reference resource for the reporting of clinically relevant sequence variants. *J. Morales[1], J.A.L. MacArthur[1], R. Tully[2], L. Gil[1], A. Astashyn[2], E. Bruford[1], R. Dalgleish[3], E. Birney[1], P. Flicek[1], D. Maglott[2], F. Cunningham[1].* 1) European Bioinformatics Institute, European Molecular Biology Laboratory, Cambridge, UK; 2) National Center for Biotechnology Information, Bethesda, MD, USA; 3) Department of Genetics, University of Leicester, Leicester, UK.

A Locus Reference Genomic record is a manually curated resource designed specifically for the reporting of variants on clinically relevant genomic loci. An LRG provides a stable and non-versioned genomic DNA sequence for a region of the human genome, along with transcripts used as reference standards and their protein products. These sequences are selected in collaboration with the diagnostic and research communities, locus specific database curators, and mutation consortia. The genomic sequence of each LRG represents a fixed version of the RefSeqGene. Only transcripts required for reporting of disease-causing variants, and for which there is currently good biological understanding, are included. Manual curation ensures that each record fulfils the unique demands of each gene. Once an LRG is made public, this core content will not change. Other content, such as mapping information, annotation of additional transcripts, overlapping genes and legacy exon numbering systems, is periodically updated. The LRG's stable nature enables unambiguous reporting of variants in LRG genomic DNA (e.g. LRG_1:g.8463G>C), LRG coding DNA (e.g. LRG_1t1:c.572G>C) or LRG protein (e.g. LRG_1p1:p.Gly191Ala) coordinates. LRGs can be viewed in the Ensembl (www.ensembl.org) and NCBI genome browsers (http://www.ncbi.nlm.nih.gov), while tools such as Mutalyzer (https://mutalyzer.nl) facilitate remapping from other coordinates to LRG coordinates. New improvements include the addition of a stable LRG-specific exon numbering system based on the transcripts included in the fixed section. So far, 653 LRGs have been created. Of these, 392 are public and 261 are pending approval. An additional 82 are under creation. Recently made public are LRGs for genes involved in cancer predisposition, such as BRCA1, BRCA2, and TP53, inherited cardiac disease, mismatch repair, the Fanconi anemia pathway, and other diseases. It is our hope that widespread use of this resource will ensure that variant reporting is consistent over time. LRGs are compiled and maintained by the NCBI and EBI, originally as part of the GEN2PHEN (http://www.gen2phen.org) and RefSeqGene (http://www.ncbi.nlm.nih.gov/refseq/rsg) projects. All LRG records are available on the LRG website (http://www.lrg-sequence.org), which also provides instructions on requesting LRGs and the complete LRG specification.

1492T

RNA-seq alignment to individualized genomes. *S.C. Munger[1], N. Raghupathy[1], K. Choi[1], A.K. Simons[1], D.M. Gatti[1], D.A. Hinerfeld[1], K.L. Svenson[1], M.P. Keller[2], A.D. Attie[2], M.A. Hibbs[3], J.H. Graber[1], G.A. Churchill[1], E.J. Chesler[1].* 1) The Jackson Laboratory, Bar Harbor, ME; 2) University of Wisconsin, Madison, WI; 3) Trinity University, San Antonio, TX.

The emergence of high throughput sequencing of RNA has yielded a wealth of information regarding transcriptional regulation. Alignment of short read sequences to a common reference genome or transcriptome is a standard first step in the analysis of RNA sequencing data. We demonstrate that genetic variation away from the reference sequence can cause reads to be assigned to the wrong location. The source of most read alignment errors is in the duplicated structure of the transcribed genome and cannot be corrected by fine-tuning the alignment algorithm. We have developed a method, implemented as the software package Seqnature, to construct the imputed genomes of individuals (individualized genomes) of experimental model organisms including genetically unique outbred animals. Alignment to individualized diploid genomes increases read mapping accuracy and improves transcript abundance estimates. In an application to expression QTL mapping, this approach corrected erroneous linkages and unmasked hidden associations. Individualized genomes will be useful for other applications of high throughput sequencing technology that currently employ a reference sequence for alignment.

1493F

Bioinformatics and Data Management of High Throughput Sequencing Data from Prospective Cohort Study at Tohoku Medical Megabank Project. *M. Nagasaki, N. Nariai, K. Kojima, Y. Yamaguchi, I. Sato, J. Yasuda, O. Tanabe, N. Fuse, K. Kengo, R. Yamashita, J. Yamagishi, I. Danjo, M. Matsumoto, K. Igarashi, K. Nakayama, F. Katsuoka, S. Saito, I. Motoike, N. Ishida, M. Shirota, M. Yamamoto.* Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) was founded to establish an advanced medical system to foster the reconstruction from the Great East Japan Earthquake. The organization will develop a biobank that combines medical and genome information during the process of rebuilding the community medical system and supporting health and welfare in the Tohoku area. The information from the brand-new biobank will create a new medical system, and, based on the findings of its analysis, the organization aims to attract more medical practitioners from all over the country to the area, promote industry-academic partnerships, create employment in related fields, and restore the medical system in Tohoku. A blueprint for Tohoku University Tohoku Medical Megabank Organization is a ten-year project including three main activities: a biobank combining medical and genome information; an online platform for the coordination of community medical information; and training program designed for a varieties of highly specialized professionals and experts such as researchers of bioinformatics and science communicators. The biobank to be developed will be utilized to analyze the local heredity information so that it can establish an advanced medical system based on genome information with cutting-edge information and communication technology. The first goal of ToMMo is to understand the genetical detailed background of population in this area including population specific rare variants. Thus, we will apply deep coverage whole genome sequencing of thousands people who applied to this prospective genome cohort project within years. This poster presents the poster main presenter's responsible part; the data management and the analysis of massive amount of high throughput sequencing data on this project and the research position availability of this very exciting project as a graduate student or a research staff.

1494W

Transcript isoform abundance estimation method with gapped alignment of RNA-Seq data by variational Bayesian inference. *N. Nariai[1], O. Hirose[2], K. Kojima[1], M. Nagasaki[1].* 1) Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Faculty of Electrical and Computer Engineering, Institute of Science and Engineering, Kanazawa University, Kakuma, Kanazawa, Ishikawa, 920-1192, Japan.

Many human genes express multiple transcript isoforms through alternative splicing, which greatly increases diversity of protein function. Although RNA sequencing (RNA-Seq) technologies have been widely used in measuring amounts of transcribed mRNA, accurate estimation of transcript isoform abundances from RNA-Seq data is challenging, because reads often map to more than one transcript isoforms or paralogs whose sequences are very similar to each other. We propose a statistical method to estimate transcript isoform abundances from RNA-Seq data. Our method can handle gapped alignments of reads against reference sequences so that it allows insertion or deletion errors within reads. The proposed method optimizes the number of transcript isoforms by variational Bayesian inference through an iterative procedure, and its convergence is guaranteed under a stopping criterion. On simulated data sets, our method outperformed the comparable quantification methods in inferring transcript isoform abundances, and at the same time its rate of convergence was faster than that of the expectation maximization (EM) algorithm. We also applied our method to RNA-Seq data of human cell line samples, and showed that our prediction result was more consistent among technical replicates than those of other methods.

1495T

Detecting epistasis in NRXN1 pathway with permutation-based variable importance measures used in Ramdom Forests(RF). *L.A. Neira Gonzalez[1], D. Tropea[1], D. Morris[1], G. Donohoe[1], M. Gill[1], F. O'Neill[2], K. Kendler[3], D. Walsh[4], B. Riley[2], A. Corvin[1], K. Nicodemus[1], WTCCC2.* 1) TCD, Ireland; 2) Queen's University Belfast, UK; 3) VCU, USA; 4) HRB, Ireland.

The RF algorithm is designed for use with high dimensional data and can be used for predicting new observations. CNVs in NRXN1 have been shown to be strongly associated with schizophrenia (SZ), so we used RF to detect epistasis in a training (80%) and independent test (20%) sample. However, recent work has shown contradictory conclusions with permutation-based variable importance measures (VIMs) used in RF as far as which VIM is the best to use when there is correlation. Therefore, we also performed a simulation study to compare the behavior of different VIMs with different numbers of predictors correlated and different strength of correlation and to assess which was most powerful. This VIM was applied that to the NRXN1 pathway to test for epistasis. In our simulation study, we generated 500 continuous and binary datasets for each correlation 0.80, 0.40 and 0.10 and for 10, 20 and 40 variables correlated. We compared 3 types of permutation VIM (PVIM), 2 scaled PVIM, 1 AUC PVIM, 1 conditional PVIM and the original Gini VIM. The unconditional PVIM outperformed all others for correlated data and it was more powerful for binary predictors than continuous. Thus, in our NRXN1 study we used it to take the top 10 predictors ranked in the top 10 of the training data across 100 runs of RF to follow up with logistic regression interaction models in our test sample. We detected two significant interactions between the genes NPAS3 and NEBL and between NRXN1 and DSP in our independent test dataset. In NPAS3/NEBL, epistasis was detected between rs1958053 and rs4556442 (LRT p-value = 0.017 and R^2 = 1,7%) and between rs7149368 and rs4556442 (LRT p-value = 0.012 and R^2 = 1.9%). Significant epistasis in DSP/NRXN1 was found between rs2237103 and the two SNPs rs972112 and rs988179 with LRT p-values 0.044 and 0.031 with R^2 0.012 and 0.013, respectively. We found significant epistasis that influenced risk of SZ in NRXN1 pathway in an independent test dataset. NRXN1 has been associated with schizophrenia, brain structure and cognition. NPAS3 is a transcription factor involved in neurogenesis, and a translocation breakpoint in this gene was observed in a family with schizophrenia. The use of RF on the training data produced a small number of interacting SNPs to assess in our test data (45 vs 17585415), thus reducing multiple testing.

1496F

CIDRVar: A Next-Generation Sequencing Database Linking Samples, Variants, and Annotations. *J.D. Newcomer, S.M.L. Griffith, E. Pugh, D.R. Leary, J.L. Goldstein, L. Watkins Jr., K.F. Doheny.* JHU/Center For Inherited Disease Research 333 Cassell Drive Triad Technology Building, Suite 2000 Baltimore, MD 21224.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. In April 2011, CIDR's software development team implemented an ANNOVAR aggregation report generator as part of our software toolbox, CIDRSeqSuite. This report aggregation tool automates the process of running ANNOVAR over a directory of VCF files, annotating these samples against an arbitrary number of flat-file databases, and aggregating the output files into summary reports. Along with several custom databases, we use the gene-, region-, and filter-based annotations provided with ANNOVAR. This tool has been in production for over two years; in that time, we have accumulated more than 50 databases against which users may annotate based on their preferences. The accumulation of many very large annotation databases has caused scalability issues. To solve these problems, we have implemented our own solution: CIDRVar. A software tool backed by a relational database, CIDRVar serves two interdependent purposes: to track the variant genotypes found for all of our sequencing samples and to store data against which those variants can be annotated. Internally tracking the variants found for all of our sequencing samples allows us to answer quickly and easily important questions such as: /Which samples had a variant at a given genomic position? /Which variants were found at a given genomic position? /What are the annotations for a particular variant or genomic location? /How many samples had each variant? /Which samples share a given genotype? Initially, report generation required server-grade hardware. Now that we are storing in a relational database all of the annotation data that were previously stored as flat-files, we can generate a variant annotation report with a much smaller resource footprint.

1497W

Improving the accuracy of novel sequence search in de novo human genomes with NSIT. *B. Pupacdi[1], A. Javed[2], M.J. Zaki[3], M. Ruchirawat[1].* 1) Chulabhorn Research Institute, 54 Kamphaeng Phet 6, Laksi, Bangkok, Thailand 10210; 2) Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672; 3) Department of Computer Science, Rensselaer Polytechnic Institute, Troy, New York, USA 12180-3590.

The idea of building the human pan-genome was first introduced in Li et al (2010). The authors estimated that each individual's whole genome harbors about 5 Mb of novel DNA sequences that are neither known repeats nor present in the human reference sequence. These were shown to be potentially functionally important and consistent with known human migration paths. When combined, the complete human pan-genome is gauged to contain as much as 19–40 Mb of novel sequences. Subsequent independent research findings also identified 2 - 3 Mb of novel sequences per individual via different computational techniques. In this work, we showed that 1) the amount of novel sequences per individual is not as high as previously anticipated, i.e., only around 1.7–2.0 Mb per person, 2) the novel sequences among different individuals largely overlap, and 3) DNA sequence contaminations from other species in a de novo assembly can get falsely reported as novel sequences. We used our software NSIT (Novel Sequence Identification Tool) to align individually 3 de novo human genome assemblies (NA18507, YH, and NA12878) to the GRCh37 reference assembly. Repeats were masked with RepeatMasker and BLASTn was used for the final refined search steps. We found 1.9 Mb, 2.0 Mb, and 1.7 Mb of novel sequences in the aforementioned de novo assemblies, respectively. A further investigation revealed that these sequences are 60 - 80% overlapped among themselves as well as with the HuRef and CHM1 assemblies. Lastly, we found that as high as 129 kb of our NA18507 novel sequence candidates did not match with any existing human genome assemblies, or any other closely related species, but aligned with extremely high confidence and near perfect sequence identity to the Epstein-Barr virus (EBV) genome instead. More importantly, these 129 kb were also included in the novel sequences of NA18507 reported in Li et al. We speculate that these DNA sequence contaminations are present because EBV was used as the transformant when generating the cell lines. Our results therefore suggest that the size of the human pan-genome may not be as large as previously expected. NSIT is a highly efficient and accurate software for the task of detecting novel sequences in a large de novo genome assembly. For the above experiments with human genomes, NSIT required <2GB of memory and finished in 1–2 hours on a commodity desktop. To the best of our knowledge, it is the only software designed specifically for this task.

1498T

ScatterShot; a Java program for creating cluster plots from Affymetrix and Illumina genotype data. *N.W. Rayner[1,2,3], N. Robertson[1,2], M.I. McCarthy[1,2,4].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, United Kingdom; 3) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

With the increased interest in rarer variants (<1% minor allele frequency) the number of these polymorphisms appearing on the newer genotyping chips from both Affymetrix and Illumina has grown dramatically. This growth, and the uncertainty around the efficacy of calling of these variants, has greatly expanded the number of genotype cluster plots that need to be examined. To address this issue we have developed a cluster plotting program, ScatterShot, that is simple to run and is more flexible than existing programs in the data input formats required and display options available. ScatterShot is a Java program run from the command line, supporting multiple OS platforms (Windows, Unix, Mac OSX), and makes use of Java's asynchronous IO libraries to facilitate parallel processing and increase throughput; the program will scale to make use of any available processors. A wide variety of input formats are supported such as the binary .chp files from Affymetrix's genotyping console as well the Final Report from Illumina's GenomeStudio. Also supported is a generic mode, taking genotype calls from a plink format ped or binary ped file coupled with a separate XY coordinate data table. The format of the XY data table is flexible allowing one row per SNP or one row per sample. An advantage of the program over others is that when using these multiple files, the ordering of the SNPs and samples between the files is unimportant. To save space all input files, including the binary ped and chp files can be read from gzipped versions. Plots are output using Scaleable Vector Graphics (SVG) which can be viewed directly using modern HTML5 web browsers, or statically rendered into images files (JPEG, TIFF) or PDF files for printing. As an XML-based format, SVG also offers the opportunity to embed other annotations such as gender, cohort and batch information. Once embedded the SVG plot can be controlled dynamically using interactive JavaScript controls allowing for dynamic filtering of the display. This can be, for example, removing samples with no genotype call assigned, or all samples from one cohort out of many. The plots use a consistent colouring tied to the genotype call and include basic SNP QC metrics such as HWE, call rate and heterozygosity. To date we have been using ScatterShot to quickly examine and investigate the cluster plots from the new Illumina Exome and Affymetrix Axiom BioBank chips both of which contain a high number of low frequency SNPs.

1499F

Functional annotation of non-coding variants. *G.R.S. Ritchie[1,2], P. Flicek[1,2], E. Zeggini[2], UK10K Consortium.* 1) European Bioinformatics Institute, European Molecular Biology Laboratory, Hinxton, Cambridge, Cambridgeshire, United Kingdom; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, Cambridgeshire, United Kingdom.

Identifying functionally relevant variants against the background of ubiquitous genetic variation is a major challenge in human genetics. For variants that fall in genic regions our understanding of splicing and the genetic code allow us to readily interpret possible functional effects, and several computational methods exist to further prioritise non-synonymous variants that are likely to affect protein function. There are, however, currently few methods to interpret variants that fall outside of coding regions and yet these are increasingly being identified as causally relevant in human disease. Efforts such as ENCODE and the Roadmap Epigenomics Project are producing a wide range of annotation in non-coding regions but it is not yet clear how to integrate these data into variation studies. To establish if any of these annotations might be informative when interpreting variants, we use the wealth of rare variants discovered in the whole genome sequencing arm of the UK10K project to quantify constraint in annotated regions across the genome, including non-coding genic elements, transcription factor binding sites and DNase1 hypersensitive sites. We identify a number of annotations, such as binding sites for the transcription factors BRF1 and BDP1 that appear to be under comparable constraint to coding sequence (mean derived allele frequency (DAF) 0.0414 (95% CI: 0.0359-0.0469) and 0.0457 (95% CI: 0.0412-0.0502) respectively, compared to 0.0379 (95% CI: 0.0374-0.0383) for coding regions and a genome-wide average of 0.0676 (95% CI: 0.0676-0.0677), Wilcoxon rank-sum $p < 1e\text{-}8$ in both cases comparing the DAF of variants falling in the binding site with variants not falling in this annotation). We use the results from this analysis to develop a classification system that prioritises variants based on the constraint on overlapping annotations. We demonstrate that scores from this system can help identify likely functional variants using a number of validation data sets, including non-coding somatic mutations from the COSMIC database where we find we can discriminate recurrent from non-recurrent mutations with an area under the receiver operating characteristic curve of 0.75.

1500W

CLINGIE: a user-friendly Clinical Genome Interpretation Engine on the web. *R. Robison, K. Wang.* Tute Genomics Inc., Los Angeles, CA.

Given the recent development of next-generation sequencing technologies on personal genomes, functional interpretation of genetic variants has become one of the major obstacles faced by biomedical researchers, clinicians and individual consumers wishing to analyze whole genome/exome sequencing data. Many research laboratories are flooded with high-throughput sequencing data sets, but the limited ability to process these data sets significantly delays the progress to infer biological insights. More and more consumers and patients are now sequencing their own genomes, but they often lack the ability to understand the hidden health-related information buried under the variant calls provided by sequencing companies or command-line software tools. Although several software tools (such as ANNOVAR) have been developed, they suffer from a number of limitations, including the lack of a user friendly interface, the lack of functionality to organize and store many VCF files, the lack of clinically oriented evidence-based annotations, and the slow speed to handle massive amounts of data. To overcome these limitations, we developed a web application called CLINGIE (Clinical Genome Interpretation Engine, see www.tutegenomics.com), to process, store and annotate personal genomes provided by researchers, clinicians and consumers. CLINGIE accomplishes this through its clinical genome interpretation pipeline using: (1) annotation, which provides simple and intuitive interface to help users determine the functional significance of variants, (2) filtering, used to identify disease causal variants, (3) probabilistic modeling, an alternative method to score all genes in a personal genome by their likelihood of causing particular Mendelian phenotypes and (4) storage, an ultrafast data querying system for real-time retrieval and organization of user files and annotations. Users can upload genome variants files (in VCF and other common formats) via a secure web interface, and obtain comprehensive annotation information from these files. The CLINGIE application can also perform case-control and family-based analysis on a collection of personal genomes, to help identify or prioritize disease-related genes and genetic variants. We believe that CLINGIE will help researchers expedite scientific discoveries, help clinicians generate clinical genetics reports, and help consumers embrace the next-generation sequencing technology to better understand their health and disease risk.

1501T

The Cincinnati Analytical Suite for Sequencing Informatics (CASSI) is a tool to identify disease-causing variants from Next Generation Sequencing studies. *A.M. Rupert[1], L.C. Kottyan[2,3], M. Kohram[1], M. Wagner[1], M.T. Weirauch[1,2,3], K.M. Kaufman[2,3], J.B. Harley[2,3].* 1) Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; 2) Center for Autoimmune Genomics and Etiology, Division of Rheumatology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, Ohio, USA; 3) Cincinnati Veterans Affairs Med. Ctr., Cincinnati, Ohio, USA.

High coverage whole genome and exome sequencing projects create large quantities of data that require storage, versioning, filtering and annotation. We are developing the Cincinnati Analytical Suite for Sequencing Informatics (CASSI) to meet these needs. CASSI is an application that seamlessly integrates file storage, metadata storage (e.g. family structure), and downstream processing with a web-based front-end that contains a user-friendly query interface. The web interface of CASSI enables biologists and clinicians without any computer science background to launch sophisticated analytical workflows to analyze next-generation sequencing data. For example, the interface allows users to directly interface with state-of-the-art annotation and filtering packages (such as vcftools, variant tools, and ANNOVAR), which are executed on a high-performance cluster. The key technical component in CASSI is the LONI pipeline engine from UCLA, which is a graphical user interface for executing complex workflows on a cluster that can be launched directly from a web browser. Query results obtained through the CASSI web interface are made available as a data source in the LONI pipeline, and users can choose from a large number of filtering and annotation workflows to analyze variant data. In particular, CASSI can be used to assess concordance between samples and to identify de novo, rare recessive, and compound heterozygous variants. The flexibility of the pipeline facilitates the implementation of new analytical strategies directly from the interface. In a particularly informative example, we uploaded the exomic sequence from a child with a severe, pediatric-onset lupus and her two healthy parents in the form of a Variant Call File (VCF) and a FAM file with pedigree information. CASSI validated the VCF file and allowed the user to pick from easily customized filters based upon the information present in the VCF file and external annotation. This analysis led to the identification of a de novo, non-synonymous variant in a DNA repair gene. Follow-up biological assays using the child's cells confirmed a defect in the radiation-induced DNA damage repair pathway and implicated the DNA-repair pathway in the etiology of lupus in this child. CASSI is a powerful analytical suite that allows users with varying degrees of programming sophistication to perform documented, reproducible studies with Next Generation Sequencing data to gain insight into the etiology of disease.

1502F

Fast and high-resolution evaluation tool of Illumina high throughput data considering spatial organization of the sequencing clusters. *Y. Sato, K. Kojima, N. Nariai, Y. Yamaguchi-Kabata, M. Nagasaki.* Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan.

We developed a tool for fast and high-resolution evaluation of Illumina sequencing data by considering spatial organization of the sequencing clusters on a flowcell. The tool was designed as a part of quality control pipeline of the Tohoku Medical Megabank Project, and analyzes the sequence data generated by the Illumina HiSeq2500 High Output mode (HCS 2.0.05) and Rapid Run mode (HCS 2.0.x), and MiSeq v2 flowcells. This tool takes fastq, Casava fastq, and SAM/BAM files as input, and outputs heatmap plots of base quality per each cycle, read density, and mapping quality. The heatmap arrangement corresponds to the shape of the flowcell used (automatically detected). Each tile within the flowcell is divided up to 100×100 resolution and shown as respective cells of the heatmap plot. The heatmap visualizes technical errors of the sequencing, e.g., air bubbles and cracks in the flowcell, providing useful information for the improvement of sequencing procedure. In addition, this tool can convert to N base from nucleotide bases obtained from low-quality spots within the tile/flowcell. Such data processing based on the errors of flowcell or sequencing procedure would highly improve the downstream analyses, such as exploration of TF binding sites or low-frequency somatic mutations that requires high-quality read and mapping information.

1503W

Exome Database. *T. Schwarzmayr[1], T. Wieland[1], A. Schmittfull[1], E. Graf[1], T. Meitinger[1, 2], T.M. Strom[1, 2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Institute of Human Genetics, Technische Universität München, Munich, Germany.

Next-generation sequencing of coding regions was successfully used to identify rare variants over the last couple of years. We developed a pipeline to analyze exome sequencing data for variations and a database to store the resulting SNVs and indels along with pedigree information. Data populated in the database can be queried interactively via a corresponding web interface. It provides a variety of standard queries to search for rare heterozygous, compound heterozygous, homozygous, de novo and somatic variants. Thus, filtering the variations by minor allele frequency and mode of inheritance allows to identify putative candidate variants for Mendelian diseases. Furthermore, all stored variants are linked with several annotations like position in transcripts, amino acid change, functional predictions, and presence in dbSNP, HGMD and the 1000 Genomes data. Quality scores, mapping scores, read depth and links to external databases are provided additionally. As controls, publically available frequency data as well as variants identified in in-house exomes of individuals with unrelated phenotypes are used. In addition to frequency based searches, one can also perform gene and disease-based searches, thus displaying the variation content of genes or focusing the search on known disease genes or HGMD mutations. Furthermore, the database comprises information of the average read depth per exon and the proportion by which each exon is covered at least 20-times. These data are useful in a diagnostic context to define the regions of a gene which have been appropriately analyzed and to identify genes carrying possible deletions of entire exons. On top of this, the database is connected with a LIMS and provides basic quality statistics, such as amount of sequence, read depth, sample contamination and discordance between specified and identified sex.

1504T

eXtasy: variant prioritization by genomic data fusion. *A. Sifrim[1,2], D. Popovic[1,2], L-C. Tranchevent[1,2], A. Ardeshirdavani[1,2], R. Sakai[1,2], P. Konings[1,2], J.R. Vermeesch[3], J. Aerts[1,2], B. de Moor[1,2], Y. Moreau[1,2].* 1) Department of Electrical Engineering, KU Leuven, Heverlee, Vlaams-Brabant, Belgium; 2) iMinds Future Health Department, Leuven, Belgium; 3) Laboratory of molecular cytogenetics and genome research , KU Leuven, Leuven, Belgium.

Massive parallel sequencing greatly facilitates the discovery of novel disease genes causing Mendelian and oligogenic disorders. However, many mutations are present in any individual genome, and identifying which ones are disease causing remains a largely open problem. We introduce a novel computational approach, called eXtasy, to prioritize nonsynonymous single nucleotide variants (nSNVs) by integrating variant impact prediction, haploinsufficiency prediction and phenotype-specific gene prioritization that allows significantly improved prediction of disease-causing variants in exome sequencing data. To train our method we use the Human Gene Mutation Database (HGMD) as our source of disease-causing variants and 3 control sets ranging from common polymorphisms to rare variation in healthy individuals. By integrating phenotype-specific gene prioritization information we are able to greatly increase the area under the receiver-operator curve (ROC AUC) by at least 30% compared to classical deleteriousness prediction methods (e.g. SIFT, Polyphen, MutationTaster).This is likely due to eXtasy's ability to discriminate between phenotype-specific and phenotype-unrelated deleterious variants. Although our performance estimates are likely overestimated due to prior information bias in a retrospective benchmark, we show that even controlling for these biases we obtain a substantial performance increase. We believe that the presented approach will greatly facilitate the analysis of exome sequencing data in human disease by efficiently prioritizing nSNVs in the light of the phenotype in question. eXtasy is publicly available at http://homes.esat.kuleuven.be/~bioiuser/eXtasy/.

1505F

PHEVOR: Integration of VAAST with Phenomizer and the Gene Ontology for accurate disease-gene identification using only a single affected exome. *M. Singleton, L. Jorde, M. Yandell.* Human Genetics, University of Utah, Salt Lake City, UT.

Accurate identification of disease-causing genes using only a single exome is major analytic challenge. For recessive rare diseases, when only a single affected individual's exome is available, VAAST and Annovar typically rank the disease-causing gene among the top 10 candidates only 15% and 2% of the time, respectively. To improve disease-gene identification, we have developed PHEVOR, an algorithm that integrates three popular tools: (1) Phenomizer, which provides powerful means to generate candidate disease-gene lists based upon phenotype data, (2) the Gene Ontology (GO), which provides function, location and biological process descriptions for more than 18,000 human genes, and (3) VAAST a probabilistic disease-gene finder that uses variant-frequency information together with amino acid substitution (AAS) frequencies to identify disease-causing variants in personal genome sequences. We benchmarked PHEVOR using 50 known recessive disease-causing variants spiked into otherwise healthy exomes. On this dataset, VAAST identifies the disease-gene as the top candidate only 2% of the time and places it in the top 10 candidates 15% of the time. Phenomizer does about as well, ranking the disease-gene first 3% of the time and in the top 10 candidates 14% of the time~but with an important caveat: Phenomizer's candidate list is restricted to 2,528 known human disease genes. If the disease-causing alleles do not reside in one of these genes, they will be missed no matter how informative the phenotype and genomics data. PHEVOR overcomes this limitation by propagating phenotype information across the Gene Ontology, extending Phenomizer's candidate list to more than 18,000 human genes. PHEVOR also provides probabilistic approaches to incorporate personal genome data using VAAST's proven statistical methodology. The improvement is dramatic. Using the same benchmark dataset described above, (note that 54/100 genes are not included in Phenomizer's gene list) and only a single affected exome, PHEVOR is able to identify 56% of disease-causing genes as the top candidate genome-wide, and 79% of the time the disease-gene is within the top 10 candidates. PHEVOR thus provides an effective new means for single exome-based diagnosis.

1506W

Multiagent-based SNP Annotation for Large-Scale Genetic Diversity Analyses: An Application to the Brazilian EPIGEN Initiative. *G.B. Soares-Souza[1], E. Tarazona-Santos[1], A. Pereira[2], M.L. Barreto[3], B.L. Horta[4], M.F. Lima-Costa[5], A. Horimoto[2], N. Esteban[2], F. Kehdy[1], W.C.S. Magalhaes[1], M.R. Rodrigues[1], The Brazilian EPIGEN Consortium.* 1) General Biology Department, Federal University of Minas Gerais, Belo Horizonte, Brazil; 2) Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of São Paulo, São Paulo, Brazil; 3) Instituto de Saúde Coletiva, Federal University of Bahia, Salvador, Brazil; 4) Universidade Federal de Pelotas, Pelotas, Brasil; 5) Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil.

Despite the great amount of genetic diversity data available, the study of variants with biomedical interest requires their association with other types of biological data, such as pathways, pharmacogenetics, GWA, and so on. Since these heterogeneous data are fragmented in different databases, it is necessary to develop tools to integrate different sources of biological data in order to move forward with genetic studies of diseases or populations. Although there are annotation tools that solve this problem, the management of large scale data is still a challenge in terms of execution time. Computer technologies that incorporate parallel execution of the code are a promising solution for processing large scale data, since they are able to reduce execution time by dividing the execution load into independent processing units. One such technology is the agent-based technology. Here, agents are autonomous programs that perform a particular task and can communicate with other agents in a Multiagent system to delegate tasks, share results or get complementary information. Agents in a Multiagent system execute concurrently, as independent threads. Given the heterogeneity of biological data and their distribution over distinct sources, we have chosen this technology to development a SNP annotation system for large scale data. In our system, Database Agents encapsulate databases of interest, such as dbSNP, pharmGKB, OMIM, UCSC and GO. The annotation process works as follows: (i) a Coordinator Agent receives a list of SNPs to be annotated and passes them on to each Database Agent; (ii) the latter Agents then retrieve, concurrently, relevant information from their encapsulated database to annotate the list of SNPs and return it to the Coordinator; (iii) after all annotated SNPs are received, the Coordinator agent merges the annotations and passes them on to the Interface Agent to generate an annotation report. Our Multiagent annotation system is being used to thoroughly annotate data from the Brazilian EPIGEN initiative, which involves the genotyping of 2.3 M SNPs in 6600 individuals, 4.3 M SNPs in 270 individuals and 30 complete genomes. So far, tests showed that the distributed architecture of the Multiagent system significantly reduces the annotation execution time. As a follow up study, we will investigate the population genetics of different SNP classes to understand the implications of its allelic spectrum on GWAS studies, and new GxG and GxE approaches.

1507T

Comprehensive network and pathway analysis of RNA sequencing of triple negative breast cancers. *J.P. Solzak[1], R. Atale[1], B.A. Hancock[1], J.N. Billaud[2], M. Radovich[1].* 1) Surgery, Indiana University School of Medicine, Indianapolis, IN; 2) Ingenuity Systems, Redwood City, CA.

Introduction: Triple-negative breast cancers (TNBCs) account for 15% of all breast cancers cases and are defined by an absence of actionable therapeutic targets (ER-,PR-,HER2-). Using RNA-seq data, we compared TNBCs to microdissected normal breast epithelium from healthy volunteers and to normal tissue adjacent to tumor followed by comprehensive network and pathway analysis. Methods: RNA-seq data from 94 TNBCs (from Indiana University and TCGA), 20 microdissected normal breast tissues (Komen Tissue Bank), and 10 adjacent normal tissues (TCGA), were merged and imported into Partek Genomics Suite. The merged transcript RPKMs were transformed, batch effect corrected, and analyzed for differential expression. Statistically significant genes were imported into Ingenuity Pathway Analysis (IPA) Spring 2013. Results: IPA analysis of differentially expressed genes of TNBCs compared to microdissected normal breast tissues identified key pathways in DNA damage, cell cycle, and immune signaling. We then employed Upstream Regulator and Causal Network Analysis which predict the activation or inhibition state of regulators based on downstream differentially expressed genes and known directionality of expression. From this analysis, two statistically significant upstream regulators with known roles in metastases were identified: IRF7 (interferon regulatory factor), and FOXM1 (cell-cycle transcription factor). In particular, FOXM1 is predicted to drive the activation of a multitude of genes involved in cancer cell proliferation. In addition, we identified two causal networks centered on inhibited SPDEF (Ets transcription factor), and on activated HOXB4 (developmental transcription factor). Of note, HOXB4 has not previously been described in TNBC and may play a role in the stem cell phenotype of TNBC. In contrast, when TNBCs were compared to adjacent normal tissue, some of the same pathways and regulators were identified, but other pathways involved in fibrosis, atherosclerosis, metabolism, and edema were also observed, likely secondary to the stromal nature of adjacent normal tissue. Conclusion: By using RNA-seq data from TNBCs and microdissected normal epithelium coupled with comprehensive network and pathway analysis we demonstrate the utility of these methods to uncover novel biological insights into TNBC biology. In particular, we observe that IPA's Causal Network Analysis has unique capabilities to identify master regulators useful for drug target discovery.

1508F

GrabBlur - a framework to facilitate the secure exchange of whole-exome and -genome sequencing data sets. *B. Stade[1], D. Seelow[2], A. Franke[1].* 1) Institute of Clinical Molecular Biology, University Hospital Schleswig-Holstein, Kiel, Germany; 2) NeuroCure Clinical Research Centre, Charité - Universitätsmedizin, Berlin, Germany.

With 'GrabBlur' we developed a tool to collect, aggregate and share SNV (single nucleotide variant) data of hundreds of samples with a special trait/ phenotype in a public database, while keeping each individual sample unidentifiable. First, 'GrabBlur' helps a submitter to combine SNV data from samples with phenotype and sequencing information (VCF files) using a local web interface or command line operations. Examples are HPO terms (Human Phenotype Ontology), sample traits, information about the applied sequencing technology and the submitting institute / laboratory. Most of the information is optional - the submitter can decide for himself what he wants to and can share. In the second step 'GrabBlur' merges and aggregates the data. For example, if it turns out that a SNV is rare - every submitter sets the proper threshold for his own data - all sample information is replaced with contact information of the submitter. In further steps the data will be highly aggregated so it can be shared while adhering to data privacy. It will not be possible to recover the exome/ genome of a sample or to identify a single individual in the dataset. After checking that the individual genetic information is sufficiently 'blurred' the data can be uploaded into a public database, where the sample sets will be aggregated again, thus adding another 'blurring' step. If another researcher finds an interesting SNV in the dataset he should now be able to get in contact with the original submitter to exchange information on the individuals, e.g. if his patient has a similar phenotype as the individual in the database. A webinterface exists for querying the database and to derive also gene-wise SNV information. In summary, our tool will facilitate large-scale data exchange between researchers as the genetic identity of individuals is sufficiently protected. The database will be of help for human geneticists who seek to distinguish between incidental or population-specific findings and truly disease-causing mutations. Moreover, the data in the database will aid to drive case-control association studies employing large-scale SNV data. Finally, exome/genome data can be shared as promiscuous as done for GWAS studies, which will help to drive well-powered analyses.

1509W

Detection of Informative Variants in Low Coverage Next Generation Sequencing Data as a Quality Control Tool to Ensure Sample Integrity. *L. Timms, P. Ruzanov, J.K. Miller, N. Buchner, D. Pasternack, J.D. McPherson.* Ontario Institute for Cancer Research, Toronto, Ontario, Canada.

Next generation sequencing (NGS) allows for the rapid generation of exome or whole genome coverage across multiple samples. Tracking of matched samples (eg tumor/normal pair) and multiple lanes of samples that are combined to generate sufficient coverage must be monitored early in the process to avoid costly sequencing of misidentified samples and lanes. Pooling of sequencing data for even a single unmatched lane within a sample severely affects downstream variant detection. Short tandem repeat (STR) analysis, such as AmpFISTR Identifiler ®, is a standard, PCR-based human genotyping method used in the field of forensics. This genotyping method has also been effectively applied as an early quality control step for NGS to prevent processing of misidentified samples; however, the STRs are not efficiently captured with short read NGS technologies and difficult to apply to fragmented DNA samples such as those derived from formalin fixed paraffin embedded material. Single nucleotide genotyping methods can also be used for identification requiring as few as 50 informative variants for identification. We have implemented a per lane level method for efficient library identification as a quality control measure prior to variant detection. We have identified a list of 150 polymorphic (MAF 0.49) SNVs that can be detected in low-coverage sequencing data. Homozygous variants can be called at a minimum of 8X, and heterozygous variants must have a minimum of two reads per allele. Pairwise comparisons are then made between sequencing lanes of the same sample or matched sets utilizing an overlapping subset of the targeted loci with sufficient coverage in both samples. Such analyses can be done as part of an initial quality control check of the data prior to starting complex analysis that requires more time and computer resources.

1510T

Consensus Genotyper for Exome Sequencing: Improving the Quality of Exome Variant Genotypes. *V. Trubetskoy[1], R. Madduri[2], A. Rodriguez[2], J. Scharf[3], P. Dave[2], I. Foster[2], L. Davis[1], N. Cox[1].* 1) Medicne/Genetic Medicine, UNIVERSITY OF CHICAGO, CHICAGO, IL; 2) Computation Institute, University of Chicago, Chicago, IL; 3) Department of Neurology, Massachusetts General Hospital, Boston, MA.

With the rise of next generation sequencing methods, there are now a variety of tools available to researchers for detection and genotyping of sequence variants. However, the concordance among variant sets between these disparate approaches has been shown to be poor. Recently, researchers in the machine learning community have shown that combining the output of multiple models can dramatically improve performance of a classifier. Collectively, these techniques are referred to as ensemble methods. Here we describe a novel variant calling approach based on an ensemble of variant calling algorithms which we call Consensus Genotyper for Exome Sequencing (CGES). Our method employs a two-stage voting scheme among a set of three algorithm implementations, GATK2.0, Freebayes, and Atlas2.0, which were used to identify variant sites and determine genotypes in the study. We apply CGES to a dataset consisting of 123 samples sequenced at the Center for Inherited Disease Research (CIDR) using the Agilent SureSelect3 Exome Capture and Illumina sequencing technology. Samples were drawn from extended pedigrees, allowing us to compare individual and family based quality metrics across all algorithms. The CGES approach is shown to outperform its constituent parts in many key quality metrics without a significant loss in the number of variant sites called. In particular we are able to achieve a threefold reduction in Mendelian inconsistencies between the best performing variant caller and our consensus approach (CGES = 240.14/trio and Atlas2.0 = 699.59/trio). For callers with comparable QUAL scores, our CGES set of variants has an average QUAL score 11% (GATK) and 70% (Freebayes) higher than the unfiltered output set of each respective variant caller. Additionally, the consensus set outperforms all individual callers in the study with regard to expected exome-wide transition-transversion ratio(CGES = 3.07 and Atlas2.0 = 2.98). For the purpose of accessible, efficient, and reproducible analysis, we provide implementation of CGES as a stand alone command line tool, as well as a set of parallel Galaxy tools and workflows for accessible and efficient use by the research community (see **Implementing a High Performance, Reusable Consensus Calling Pipeline for Next Generation Sequencing using Globus Genomics**, Madduri et al., ASHG 2013).

1511F

DRAW+SneakPeek: Analysis Workflow and Quality Metric Management for DNA-Seq Experiments. *O. Valladares[1,2], C.-F. Lin[1,2], D.M. Childress[1,2], E. Klevak[3], E.T. Geller[1], Y.-C. Hwang[2,4], E.A. Tsai[4,5], G.D. Schellenberg[1], L.-S. Wang[1,2].* 1) Department of Pathology and Laboratory Medicine, University of Pennsylvania. Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Department of Physics, University of Washington, Seattle, WA; 4) Genomics and Computational Biology Graduate Group, University of Pennsylvania. Philadelphia, PA; 5) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA.

We report our new DRAW+SneakPeek software for DNA-seq analysis. DRAW (DNA Resequencing Analysis Workflow) is based on commonly used open source programs for analyzing sequencing experiment data. DRAW implements the Best Practice Variant Detection from Genomic Analysis Toolkit (GATK) and automates the workflow of processing raw sequence reads including quality control, read alignment, and variant calling on High-Performance Computing (HPC) facilities such as Amazon Elastic Compute Cloud (EC2). SneakPeek provides an effective interface for reviewing dozens of quality metrics reported by DRAW, so users can assess the quality of data and diagnose problems in their sequencing procedures. We evaluated DRAW on Amazon EC2 using a single flowcell WES dataset of 350.2G nucleotides from 34 multiplexed samples using Nimblegen SeqCap EZ Human Exome Library, 100bp pair-end reads on a HiSeq 2000 sequencer. DRAW processed the whole dataset in 1,943.2 core-hours, or 17.4 hours/core on 14 Quadruple Extra Large Instances with 112 cores, generating 1.1TB of data. The total cost was $528 including storage, computing, and data download. Both DRAW and SneakPeek are freely available under the MIT license, and are available as Amazon Machine Images to be used directly on Amazon Cloud with minimal installation. Instructions and codes can be obtained from the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) at https://www.niagads.org/content/drawsneakpeek.

1512W

Performance of two imputation methods on large scale data: experiences in the eMERGE network. *S. Verma[1], G. Armstrong[1], M. Ritchie[1], D. Crawford[2], Y. Bradford[2], M. Andrade[3], I. Kullo[3], G. Tromp[4], H. Kuivaniemi[4], L. . Armstrong[5], G. Hayes[5], B. Keating[6], D. Crosslin[7], G. Jarvik[7], B. Namjou[8], E. Bookman[9], R. Li[9], eMERGE Network.* 1) Center for Systems Genomics, The Pennsylvania State University, University Park, PA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Mayo Clinic, Rochester, MN; 4) Geisinger Health System, Danville, PA; 5) Northwestern University, Chicago, IL; 6) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 7) University of Washington, Seattle WA; 8) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 9) Division of Genomic Medicine, National Human Genome Research Institute.

The eMERGE network, an NHGRI funded initiative comprises nine sites each with DNA biobanks linked to electronic health records (EHRs). Approximately 39,206 unique DNA samples have been genotyped using either Affymetrix or Illumina genome-wide SNP arrays. Led by the Coordinating Center and the eMERGE genomics workgroup, we have developed an imputation pipeline for merging genomic data across the different SNP arrays used by the eMERGE sites, to maximize sample size and the power to detect associations. We performed imputation using the 1000 Genomes Cosmopolitan reference panel - which includes 1092 individuals and over 36 million SNPs. We compared accuracy of imputation results from two software packages - Beagle and Impute2 (phasing performed with ShapeIt2). For the comparison we used the following metrics: accuracy of imputation, allelic R2 (estimated correlation between the imputed and true genotypes for all imputed SNPs), and relationship between allelic R2 and minor allele frequency across all imputed SNPs. Since imputation is computationally intensive, we compared computation time and computing resources required by the two software packages. We have also outlined the major challenges and lessons learned due to the complexity of using these two approaches and different platforms. Finally, we will present lessons learned as we conducted our quality control process, as many of the pipelines are unique and cleaning of imputed data is essential for further analysis. This imputed dataset will serve as a valuable resource for variant discovery, leveraging the wide range of medically relevant phenotypes that can be mined from the EHR.

1513T

FILTUS - a versatile and user-friendly program for filtration and statistical evaluation of variants in high-throughput sequencing projects. *M. Vigeland.* Medical Genetics, Oslo University Hospital, Oslo, Norway.

We present FILTUS, a program for working with variant files resulting from sequencing projects, e.g. exome sequencing. Various options for filtering and summarizing statistics are available, as well as exploratory tools like pairwise sharing among the samples. Several statistical methods for evaluation of variants are implemented, aiding the identification of causal variants in Mendelian disease projects. Both case-control and family based sequencing designs are supported, as well as dominant/simple recessive/compound recessive disease models. Homozygous regions can be identified (via PLINK), and the program offers conversion of variant files to MERLIN format, facilitating linkage analysis. FILTUS accepts variant files in any format, as long as the user identifies essential columns like chromosome, position, gene and genotype. Files in VCF format are automatically recognized and given special treatment. The implementation allows several hundred complete exomes to be analyzed simultaneously on a standard laptop. FILTUS has an easy-to-use GUI accompanied with a user manual, and does not require particular bioinformatic skills. It is written in Python and runs on Windows, Mac and Linux.

1514F

DeMix: Deconvolution for Mixed Cancer Transcriptomes Using Raw Measured Data. *W. Wang[1], J. Ahn[1], Y. Yuan[2].* 1) Bioinformatics & Comp Biology, Univ Texas MD Anderson Cancer Ctr, Houston, TX; 2) Biostatistics, Univ Texas MD Anderson Cancer Ctr, Houston, TX.

Tissue samples of both tumor and stromal cells cause underdetection of gene expression signatures associated with cancer prognosis or response to treatment. In silico dissection of mixed cell samples is essential for analyzing expression data generated in cancer studies. Currently, a systematic approach is lacking to address three challenges in computational deconvolution: 1) violation of linear addition of expression levels from multiple tissues when log transformed microarray data are used; 2) estimation of both tumor proportion and tumor-specific expression, when neither is known a priori; and 3) estimation of expression profiles for individual patients. We have developed a statistical method for deconvolving mixed cancer transcriptomes, DeMix, to address the above issues in array-based expression data. We demonstrate the performance of our model in synthetic and publicly available real datasets. Our method can be applied to ongoing biomarker-based clinical studies, as well as to the vast expression datasets previously generated from mixed tumor samples.

1515W

üRRBS-Predictor, a webtool to guide enhanced reduced-representation bisulfite (RRBS) analysis of genomic DNA methylation patterns. *T.R. Ward[1, 2], X. Zhu[1, 2], A.E. Urban[1, 2].* 1) Genetics Dept, Stanford, Stanford, CA; 2) Dept of Psychiatry, Stanford, Stanford, CA.

üRRBS-Predictor is a bioinformatics tool that predicts which parts of a given genome sequence will be included in an RRBS experiment if a given size-fraction of DNA fragments is chosen during the RRBS sample preparation step. Genome-sequencing after bisulfite conversion is a powerful approach to understanding genomic DNA methylation. But whole genome bisulfite sequencing is costly and inefficient at today's sequencing and computing speeds, due to GC poor regions in many genomes that will be included in the analysis without contributing much if anything at all in terms of insight regarding the question at hand. RRBS uses restriction cutting at recognition sites containing CpG dinucleotides (typically with the restriction endonuclease MspI) followed by size selection for genomic DNA fragments in order to enrich for the CG-rich parts of a genome before next-generation sequencing. However, the size fraction taken during the selection step of standard RRBS protocols seemed somewhat arbitrarily chosen and allows for only a very small portion (often just around 1%) of a typical genome to be included in the analysis, with little or no ability to assure that given regions of interest such as certain genes or regulatory regions are covered. üRRBS-Predictor is centered around an algorithm that will analyze the whole genome and report back where in the genome MspI will cut (fragment start and stop positions), the fragment lengths and the number of CpGs found in each fragment. The user can then explore which portion and percentage of a genome would be covered by RRBS sequence data if a given size fraction of DNA fragments is selected during RRBS sample preparation (in 50 bp increments between 50 and 800 bp, based on experimental limitations). At first this can be done for the human genome and the genomes of the most common model organisms such as mouse, rat, fruit fly, zebrafish, C. Elegans (later-on additional genomes can be included in the webtool). The script produces a BED file that can be uploaded to the UCSC Genome Browser that will display the fragments in their locations and color-coded based on the size of the fragment. Alternatively, the user will be able to select a genomic region of interest and will be given all the fragments that cover that region along with their sizes. [üRRBS-Predictor is just a working title, ü for über, could be enhanced-RRBS-predictor/eRRBS-predictor, variable-RRBS/vRRBS - or something entirely different].

1516T

Analysis of archived residual newborn screening bloodspots after whole genome amplification, using Genformatic's genomic medicine tool suite. *D. Weaver[1], B. Cantarel[1], J. Reese[1], R. Finnell[2, 3].* 1) Genformatic, Austin, TX; 2) University of Texas, Austin; 3) Dell Pediatric Research Institute.

A wealth of genomic health information is archived in residual newborn blood samples originally collected and used for routine newborn screening for common genetic diseases. Where legally and ethically permissible, exome or whole genome analysis of archived samples could potentially reveal valuable epidemiological, population genetic and personal genomic information. However, the quantity of residual sample material in many biobanks is relatively small and may present an impediment to whole genome or exome analysis. Nonetheless, the information that could be gleaned from large-scale analysis of archived residual newborn blood is tremendous, provided current methodological impediments can be overcome. In our study, we compared the recovery of variant calls from NGS of amplified DNA to the variant calls generated in standard exome analysis of non-amplified material from the same individuals, using Genformatic's cloud-based genome analysis suite. While some data fall-out from whole genome amplified DNA is evident, the extent of concordance between variant call sets demonstrates that valuable genomic medicine information may be obtained from sequencing residual newborn screening bloodspots after whole genome amplification. The information which could be generated in a broad survey of currently archived newborn samples is vast and complex, requiring extensive computing capabilities, including large amounts of storage space and memory. We offer a cloud-based genome analysis and comparison suite of tools enabling users to (i) detect variation, (ii) annotate and filter SNPs, (iii) share encrypted genome data with collaborators, and (iv) compare multiple genomes to identify shared ancestry or common variation. Our analysis tools allow users to analyze their data in a secure and scalable fashion, without building a complicated server infrastructure. Applying our methods, we have determined that approximately 80% of the SNPs are concordant between the amplified and the PCR-free sequence data. Additionally, we evaluated SNPs predicted from the same individual using alternative capture kits, and discover a surprisingly high level of discordance in SNP calls - only 15-70% SNP concordance, depending on which SNP detection method is employed.

1517F

Visualizing Multiple Types of Genomic Information Across Chromosomes With PhenoGram. *D. Wolfe, S.M. Dudek, M.D. Ritchie, S.A. Pendergrass.* Center for Systems Genomics, The Pennsylvania State University, University Park , PA.

With the abundance of information and analysis results being collected for genetic loci, user-friendly and flexible data visualization approaches can inform and improve the analysis and dissemination of these data. An ideogram is a graphic representation of chromosomes, and these plots have been used with the addition of overlaid points, lines, and/or shapes, to provide summary information of various kinds coupled with genomic location information. For instance, the results of multiple published genome-wide associations have been plotted on ideograms (www.genome.gov/gwastudies), where lines identify chromosomal location and colored circles indicate associations with different phenotypes/traits. We have developed a flexible software tool PhenoGram, which exists as a web-based tool and also a command-line program, providing a way to visualize data in multiple ways via ideograms. Initially conceived as a method to highlight SNP-phenotype association results across the genome through the use of color-coded circles linked by lines to genomic locations like the NHGRI GWAS catalog plots, the software has been expanded with many unique features. With PhenoGram users can plot lines at specific base-pair locations, or base-pair to base-pair regions of chromosomes, with the use of color, with or without other annotation. This feature has been used in several ways to highlight locations or regions of interest. For example, we have used PhenoGram to produce plots showing the genomic coverage of SNPs from a genotyping array, plots highlighting the chromosomal coverage of imputed SNPs, as well as plots showing the location of sequenced loci. We have also successfully used this software to plot regions covered by two different copy-number variant detection methods, visually contrasting overlapping regions between the two approaches. PhenoGram allows users to annotate chromosomal locations and/or regions with shapes in different colors. This feature can be used to indicate different traits associated with specific loci, and users can choose different shapes to highlight ancestry or another study attributes related to specific data points. Further, users can annotate an ideogram with gene identifiers or other text, as well as create plots showing expanded detail for specific chromosomal locations. PhenoGram is a versatile software, further fostering the exploration and sharing of genomic information. For full details see: http://visualization.ritchielab.psu.edu.

1518W

Massively parallel sequencing as a tool for HLA typing. *W. Yang[1,2], Y. Huang[1,2], J. Yang[1,2], P.C. Sham[2], Y.L. Lau[1].* 1) Paediatrics & Adolescent Med, Univ Hong Kong, Hong Kong, 10000, Hong Kong; 2) Centre for Genomic Sciences, the University of Hong Kong.

It has been shown that massively parallel sequencing(MPS) for HLA typing has the potential to become a more accurate, high throughput, and highly flexible HLA typing platform for both clinical laboratories and research settings. We introduce the two areas we have made progresses in using MPS for HLA typing, including improvement in using genome-wide sequencing (GWS) data to make HLA calls and pooling of a large number of samples for HLA typing using MPS technology. HLA typing from GWS data using available tools has many issues, often because they don't fully account for the polymorphic nature of this group of genes and enormous allelic differences. Thus we developed a multiple references-based mapping approach for defining HLA types from GWS and identified large number of previously unmapped sequencing reads, which allowed a much more complete HLA calling from GWS data. A toolkit was developed to facilitate using this method for HLA typing from GWS. Making use of the available whole exome sequencing data we have and using this novel HLA mapping and calling method, we were able to gather detailed information on the sequences of the polymorphic exons of the major HLA genes and to design PCR primers for amplifying these exons. The primers are particularly suited for East Asian populations. In research laboratories, we often need to type certain HLA genes on a large number of samples, which is quite different from the practices in clinical laboratories. This is especially important for genetic studies of autoimmune diseases and population screening for certain HLA gene alleles to avoid adverse drug responses. Making use the method and primers we have developed, together with an innovative barcoding system to allow pooling of up to a thousand PCR amplicons together for next generation MPS sequencing, we were able to increase throughput of HLA typing and reduce sequencing cost significantly, while providing flexibility and automation needed for research projects. These innovations will enhance the development of using MPS technology for HLA typing for both clinical laboratories and research settings, and the results are particularly suited for East Asian populations.

1519T

PHV: A highly accurate SNP and INDEL variant detector based on the Profile Hidden Markov Model. *M. Zhao, W. Lee, G.T. Marth.* Biology, Boston College, Chestnut Hill, MA.

Accurately calling single nucleotide polymorphisms (SNPs), insertions and deletions (INDELs) is crucial to medical and population genetics. A variety of algorithms have been developed to detect SNPs and have been successfully applied to large-scale re-sequencing projects, but detecting INDELs still presents significant problems. Moreover, most of the existing algorithms are designed for high-quality short-read data with a low INDEL sequencing error rate whereas upcoming technologies such as Ion Torrent sequencing technique produces long reads at the cost of an increased sequencing error rate. This leads to more challenging variation detection, and is beyond the scope of present variant detection algorithms. We have developed PHV based on the Profile Hidden Markov Model (PHMM) for accurately calling SNPs and INDELs from data with varying quality. This method jointly models sequencing and alignment uncertainties, which are overlooked or not properly modeled by most existing variant calling algorithms. The high time complexity of the PHMM algorithm is reduced by the banded approach, so at the same time of achieving high accuracy, the running time of PHV is comparable with other variation detection tools. While this tool is designed for long read technologies, it also demonstrates an improvement in variation detection when deployed on currently available short read technology data. We used both simulated Illumina data and a single sample (NA12878 from the 1000 Genomes Project) with deep sequenced whole genome data (Illumina and LS454) to test the algorithm. Comparing with SNPs called by the Genome Analysis Toolkit (GATK), Samtools and Freebayes, PHV's results have the best receiver operating characteristic (ROC) curve on the simulated data. Moreover, comparing with the other tools, PHV shows higher sensitivity and specificity for INDEL detection and higher specificity for SNP detection for the real data, especially in the genomic regions with poorly aligned reads. PHV will be further tested with real PacBio and Ion Torrent data.

1520F

Resolving complex structural genomic rearrangements using a randomized approach. *X. Zhao[1], S.B. Emery[2], J.M. Kidd[1,2], R.E. Mills[1,2].* 1) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Genetics, University of Michigan, Ann Arbor, MI.

Structural variants (SVs), defined as the deletion, duplication, insertion, inversion or translocation of genomic regions, are both a major source of genetic diversity in human populations and are also directly responsible for the pathogenesis of numerous diseases. Many studies have been conducted in the past decade to discover and analyze SVs, however these have predominantly focused on unbalanced (copy number variant) events involving only one or two breakpoints. In contrast, more complex rearrangements resulting from multi-step or overlapping events involving three or more breakpoints have received considerably less attention or have been incorrectly interpreted. We have developed an algorithm, metroSV, which accurately analyzes and resolves complex rearrangements using whole genome, paired-end sequencing data. This method first identifies regions of the genome suspected to involve a complex event and then delineates putative breakpoints using aberrant sequence alignments. The resulting segments are then iteratively rearranged in a randomized fashion and scored against expected models of sequence characteristics using a Markov chain Monte Carlo approach to infer the underlying architecture of these variants. We have applied our algorithm to well-characterized genomes including the high coverage Illumina Platinum CEPH pedigree that has been deep sequenced and compared our results to previously reported complex events in these samples. We are in the process of broadening the scale of our algorithm to assess the entire genome concurrently and are also expanding our analysis to other large-scale, publicly available datasets as well as pathogenic samples associated with germline or somatic chromothripsis-like events. We believe metroSV represents a significant advancement towards resolving these complex chromosomal structural rearrangements and furthering our understanding of their mechanistic origins and functional impact.

1521W

Systematic integration of functional and computational genomics suggests that the indel rs79240969 in the DNM3 gene influences both bone- and obesity-related traits. *M. Claussnitzer[1], X. Chen[1], D. Karasik[1], LA. Cupples[2], DP. Kiel[1], YH. Hsu[1].* 1) Inst Aging Res, Hebrew Senior Life, Harvard Medical School, Boston, MA; 2) Boston Univ Sch Pub Hlth, Boston, MA.

Recent studies suggest shared etiologies of bone and adiposity phenotypes. Previously, we conducted a multivariate genome-wide association study (GWAS) of bone mineral density (BMD) at different skeletal sites and metabolic syndrome risk factors, and identified 31 bivariate genome-wide significant loci ($p<5\times10\text{-}8$, p(bivariate)<p(univariate)/10). However, signals emerging from GWAS are merely markers for large genomic regions in linkage disequilibrium (LD), harboring the disease-causing variant. Identifying the shared causal variants is central to elucidate the molecular mechanisms underlying the genetic correlation between bone and fat. We applied an integrative bioinformatics approach leveraging tissue-specific functional and computational genomics and sequence data to narrow-down the potential causal variants. We identified all reported sequence variants (from the 1000 Genomes Project) within the identified bivariate GWAS loci (physical boundaries LD r2>0.7). We merged data on sequence variants with bone- and adipose-specific genome-wide epigenomic profiling data, reported from the ENCODE project, that allow for chromatin state-dependent analyses of regulatory variation. Within predicted regulatory regions, we discovered cis-regulatory modules (CRMs) by analyzing cross-species conserved patterns of transcription factor binding sites (TFBS) across 16 vertebrate species. Using our approach, we identified potential causal variants within the bivariate association loci that may be responsible for the association with BMD and obesity traits. One example is the bivariate signal at DNM3, 1q24.3 (GWAS SNP rs10489290, bivariate association test p=8.4 x 10-11 for femur neck BMD with waist-to-hip ratio). We pinpointed an intronic insertion/deletion variant rs79240969 (-/TCA, MAF=0.296, LD r2=1.0), specifically mapping within predicted osteoblast and adipocyte gene regulatory regions. The cell-type specific TFBS pattern analysis revealed rs79240969 localizing within a cross-species conserved CRM relevant to bone and fat. The rs79240969 insertion allele creates a perfect binding site for the zinc-finger protein Zfp521 which has been previously shown to control bone mass. Our bioinformatics analysis may represent a useful step toward pinpointing causal variants from potentially pleiotropic loci for direct laboratory validation and ultimately for improving therapeutic strategies addressing shared etiological mechanisms of BMD and obesity-related phenotypes.

1522T

The Developmental Brain Disorders Database (DBDB): A curated neurogenetics knowledge base with clinical and research applications. *G. Mirzaa[1], AJ. Barkovich[2], WB. Dobyns[1], KJ. Millen[1], AR. Paciorkowski[3].* 1) Deparmtent of Human Genetics, Seattle Children's Hospital, Seattle, WA; 2) Department of Pediatrics, University of California San Francisco, San Francisco, California; 3) Depts of Neurology, Pediatrics & Biomedical Genetics, Center for Neural Development & Disease, University of Rochester Medical Center, Rochester, NY.

The number of genes associated with neurodevelopmental disorders has risen dramatically over the past decade, with the discovery of many genes associated with intellectual disability, autism, epilepsy, and others. The maturation of copy number studies and the use of whole exome sequencing (WXS) have further increased the number of genes associated with neurodevelopmental phenotypes. For geneticists, this rapid expansion of knowledge increases the difficulty of arriving at diagnoses in a timely and cost-effective manner. Given rapidly advancing technology, the genetics community has moved most of its molecular and clinical resources online to assure more rapid dissemination of information. The currently available resources, while encyclopedic, are limited by the state of organization of the knowledge base. There is no system of levels of evidence for gene-phenotype associations, making it difficult to judge which genes ought to be tested first. Also lacking is a neurodevelopmental phenotype ontology. Finally, laboratories working with WXS data would benefit from a well-curated source of genes associated with disorders of brain development. To address these issues, we designed and implemented a publicly available web-based tool that curates the body of knowledge regarding genes associated with neurodevelopmental phenotypes, assembled the first ontology of those phenotypes, and developed a system of levels of evidence for gene-phenotype associations. This tool is called the Developmental Brain Disorders Database (DBDB), an on-line curated repository of genes, phenotypes, and syndromes associated with neurodevelopmental disorders available at https://www.dbdb.urmc.rochester.edu/home. The current release contains over 700 evidence-ranked gene-phenotype associations, 70 neurodevelopmental phenotypes, and 150 syndromes. While DBDB augments existing web-based resources such as OMIM and GeneReviews, it uniquely offers an online evidence system for neurodevelopmental disorders that is transparent, dynamically linked to the literature, and curated by recognized experts. As new literature emerges, DBDB is easily updated. References remain current as they are served directly from PubMed. When used with other resources, DBDB will streamline the genetic workup of children with relevant disorders for clinicians. For researchers, it provides an evidence-ranked gene list against which results of WXS may be filtered, and facilitates more advanced pathway-based analyses.

1523F

Identifying putative functional variants from GWAS by utilizing ENCODE Consortium data. *J. Hayes[1,2,3], X. Xu[1,2,4], J. Farber[1,2], M. Setty[5,6], A. González[6], A. Perez[6,7], V. Joseph[1,2], K. Offit[1,2], S. Raychaudhuri[8,9], C. Leslie[6], R. Klein[1,2].* 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 2) Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 3) Biochemistry, Cellular, and Molecular Biology Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 4) Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 5) Physiology, Biophysics and Systems Biology Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA; 6) Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 7) Tri-institutional MD-PhD Program, Weill Cornell Medical College, New York, NY, USA; 8) Divisions of Genetics & Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 9) Partners Healthcare Center for Personalized Genetic Medicine, Harvard Medical School, Boston, MA, USA.

The first genome-wide association study (GWAS) was published in 2005. Since then, there have been over 1,300 published GWAS papers spanning over 200 diseases and more than 3,000 disease susceptibility loci identified. However, one of the major challenges of the field is that over 90% of these identified risk variants do not result in an amino acid change, thus failing to explain the mechanism for increased risk. We hypothesize that the GWAS-identified risk-SNPs, or those variants highly correlated with them, increase risk is by altering transcriptional regulatory elements. We tested this hypothesis by analyzing identified risk-SNPs for both lymphoma and chronic lymphocytic leukemia (CLL) together. The lymphoblastoid cell lines (LCLs) in ENCODE make for an ideal dataset in which to test these risk SNPs for B-cell malignancies as LCLs are derived from B cells. When combined, there are 29 non-HLA lymphoma risk SNPs published to date. When we identify tagged SNPs that are in LD with the original SNP with an $r^2 \geq 0.8$ the number of SNPs to be analyzed increases to 352. When we compare these with DNase hypersensitivity data, 23 SNPs lie in hypersensitivity regions. We then overlapped these DNase-SNPs with ChromHMM data and found 11 SNPs that may be functional: 3 SNPs were classified in 'Active Promoter' sites and 8 SNPs classified in 'Strong Enhancers.' The analysis is not limited to DNase-seq and ChromHMM data; the system is being built with flexibility to take in many different types of ENCODE data, including transcription factor ChIP-Seq and histone methylation marks. Additionally, we are developing discriminatively trained transcription-factor binding models to predict the impact that these variants may have on binding. We plan to identify which regulatory marks, or combination thereof, are important in identifying putative functional SNPs and provide a framework by which other groups can accurately and quickly select these SNPs for preliminary experimental follow-up.

1524W

Prioritization of Pathways, Genes and Polymorphisms for Dental Caries. *A. Nicolaou[1,2], M. Shibata[2], SR. Diehl[1,2].* 1) Center for Pharmacogenomics and Complex Disease Research, New Jersey Dental School, UMDNJ, Newark, NJ; 2) School of Health Related Professions, UMDNJ, Newark, NJ.

The main goal of this study was to identify genetic variation that may explain why some individuals are more susceptible than others to dental caries (tooth decay or 'cavities'). The discovery of causal genes and awareness of risk factors is essential for development of personalized prevention and early detection of this major oral disease. Although oral bacteria play an important role in the development of caries, there is strong evidence for genetic risk of dental caries from human twin and family studies. Our study used two sources for selection of genes and Single Nucleotide Polymorphisms (SNPs): 1) the biological roles of genes as reported in Gene Ontology (GO) and 2) genetic association studies reported in literature. We found that GO was limited in the coverage of oral biology. However, good candidate genes were found for enamel, antimicrobial and saliva pathways. After the first Genome Wide Association Study (GWAS) for caries in children was published, an in silico validation of the selected genes/SNPs was done by evaluating the findings. Recently, we have extended this to data from a new GWAS publication for caries risk in adults. No overlap was found between the selected genes/SNPs from our study and those obtained from the GWAS. However, it should be noted that the associations found in the GWAS show only weak to moderate statistical support. Therefore, it appears likely that common variants with small effects cannot account for the high estimated heritability for caries and the candidate genes obtained in our study may be helpful for prioritization of rare variants identified by Whole Exome Sequencing (WES) approaches.

1525T

The human Genome Clinical Annotation Tool (h-GCAT): enabling clinicians with little bioinformatics background to analyze whole exome data. *K. Wierenga[1], Z. Jiang[2], N.F. Tsinoremas[2].* 1) Pediatrics, OUHSC, Oklahoma City, OK; 2) Center for Computational Science, University of Miami, Miami, FL.

Background. Whole exome sequencing (WES) and whole genome sequencing (WGS) are the most powerful technologies currently available to clinicians to provide patients and families with an etiologic diagnosis that remained elusive until then. Currently, only a few laboratories provide WES for clinical diagnostics. The clinician involved in the care of patient(s) and family is typically not involved in the data analysis. These developments risk demoting the clinical team to bystanders, and could disrupt physician engagement. Approach/Results. We developed the human Genomic Clinical Annotation Tool (h-GCAT), a server-based software tool that allows the user to upload a variant call format file (vcf), containing the variants of patient(s) and relevant family members. The tool allows quick selection of relevant cut-offs for sample sequencing depth and genotype quality, minor allele frequencies (using 1kG and ESP5600), and of relevant chromosomes. Subsequently, the user continues to filter using pedigree information and type(s) of mutation(s) to be ascertained (known/novel SNPs, gene/exon/coding region, synonymous/non-synonymous SNPs). The user can also filter by clinical annotation, using tools as HGMD, DO, OMIM and HPO. Regions of homozygosity can be analyzed in cases with consanguinity. At the end of this process, the user can review the calls that 'survived' these iterative filters. The result page provides relevant information of these calls: genomic location of the candidate calls, dbSNP status, cDNA location of calls for the various transcripts, as well as residue change by protein isoform(s), if applicable. Link Outs to the NCBI, UCSC Genome Browser, OMIM, HGMD and DO entries are provided for review. We tested our tool on various training sets given certain inheritance patterns. Within less than 10 minutes results similar to other WES analysis tools were obtained. Conclusion. While the software underlying the various algorithms of filtering activities in h-GCAT is intricate, the process of filtering out (likely) irrelevant calls and performing a directed search for relevant mutation(s) and associated phenotypes is easily understood and accomplished by a clinician with little bioinformatics background. The clinician is aware of key clinical features that can inform phenotyping, becoming more important with continued identification of Mendelian disorders. It is our opinion that variant call files should be made available to the clinician when requested.

1526F

SeqSIMLA: An efficient simulation software for generating sequences and phenotypes for complex disease studies. *R.H. Chung, C.C. Shih.* National Health Research Institutes, Zhunan, Taiwan.

Association studies based on next-generation sequencing (NGS) technology have become popular, and statistical association tests for NGS data have been developed rapidly. A flexible tool for simulating sequence data in either unrelated case-control or family samples with different disease and quantitative trait models would be useful for evaluating the statistical power for planning a study design and for comparing power among statistical methods based on NGS data. To our knowledge, the software SimRare is the only tool designed specifically to simulate sequence data with phenotypes. However, SimRare focused on generating unrelated samples. As family-based association studies using NGS are also important, software that can simulate sequence data in families will be very useful for evaluating the properties of family-based NGS analysis. We developed a simulation tool, SeqSIMLA, which can simulate sequence data with user-specified disease and quantitative trait models. We implemented two disease models, Models 1 and 2, based on logistic functions. In Model 1, the user can flexibly specify the odds ratios of disease variants and disease prevalence. In Model 2, the user can specify the overall population attributable risk for disease variants and the baseline penetrance. Risk and protective variants can also be simulated. We do not have restrictions on the number of disease loci to be simulated. We also implemented a quantitative trait model, in which the user can specify the number of quantitative trait loci (QTL), proportions of variance explained by the QTL, and genetic models. We compiled recombination rates from the HapMap project so that genomic structures similar to the real data can be simulated. Both unrelated case-control and three-generation family data can be simulated. Java threads are used to parallelize the code. SeqSIMLA can efficiently generate a replicate of 500 extended families in a 5MB region containing 50 haplotype blocks in 4 minutes. SeqSIMLA will be very useful for evaluating statistical properties for new study designs and new statistical methods using NGS. SeqSIMLA can be downloaded at http://seqsimla.sourceforge.net.

1527W

RegScan: a tool for rapid estimation of allele effects on continuous traits and their combinations. *T. Haller[1], M. Kals[1], T. Esko[1,2,3], R. Mägi[1], K. Fischer[1].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Broad Institute, Cambridge, MA, USA; 3) Children's Hospital Boston.

Genome-wide association studies (GWAS) are becoming computationally more difficult with the ever-increasing number of samples, markers, and traits. The combinatorial traits (for example the ratios of metabolite concentrations) can increase the data size beyond what the current GWAS tools can handle. At the same time the scientific interest to study combinatorial traits is rapidly increasing and the field needs a quick analytical tool. We addressed the issue of increasing need for computational power in GWAS by creating an application called RegScan. Our tool is designed for performing basic linear regression analysis with very large data sets maximally fast. It performs association analysis between markers and continuous combinatorial traits ten to hundreds of times faster than the other tools. RegScan specifically targets association analysis of combinatorial traits in metabolomics, however, it is not limited to that. Another area of use is studying marker associations with gene expression. RegScan can automatically create the combinatorial traits for analysis. It also comes with several supporting functions used for filtering of the results and additional analysis of the detected associations with the focus to determine their significance. RegScan has been successfully tested in linear regression analyses of large data sets such as 1000 Genome imputed data for thousands of individuals with over 6000 combinatorial traits. The analyses that have typically taken several weeks or months can now be carried out in just days. The results allow researchers to investigate the links between genetic markers and biological pathways. RegScan is an open source project. It can be freely downloaded together with the instructions and examples at www.biobank.ee/regscan.

1528T

Shortening the Diagnostic Odyssey: Integrating Genomic, Structural, and Phenotypic Information to Reduce Time of Rare Disease Diagnosis. *R. Hariharan[1], M. Bhat[2], S. Agarwal[1], S. Krishna[1], B. Panda[1], V. Veeramachaneni[1].* 1) Strand Life Sciences, Bangalore, India; 2) Center for Human Genetics, Bangalore, India.

Genomic sequencing for diagnosis of rare, and often complex, diseases is gaining traction to help end what is often a diagnostic odyssey for families and patients. However, knowledge of genetic variants alone is often not enough to diagnose disease - understanding the effects of variants on gene and protein structure inform how the variant may affect the role in biological function while incorporating phenotypes and symptoms help explain how the variants manifest in patients. To that end, we are developing software that combines our own tool for analysis of NGS data (AVADIS NGS) with systematically curated literature content and bioinformatics databases (to integrate genomic, structural and phenotypic information) to enable data to report generation in one single step. This will substantially compress the time needed for clinicians to correctly diagnose patients and start them on the right treatment regimen. Here, we present a case study wherein, through collaboration with a clinical geneticist, we have sequenced and applied our approach to a family in India. The young couple, of a consanguineous marriage, had previously given birth to two boys with pulmonary hypertension and respiratory problems who died suddenly under two years of age. The couple was again pregnant and wanted to determine whether this child would experience similar outcomes. Through analysis of whole exome sequencing data of the parents, fetus, and one of the deceased brothers with our software, we were able to provide an answer.

1529F

Gene Expression Deconvolution using Single-cells. *J. Lindsay[1], I. Mandoiu[1], C. Nelson[2].* 1) Department of Computer Science and Engineering, University of Connecticut. Storrs CT; 2) Department of Molecular and Cell Biology. University of Connecticut. Storrs CT.

Obtaining whole-transcriptome expression profiles of closely related cell types is challenging for stem-cell biologists. Here we present an approach that utilizes single-cell qPCR probing of a small number of genes to aid in the deconvolution of whole-transcriptome profiles of mixed samples. Typically the expression profiles of a given mixture of cells is modelled as linear combination of the signature of its constituent cells multiplied by the concentration of each each cell type in the mixture. Existing approaches to deconvolution methods attempt to estimate both the cell type signatures and concentrations simultaneously, or separately if knowledge of one is know beforehand. Our method first obtains a reduced profile of constituent cell-types from single-cell samples by using k-means clustering and then averaging all cell-types in each cluster. Then we apply a robust quadratic programming method to inferring mixture proportions of mixed sample. Finally we have implemented a second quadratic program for inferring cell-type specific expression levels of genes not measured directly in single-cells based on mixture proportions derived for each mixed sample. Using real single-cell data obtained from the posterior Node-Streak-Border region of a mouse embryo and 100 simulated mixtures, a leave-one-gene-out experiment found our method estimates of concentrations had a RMSE of 0.03 and the missing gene estimates had a correlation of 0.997.

1530W

Implementing a High Performance, Reusable Consensus Calling Pipeline for Next Generation Sequencing using Globus Genomics. *R.K. Madduri[1], A. Rodriguez[1], V. Trubetskoy[2], L.K. Davis[2], P.J. Dave[1], N.J. Cox[2], I.T. Foster[1].* 1) Computation Institute, University of Chicago, Chicago, IL; 2) Section Genetic Medicine, University of Chicago, Chicago, IL.

We developed Globus Genomics (http://globus.org/genomics/), an end-to-end hosted service designed to efficiently and easily analyze large quantities of Next Generation Sequencing (NGS) data using state of the art algorithms, efficient data management tools, a graphical web-based workflow environment and on-demand computing infrastructure.Globus Genomics leverages a collection of existing cloud-based services. Globus Genomics users, however, can build new analysis workflows from scratch. Users can analyze large amounts of data using computationally efficient analytical pipelines and cutting edge tools that leverage the power and flexibility of on-demand cloud computing resources~without being exposing to the complexities of managing large scale infrastructure; deploying and configuring analysis tools; transferring data between sequencers, analysis nodes and storage systems; or managing their own users and groups. To this end, we use elastic computational infrastructure provided by Amazon Web Services. We use the Condor scheduler to manage a dynamically assembled pool of hosts. We outsource high performance data transfer and user, group and credential management to Globus Online, a platform as a service (PaaS) provider also developed and operated by our team. Finally, we host a Galaxy workflow system to enable easy to use graphical workflow orchestration. We created computational profiles for multiple variant calling and genotyping algorithms available for academic use (i.e., GATK2.0, Atlas2.0, and FreeBayes toolkits). These profiles enable high performance, scalable execution of algorithms on hundreds of raw data sets. We built reusable, robust pipelines using different computational modalities that best suited the underlying analysis. The resulting variant calls from each pipeline can then be fed to a consensus-calling algorithm (Consensus Genotyper for Exome Sequencing CGES; see Trubetskoy et al., ASHG 2013) resulting in high quality variant and genotype calls. We have run these three pipelines in parallel calling variants on over a hundred raw BAM files in the course of three days. Atlas2.0 and the GATK pipelines took a little over two days to finish execution while Freebayes pipeline took a little over three days. In conclusion, we present the workings of Globus Genomics, a robust, powerful, and user-friendly suite of tools for NGS analysis empowering geneticists and enabling translational discovery relevant to human disease.

1531T

Genome-wide structural variation analysis with genome mapping on nanochannel arrays. *A.C.Y. Mak[1], J.J.K. Wu[2], Y.Y.Y. Lai[1], K.Y. Yip[3,5], T.F. Chan[4,5], E.T. Lam[1], T.P. Kwok[3], J.W. Li[4,5], A.K.Y. Leung[4], A.K.Y. Yim[4,5], M. Xiao[6], P.Y. Kwok[1], S.M. Yiu[2].* 1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 2) Department of Computer Science, The University of Hong Kong, Hong Kong; 3) Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong; 4) School of Life Sciences, The Chinese University of Hong Kong, Hong Kong; 5) Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong; 6) School of Biomedical Engineering, Science & Health Systems, Drexel University, Philadelphia, PA.

Despite recent advances in next-generation sequencing technology, genome-wide structural variation (SV) detection using 'short reads' remains challenging. Detection of large and/or balanced SV such as inversion or translocations is difficult, if not impossible. To overcome the limitations of short reads, we generated genome maps using a novel approach that allows very long DNA molecules (> 150kb) fluorescently labeled at Nt.BspQI sites (GCTCTTCN/) to be linearized and imaged in highly parallel nanochannel arrays [1]. We obtained data to 50X genome coverage on NA12878, a member from a CEPH CEU trio extensively sequenced in the 1000 Genomes Project. To detect structural variations from these genome maps, we have developed a BreakPoint sensitive Optical Map Dynamic Programming algorithm (BP-OMDP). Data for each DNA molecule is represented as an array of inter-label distances or segments. BP-OMDP allows molecules to be partially aligned to the hg19 in silico Nt.BspQI reference with mismatching segments on either end of the molecules. Potential SV regions and their breakpoints are identified by observing a clear drop in the number of aligned molecules to the reference. SV regions are then reconstructed and verified by de novo assembly and re-alignment of partially mapped and originally unmapped molecules. Using BP-OMDP, we detected 78 putative inversions, of which 59% overlap with known inversions found in NA12878. Extension of our method to detect other balanced and unbalanced SV is underway.

1532F

GeneZip: A gzip-based C library for compact imputed genotype storage in RAM. *C. Palmer, I. Pe'er.* Center for Computational Biology & Bioinformatics, Columbia University Medical Center, New York, NY.

Genome-wide association studies directly assay ~10^6 single nucleotide polymorphisms (SNPs) in a study population. With access to phased reference samples of common ancestry that have genotype information at a denser panel of variants, a sample of 10^6 SNPs can be extended probabilistically by a factor of 10- to 40-fold. These resulting datasets are standardly stored as text files that detail imputed genotype probabilities, requiring 12 characters per SNP per sample, thus are too large to be simultaneously stored in RAM losslessly as is. Instead, the maximum likelihood genotype at each site is used, or SNPs are processed and discarded one at a time. Neither of these adaptations are optimal: missingness at individual sites may be informative to a trait, and multi-locus testing is complicated. Here, I present a C/C++ library that dynamically compresses probabilistic genotype data as they are loaded into memory. This method uses a customization of the DEFLATE (gzip) algorithm, and maintains constant-time access to any SNP (but linear O(# individuals) access to a given genotype). Briefly, the probabilistic genotype matrix is stored one SNP at a time, with each SNP data being compressed. Genotype calls for individuals for a particular SNP require serial access, but indexing the compressed matrix for where does information for each SNP reside allows random access to any marker. Average compression ratios of ~90% are observed in test data.

1533W

VariantMaster: a novel platform to identify causative variants from HTS data in familial, denovo and somatic genetic disorders including cancer. *F.A. Santoni[1], P. Makrythanasis[1], S. Nikolaev[1], M. Guipponi[2], D. Robyr[1], A. Bottani[2], S. Antonarakis[1,2,3].* 1) Department of Medical Genetics and Development, University of Geneva, Geneva, Switzerland; 2) Geneva University Hospital - HUG; 3) . iGE3 institute of Genetics and Genomics of Geneva, Switzerland.

There is a fast growing interest in clinical genetics to the employment of High Throughput Sequencing data for accurate diagnosis of monogenic diseases. Furthermore, whole exome sequencing data introduced a significant advance in the comprehension of cancer development by the identification of driver somatic variants. To the aim of providing an accurate and efficient methodology in these contexts, we developed VariantMaster, an original and comprehensive methodology that extracts likely causative variants in familial and sporadic genetic diseases considering different modes of inheritance: X-linked, autosomal dominant, and recessive (homozygosity or compound heterozygosity), denovo germline and somatic mutations including cancer. For the highest accuracy the algorithm takes into account predicted variants and integrates the phenotypes with family pedigrees. A probabilistic approach based on row data is furthermore applied to robustly evaluate the likelihood of the occurrence of a putative causal variant in each family member. Additionally, VariantMaster can incorporate several layers of information as, for example, genotypes, allelic frequencies and damaging scores. VariantMaster is already employed as an effective tool in the clinical genetics department of the Geneva University Hospital (HUG) and in several research projects as the identification of the burden of denovo mutations in a cohort of 58 trios to elucidate the molecular basis of schizophrenia and in the identification of novel causative genes in Primary Ciliary Diskinesia (PCD). Furthermore, to prove the effectiveness of the methodology on the detection of somatic variants we extract 9 random samples of colorectal cancer tumors from the TCGA consortium. Remarkably, 80% of the variants identified by VariantMaster were validated and annotated in COSMIC. Moreover, VariantMaster detected two novel putative 'driver' variants. In general VariantMaster demonstrated to be more flexible and accurate compared with previously published algorithms and we believe it has the potential to become an indispensable tool in the investigation of genetic diseases and molecular cancer profiles.

1534T

Accuracy of predictions of deleterious effects on protein function of multiple dominant negative and activating mutations in STAT3: comparison of nine computational methods. *M.S. Smith.* Research and Development, Viracor-IBT Laboratories, Lee's Summit, MO.

Purpose: STAT3 is activated in response to external signals. After tyrosine phosphorylation of the receptor, STAT3 is recruited by its SH2 domains, is phosphorylated, dimerizes and translocates to the nuclease and binds to specific promoter sequences. Many dominant-negative STAT3 mutations have been identified from cases of hyper IgE syndrome, and activating mutations from large granular lymphocytic leukemia (T-LGL) and chronic NK cell lymphoproliferative disorder (CLPD-NK). We investigated the ability of 9 computational methods to predict the effect of these mutations on protein function. Methods: Nonsynonymous mutations were initially detected in two patient specimens by STAT3 sequencing. These and other known STAT3 hyper IgE syndrome mutations, as well as activating mutations from T-LGL and CLPD-NK cases were examined using programs designed to predict the pathogenicity of amino acid substitutions. The programs were PolyPhen-2, SIFT, MutPred, fathhm, SNPs&Go, nsSNP Analyzer, EvoD, SNAP, and PON-P. These methods vary widely in approach, from protein alignments, statistical approaches, and biochemical aspects such as solvent accessibility and predicted residue flexibility. In the literature, many of these programs were previously tested by benchmarked experimental data sets of neutral and pathogenic mutants from proteins such as E. coli LacI repressor, phage T4 lysozyme, and HIV-1 protease mutations. Summary: For 32 hyper IgE mutations in the STAT3 DNA-binding, SH2, and transactivation domains (the majority reported in multiple patients each) and 10 mutations in the dimerization interface of the SH2 domain from leukemia patients, we examined agreement across all programs or among a subset of programs with reported higher accuracy levels. For Hyper IgE syndrome, the use of an overall score from 5 (SIFT, MutPred, SNPsGo, SNAP, and PON-P) or all 9 programs was in general agreement for the most common mutations, including some predicted in contact with DNA or involved in SH2 domain dimerization; however, not all common mutations showed a significant score in this way. For the somatic mutations seen in T-LGL, CPLPD-NK, and hepatic adenomas, where the mutations are thought to be activating, there was poor agreement with the scoring by these programs. Thus, for a large set of dominant negative and activating mutations of STAT3, no single or small group of analysis programs were found to be able to accurately predict the effect of a mutation.

1535F

A Graphical Quality Control Tool for Next Generation Exome Sequencing. *J.D. Smith, S. McGee, D.A. Nickerson.* Genome Sciences, University of Washington, Seattle, WA.

The wide-scale application of massively parallel sequencing requires the development of new approaches to automate data quality control. These approaches should accurately and efficiently pinpoint potential problems and evaluate quality throughout the process. Additionally, the quality metrics associated with the data need to be presented in a concise but informative format to facilitate rapid decision-making about whether to proceed to interpretation, or to remove or hold a sample for further assessment. To facilitate this process, we have developed a "QC Dashboard" that presents an overview of quality metrics, and is currently in use for exome sequencing at the Northwest Genomics Center (NWGC). The QC Dashboard is a graphical and textual display of statistical information extracted from BAM files, and uses multiple open-source analysis tools. Plots include overall number of reads, reads mapped to target and number of unique reads; allele distributions at each sequencing cycle; plots for read depth coverage, uniformity and insert size distribution. Per sequencing cycle error rates and read qualities are also plotted and summarized. Sample complexity is quickly and accurately extrapolated from a subset of data and displayed. Each dashboard is adaptable to individual- or merged-lane data and can easily be implemented in any analysis pipeline for automatic display.

1536W

Evaluation of imputation method for classical HLA-DRB1 using a Finnish dataset. *E. Vlachopoulou[1], E. Lahtela[1], A. Wennerström[1], A.S. Havulinna[3], P. Salo[3,4], M. Perola[3], V. Salomaa[3], M.S. Nieminen[2], J. Sinisalo[2], ML. Lokki[1].* 1) Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland; 2) HUCH Heart and Lung Center, Division of Cardiology, Helsinki University Central Hospital, Finland; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland.

HLA genes are located at the MHC region (6p21) revealing high degree of gene density and polymorphism. Allelic structures of HLA genes show broad linkage disequilibrium and are inherited as haplotypes with significantly different frequencies between populations and ethnicities. HLA genes have been associated with many autoimmune, infectious and inflammatory diseases. Due to vast amount of alleles, the high-resolution HLA typing is expensive and time-consuming. Scientists have attempted to develop computational approaches to define HLA alleles with high confidence. We tested the reliability and performance of a widely used software HLA*IMP (versions HLA*IMP:01 and HLA*IMP:02) for imputing classical HLA-DRB1 alleles in the Finnish material. Unrelated samples (n=161) from Finnish subjects were selected from the FINRISK 1997, 2002 and 2007 cohorts. The samples were SNP genotyped with Illumina 610K genotyping chip and the genotypes were called in the quality control (QC) using Illuminus. The cleaned dataset was imputed with MACH 1.16 using HapMap 2 release 22 CEU reference haplotype set. The results of imputed HLA-DRB1 alleles were compared with the high-resolution results obtained from sequence-based typing. The frequencies of the imputed HLA-DRB1 alleles were similar to the frequencies from sequence-based typing. However, the per-individual success rate was 27.64% using either HLA*IMP:01 or HLA*IMP:02 having mean Q and Q2 posterior probabilities for correct inferences of 0.9539 and 0.9286, respectively. When we used the posterior probability of 0.95 as a threshold for both imputed alleles simultaneously, the per-individual success rates were as low as 23.74% and 23.39% with HLA*IMP:01 and HLA*IMP:02, respectively. Hence, the higher confidence decreased the correct inferences. One of the most prominent example was HLA-DRB1*01:01 allele showing approximately 30% success rate while being the most common wrongly imputed allele. In Finland, isolation and migration history have shaped the gene pool narrower showing HLA haplotype frequencies typical to Finnish population when compared to Europeans. When we used HLA*IMP having different European populations as references, the imputation success for HLA-DRB1 alleles was very low pointing to the importance of population specific reference material.

1537T

An adaptive permutation procedure to estimate the significance threshold for the minimum p-value of multiple permutation tests. *P. Yajnik, M. Boehnke, H. Jiang.* Dept Biostatistics, University of Michigan, Ann Arbor, MI.

The analysis of modern genetic association studies often involves testing multiple hypotheses. The Bonferroni correction is commonly used to approximate p-value significance thresholds to maintain the familywise type I error rate. When tests are correlated, the Bonferroni correction may be conservative and permutations can be used to estimate significance thresholds.

Permutation based estimates are computationally expensive. The computational burden is exacerbated if p-values of the tests also need to be obtained by permutation when analytical values based on asymptotic theories are unavailable/inaccurate. We propose an adaptive permutation procedure which can greatly reduce the total number of permutations needed to estimate the minimum p-value of multiple permutation tests. At each iteration, the pool of tests included in the procedure may be reduced by discarding tests whose estimated p-values are much larger than the currently estimated minimum p-value. The likelihood-ratio test (LRT) statistic is used to discard tests. The proposed estimator is the minimum estimated p-value amongst the tests that remain at the end.

We performed 10000 simulations to assess the performance of the procedure. The simulation parameters included the number of multiple tests (100, 200, 500, 1000 or 10000), the correlation between blocks of tests (0, 0.2, 0.4, 0.6, 0.8 or 0.95) and the stringency used to discard tests (LRT statistics exceeding 2.7, 3.8, 6.6 or 10.8). At the highest stringency, the proposed method required on average only 0.2%-5% as many permutations as the naive method (the relative efficiency increasing with number of multiple tests). At this stringency, the proposed method achieved the exact minimum p-values found by the naive method in 99% of the simulated cases.

We also applied this procedure to a real dataset. A burden test (SKAT) was performed with each of 13 correlated phenotypes regressed on variants from 62 genes (806 tests with a Bonferroni corrected threshold of 6e-05). It took our procedure 12 hours with 40 CPUs to provide an estimated threshold of 7e-05. Per our simulations, the naive approach would take 20~500x longer than the proposed approach. Based on a modest estimate of 100x, the naive approach would require 1200 hours.

1538F

BDgene: a genetic database for bipolar disorder and its overlap with schizophrenia and major depressive disorder. *S. Chang, L. Guo, J. Wang.* Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences. 16 Lincui Road, Chaoyang District, Beijing, China.

Bipolar disorder (BD) is a common psychiatric disorder with complex genetic architecture. It shares overlapping genetic influences with schizophrenia (SZ) and major depressive disorder (MDD). Large numbers of genetic studies of BD and cross-disorder studies between BD and SZ/MDD have accumulated numerous genetic data. There is a growing need to integrate the data to provide a comprehensive data set to facilitate the genetic study of BD and its highly relevant diseases. To fulfill this demand, BDgene database was developed to integrate BD-related genetic factors and shared ones with SZ/MDD from profound literature reading. Through depth-mining of 796 papers, BDgene contains multiple types of literature-reported genetic factors of BD with both positive and negative results, including 797 genes, 3119 SNPs, and 789 regions. Shared genetic factors such as SNPs, genes, and regions from published cross-disorder studies among BD and SZ/MDD were also presented, including 285, 120 and 49 shared genes for BD-SZ, BD-MDD and BD-SZ-MDD respectively. On the basis of data from the literature, in-depth analyses were performed for further understanding of the data: gene prioritization analysis for literature-origin genes obtained 43 BD core genes; pathway-based analysis for genome-wide association study data identified 70 BD candidate pathways; intersection analysis of multidisease candidate genes got 127, 79, and 107 new potential cross-disorder genes for BD-SZ, BD-MDD, and BD-SZMDD respectively; pathway enrichment analysis for BD core genes showed the majority of the pathways were involved in synaptic transmission, membrane and ion channel activity; pathway enrichment analysis for literature-reported positive genes shared by BD-SZ and BD-MDD showed two rhythmic related pathways were enriched by both gene lists. As a central genetic database for BD and the first cross-disorder database for BD and SZ/MDD, BDgene provides not only a comprehensive review of current genetic research but also high-confidence candidate genes and pathways for understanding of BD mechanism and shared etiology among its relevant diseases. To facilitate better usage of the database, BDgene provided powerful search tools to access the data, and a forum to share or exchange ideas. BDgene will be updated quarterly for both the literature-origin data and in-depth data analyses result to maintain an update-to-date resource. BDgene is freely available at http://bdgene.psych.ac.cn.

1539W

PhenoVar: an innovative approach in clinical genomics for the diagnosis of polymalformative syndromes. *C. Buote[1], Y. Trakadis[2], JF. Therriault[1], H. Larochelle[3], S. Lévesque[1].* 1) Dept. of Paediatrics, division of medical genetics, Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, Canada; 2) Dept. of Medical Genetics, McGill University Health Center, Montreal, Canada; 3) Dept. of Informatics, Faculty of Sciences, University of Sherbrooke, Sherbrooke, Canada.

BACKGROUND: Widespread clinical use of Exome Sequencing (ES) requires an approach addressing issues such as incidental findings and accurate prediction of causal mutations among massive amount of variations. We have previously proposed a phenotype-driven analysis of exome data to address these issues and predict diagnosis in the clinical setting. Here we provide a more extensive validation of our software PhenoVar and present a pilot study of 15 patients with undiagnosed polymalformative syndromes. METHODS: We tested the efficiency of our software, PhenoVar, in predicting diagnoses of simulated test-patients with known polymalformative syndromes. Control exome variants files, modified to include a previously published mutation, were combined to the corresponding published phenotypic data of twenty patients with varied polymalformative syndromes. For each test-patient, modified variant file was input along with blindly selected phenotypic traits in PhenoVar for diagnosis prediction. PhenoVar calculated phenotypic similarity, using Human Phenotype Ontology terms, between the test-patient and a database consisting of simulated and real patients for each OMIM phenotype with known molecular basis. Final diagnostic score, assigned to a given OMIM entry, took into consideration both the patient's phenotype and variations. Resulting OMIM entries list was sorted according to diagnostic score and filtered using a minimal phenotypic weight threshold to prevent undesired discovery of incidental findings. In the pilot study, we recruited 15 patients with unknown polymalformative syndrome, after chromosome microarray and standard molecular investigations were done. ES was performed on a HiSeq 2000 (Illumina), following target enrichment using Sureselect All Human exon v4 (Agilent). Variations were called using GATK pipeline. Resulting variants file and phenotypic traits were input in PhenoVar to perform diagnostic prediction as described. RESULTS: In 15/20 simulated test-patients, the expected diagnosis ranked within top 5 predicted most likely diagnoses and this was the case in 18/20 test-patients when known pathogenic variants were priorized. No incidental findings were found using our minimal phenotypic threshold. The analysis of the pilot study 15 patient's exomes is underway. CONCLUSION : Results suggest that this 'phenotype-driven' approach could be applied effectively in the clinical setting for the diagnosis of polymalformative syndromes.

1540T

A joint latent factor model for gene expression data with confounders. *C. Gao[1], C.D. Brown[4], B.E. Engelhardt[1,2,3].* 1) Institute for Genome Sciences and Policy, Duke University; 2) Department of Biostatistics and Bioinformatics, Duke University; 3) Department of Statistical Science, Duke University; 4) Department of Genetics, University of Pennsylvania.

With high dimensional sequencing data being generated at an exploding rate, interpretability is especially critical for problems in genome sciences. One important goal is to understand the relationships between genes from gene expression data, where there may be substantial biological and technical noise associated with measurements of expression levels. Sparse latent factor models have been used successfully to extract interpretable relationships from high dimensional data, but generally the data are preprocessed using principal components, for example, to control for these confounders. In this work, we develop a Bayesian sparse latent factor model that uses a three parameter beta prior to flexibly model regularization on the loading matrix. We add a simple two-component mixture to model each factor loading as a sparse or a dense component. In this setting, sparse factors select small numbers of co-regulated genes, such as genes that are co-regulated by a single transcription factor, whereas dense factors identify confounding factors that impact the expression levels of large numbers of genes simultaneously, such as sex or batch effects. This framework has three advantages over the traditional approaches: i) it represents a joint model of confounders and signal; ii) it avoids the problem of overfitting by regularizing both sparse and dense factors; iii) each factor has an explicit posterior probability for whether it is dense or sparse. Furthermore, our approach automatically estimates the number of sparse and dense factors from the data. To address the known identifiability issues for factor analysis models, we developed measures to quantify robustness for both sparse and dense factors. Using simulated data featuring both sparse and dense structure, we found that we are much more successful at recovering the true latent structure relative to two-stage approaches and other factor models. We applied our model to gene expression data and found that it identifies known covariates and small groups of co-regulated genes. We use additional methods to validate the identified sparse and dense clusters, and find interesting overlap in the genes involved in sparse clusters. Finally, we performed an eQTL analysis to the identified co-related subsets of genes, and we find that we can identify a large number of eQTL hotspots that are not identified in a univariate analysis.

1541F

A Clinical History Weighting Algorithm Accurately Classifies BRCA1 and BRCA2 Variants. *K.R. Bowles[1], B. Morris[2], E. Hughes[2], J. Eggington[1], L. Esterling[1], B. Robinson[1], A. Van Kan[1], B. Roa[1], E. Rosenthal[1], A. Gutin[2], R. Wenstrup[1], D. Pruss[2].* 1) Myriad Genetic Laboratories, Inc., Salt Lake City, UT; 2) Myriad Genetics, Inc., Salt Lake City, UT.

Genetic testing, including full gene sequencing and large rearrangement analysis for germline *BRCA1* and *BRCA2* mutations, can identify individuals with Hereditary Breast and Ovarian Cancer syndrome. Current genetic analysis identifies *BRCA1* and *BRCA2* deleterious mutations as well as variants of unknown clinical significance. Reclassification of uncertain variants to more clinically interpretable categories is critical for patient management. We have developed a statistical algorithm that aids in the assignment of clinical classifications to uncertain variants. This algorithm is based on the premise that disease-associated mutations will be observed more often in individuals at high risk for carrying a mutation, as determined by personal and family history. Statistical analysis weights the family histories of each proband carrying the variant of interest and compares these histories to those of control probands carrying variants known to be benign or deleterious. Data from over 400,000 probands were utilized for algorithm development. This technique was validated by and used to analyze over 6000 *BRCA1* and *BRCA2* variants. The algorithm successfully classified well-documented variants such as *BRCA1* c.181T>G (Deleterious), *BRCA1* c.1065G>A (Polymorphism), and *BRCA2* c.2808_2811del (Deleterious). The *BRCA1* c.5096G>A (Suspected Deleterious with reduced penetrance) and *BRCA2* c.7878G>C (Suspected Deleterious with reduced penetrance) mutations were classified as "Not Callable" by the algorithm, consistent with their previous hypomorphic interpretations. This 'history weighting' algorithm allows for the accurate reclassification of *BRCA1* and *BRCA2* uncertain variants and improved clinical management of at-risk patients. The history weighting algorithm has also been successfully applied to reclassification of variants in the Lynch syndrome-associated genes *MLH1* and *MSH2*, and it is currently being modified to allow for analysis of germline variants in *MSH6*. With additional modifications, this algorithm is expected to be applicable to other autosomal dominant cancer-associated and non-cancer-associated genes.

1542W

GenAP: automated analysis of genetic HTS data for clinical diagnostic use. *M.C. Eike[1], T. Håndstad[1], E. Nafstad[2], T. Hughes[1], T. Grünfeld[1], D.E. Undlien[1].* 1) Department of Medical Genetics, Oslo University Hospital, University of Oslo, Oslo, Norway; 2) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.

High-throughput sequencing (HTS) presents both great opportunities and challenges for use in clinical diagnostic settings. One of the most pressing needs is efficient analysis of the increasing volumes of generated data, without compromising quality. In response to this, we are developing a system for rules-based, semi-automated analysis of sequence variation data for diagnostic use, as part of the project *Norwegian clinical genetic Analysis Platform* (genAP). The aim is not only to speed up the analytic process, but also to achieve higher levels of reproducibility and integrity of results.

The input consists of annotated sequence variants, including various sequence and variant calling quality metrics and annotation data from external sources (including frequency data, pathogenicity prediction, mutation databases and literature references). Importantly, when available, variants are also annotated with results from previous analyses stored in an in-house database, which are given a higher priority than external data. The system then uses this information to classify variants along the axis pathogenic-neutral, generating suggested conclusions for each variant. The rules applied in this classification process are specific for each clinical question (but easily editable) and are based on standard operating procedures at our department. The generated results are presented in a graphical user interface, which includes a structured display of all annotation information. Here, variants are prioritized according to their likely involvement in disease, and key information underlying the suggested classification is highlighted. The user then has the choice to edit or approve the results, which upon approval are added to the in-house database.

We aim to make the source code for the system publically available, with a beta release planned by the end of 2013. Although the release will be tailored for our department and particular clinical questions (currently breast/ovarian cancer and cardiomyopathy), it is our goal to make the system flexible enough to be adapted by other parties.

This presentation will demonstrate the logic of the system, with specific examples and an overview of the general architecture.

1543T

Variation data services at NCBI: archives, tools, and curation for research and medicine. *S. Sherry, K. Addess, V. Ananiev, C. Chen, D. Church, M. Feolo, J. Garner, T. Hefferon, D. Hoffman, B. Holmes, M. Kholodov, A. Kitts, J. Lee, J. Lopez, D. Maglott, R. Maiti, L. Phan, G. Riley, W. Rubinstein, D. Rudnev, Y. Shao, E. Shekhtman, K. Sirotkin, D. Slotta, R. Tully, R. Villamarin-Salomon, Q. Wang, M. Ward, H. Zhang, C. Xiao.* National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

NCBI operates several archives for sequence variation data that are increasingly relevant for medical research. These archives contain general variation data in the public domain and high density surveys of genetic diversity in medical study populations. Small scale variations are accessioned and distributed through dbSNP at http://www.ncbi.nlm.nih.gov/snp/ and larger structural variations (greater than 50bp) are accessioned and distributed through dbVar at http://www.ncbi.nlm.nih.gov/dbvar/. Research participant genotypes, participant phenotypes and analysis results are distributed through the dbGaP controlled access system http://www.ncbi.nlm.-nih.gov/dbgap/. Assertions of clinical significance for variants and alleles are accessioned and distributed through ClinVar at http://www.ncbi.nlm.nih.gov/clinvar/. Tools for exploring and visualizing variation data include the 1000 genomes browser http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/, the 1000 genomes data slicer http://trace.ncbi.nlm.nih.gov/Traces/1kg_slicer/ and the phenotype-genotype data integrator http://www.ncbi.nlm.nih.gov/gap/PheGenI. The clinical remap tool at http://www.ncbi.nlm.nih.gov/genome/tools/remap#tab=rsg will provide sequence coordinates for variations on a clinical RefSeqGene record, and the variation reporter service at http://www.ncbi.nlm.nih.gov/variation/tools/reporter will provide a list of known variants and the functional consequences for a region of interest in BED format or set of variants in HGVS or GVS format. The presentation will show how these services are organized in relation to the primary data archives of GenBank, SRA, GEO, dbSNP, dbVar and dbGaP, and how they can be used as input into external services for research, clinical practice or cloud-based analysis projects.

1544F

T2D2seq: An enhanced, novel algorithm to determine differential expression in RNA-Seq data. *E.D. Au, M.H. Farkas, J.A. White, E.A. Pierce.* Ocular Genomics Institute, Berman-Gund Laboratory, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA.

As RNA-Seq methods continue to evolve, and the technology continues to be more widely used, it is imperative that data analysis becomes more uniform. RNA-Seq studies are unmatched in terms of the wealth of information they provide; the ability to study differential expression on the level of individual exons and introns allows for a more precise analysis of the effect of mutations and has promise to offer new insight into the potential cause of disease. The amount of data generated by sequencing studies and the potential knowledge that can be gained is constantly increasing, but analysis algorithms have not kept up with the high demand of these studies to date. While many algorithms are available, the discordance among the results produced is highly concerning. Using real RNA-Seq data as a guide, we have developed a novel method to determine differential expression that is both efficient and precise. A limiting factor in the development of analysis algorithms is the quality of model datasets used to test their performance. Currently published synthetic datasets are poor models of RNA-Seq data: they contain far fewer features than found in real transcriptomes, and fail to properly control background levels of noise, levels of differential expression, and outliers. To address this, we use real RNA-Seq data as a model to generate test datasets, following a negative binomial distribution, with five levels of noise. We introduce 5% true positive features with fold changes between 2 and 20X, and 2% false positive outliers, while verifying that the remaining features are indeed true negatives. Using these datasets, we compare the performance of our algorithm with that of currently available algorithms. We are able to detect 99.5% of true positive features in the dataset with the lowest level of variation, and 83.5 % true positives at the highest level of variation, with less than 0.5% false positive rate. While other algorithms perform well at low levels of noise, with the best finding 99% true positives, most also call 50% of outliers as differentially expressed. Further, at higher levels of noise, they are unable to properly distinguish true and false positives, with the most accurate finding only 64% true positives, and most calling greater than 25% false positives. When compared to a variety of currently available algorithms, T2D2seq shows higher sensitivity and specificity, and is more computationally efficient.

1545W

Statistical model for the prioritization of causal genetic variants in next generation sequencing data. *J.A. Chen[1], G. Coppola[1,2].* 1) Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA; 2) Department of Psychiatry and Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA.

The majority of the 10,000+ risk-associated SNPs identified in genome-wide association studies (GWAS) have small effects, and much of the postulated genetic contribution to most diseases or traits is still unaccounted for. Most of the GWAS SNPs are hypothesized to tag common and rare variants with true functional effect. Under these assumptions, next generation sequencing (NGS) approaches can be used to identify all variants in a region with single base pair resolution. However, few of these associations have been defined and their existence is largely theoretical, as there are no rigorous statistical methods to uncover causal variants and disentangle their respective contributions to disease risk. We propose the use of structural equation modeling and Bayesian model selection procedures to prioritize likely causal variants from NGS data in GWAS-associated regions. The statistical method was used on synthetic data that was generated using the 1000 Genomes Project data, and on NGS data from neurodegenerative disease. The method identified the causal variants with high accuracy in the synthetic data, and prioritizes variants for further follow-up from the neurodegenerative disease data.

1546T

The database of Genotypes and Phenotypes: dbGaP. *M. feolo, R. Bagoutdinov, S. Dracheva, L. Hao, Y. Jin, M. Kimura, M. Lee, J. Mena, N. Popova, S. Pretel, N. Sharopova, S. Stefanov, A. Stine, A. Sturcke, K.T. Tryka, Z. Wang, M. Xu, L. Ziyabari, S.T. Sherry.* National Center for Biotechnology Information, National Institutes of health, Bethesda MD.

The National Center for Biotechnology Information's database of Genotype and Phenotype (dbGaP) is an NIH sponsored repository charged to archive, curate and distribute information produced by studies investigating the interaction of genotype and phenotype. The data submitted to dbGaP includes: individual level molecular and phenotype data; analysis results; medical images; general information about the study; and documents that contextualize phenotypic variables, such as research protocols and questionnaires. The molecular data includes array-based, sequence-based or imputed genotypes; expression, as well as next-generation sequencing (NGS) performed to produce whole exome; whole genome; RNA seq; and epigenomic data. NGS sequence submitted to dbGaP, mostly in the form of BAM files, are processed by NCBI's Short Read Archive (SRA). Phenotypic data pertaining to study participants or samples consists of any combination of cross-sectional or longitudinal demographic; clinical; laboratory; exposure; or treatment variables. Medical images such as CT; MRI; and retinal scans of the eye are also stored and distributed by dbGaP. Analysis results include summary-level statistical evaluations of the association between the phenotypic variables and the molecular data. All documents, analysis, phenotypic and molecular data are accessioned with stable, unique identifiers that support update of the same clinical study with successive genetic investigations. For example, high-throughput genotyping by one research group, followed by exome sequencing by another group. These identifiers also make it possible to cite the primary data used by published genome association studies in a very specific and stable way. The dbGaP provides unprecedented access to very large genetic and phenotypic datasets, both funded by National Institutes of Health, and other funding agencies worldwide. Public access includes summary data on specific phenotype variables which is linked to study documents; statistical overviews of the genetic information and the position of published associations on the genome. Through controlled access, approved researchers from across the globe may obtain complete statistical analyses, individual phenotype and molecular measures. This presentation will review new dbGaP features that support high-throughput submission, open access data sets, improved query services and tools to download subsets of data for selected individuals.

1547F

Explore genetic components underlying dental caries through geneset- and network-assisted approaches. *Q. Wang[1], P. Jia[1], K.T. Cuenco[2,3], Z. Zeng[4], E. Feingold[2,4], M.L. Marazita[2,3], L. Wang[5,6], Z. Zhao[1,6,7,8].* 1) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 3) Center for Craniofacial & Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 4) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 5) Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN; 6) Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN; 7) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN; 8) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN.

Dental caries is a common, chronic, and complex disease leading to a decrease in quality of life worldwide. The impact of genetic factors in dental caries has been recognized for a long time. Here we adopted two strategies attempting to dissect the genetic components underlying dental caries, namely the gene set enrichment and protein-protein interaction (PPI) network analyses. For the former strategy, we applied four complementary gene set enrichment methods to a major dental caries genome-wide association study (GWAS) dataset, which consists of 537 cases and 605 controls. After analyzing 1331 Gene Ontology (GO) terms, we identified 13 significantly associated and 17 marginally associated gene sets, including 'Sphingoid metabolic process,' 'Ubiquitin protein ligase activity,' 'Regulation of cytokine secretion,' and 'Ceramide metabolic process.' For the network-assisted strategy, we first prioritized 1214 candidate genes that were collected and curated from four major genetic or genomic approaches (association studies, linkage scans, gene expression analyses, and literature mining) according to the magnitude of evidence related to dental caries. Then we searched for dense modules in the prioritized list of candidate genes by incorporating PPI data, and discovered 23 significant modules that are enriched with genes of great interest. Three major gene clusters were observed among the 23 dense modules, including cytokine network relevant genes, matrix metalloproteinases (MMPs) family, and transforming growth factor-beta (TGFB) family. Both the GO terms and dense modules we identified encompass broad functions that potentially interact and contribute to the oral environment related to caries development. Collectively, our gene set enrichment and PPI network analyses provided complementary insights into the molecular mechanisms and polygenic interactions in dental caries.

1548W

Variant Association Tools for association analysis of large scale sequence and exome genotyping array data. *G. Wang[1], B. Peng[2], S.M. Leal[1].* 1) Baylor College of Medicine, Houston, TX; 2) The University of Texas MD Anderson Cancer Center, Houston, TX.

Currently there is great interest in detecting associations of complex human traits with rare single nucleotide variants (SNVs) using large scale sequence and exome genotyping array data. The analysis and quality control for rare variant data obtained from sequencing is quite different from the analysis of common SNVs in traditional genome wide association studies. . Data quality control for sequence data uses information from a variety of difference quality matrices, e.g. genotype quality score, read depth. Statistical association tests for the analysis of rare variants aggregate variants across a genetic region which is usually a gene. Within the gene region the variants which are analyzed is usually limited to missense, nonsense and splice site variants; therefore before the commencement of association testing the variants must be annotated To address the specialized analysis of rare variants we developed variant association tools (VAT), a pipeline that implements best practices for rare variant association studies. Major features of VAT include variant site/call level quality control, summary statistics, phenotype/genotype based sample selections, variant annotation, selection of variants for analysis and the implementation of rare variant association methods for analysis of qualitative and quantitative traits. The association testing framework implemented in VAT is regression based which readily allows for flexible construction of association models with multiple covariates, weighting (based on allele frequencies or predicted functionality), interactions terms and models for pathway analysis. VAT is capable of rapidly scanning through data using multi-processes computation, adaptive permutation and simultaneously conducting multiple association tests. Results can be view as text or graphically. Additionally a programming interface is provided to readily facilitate user implementation of novel association methods. The VAT pipeline can be applied to sequence, imputed and genotyping array, e.g. exome chip data. VAT can perform association analyses on small to large scale complex trait studies making use of the latest genotyping and sequencing technologies.

1549T

Variant calling in low-coverage whole genome sequencing of a Native American population sample. *C. Bizon[1], M. Spiegel[1], S. Chasse[2], I.R. Gizer[3], Y. Li[2], E. Malc[4], P. Mieczkowski[4], J. Sailsbery[1], X. Wang[1], C.L. Ehlers[5], K.C. Wilhelmsen[1,2].* 1) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Psychological Sciences, University of Missouri-Columbia; 4) University of North Carolina High Throughput Sequencing Facility; 5) Department of Molecular and Cellular Neuroscience, The Scripps Research Institute.

For a given sequencing budget, the Low-Coverage Whole Genome Sequencing (LCWGS) strategy allows more samples to be sequenced, and therefore potentially finds a greater number of rare variants. However, the low coverage makes variant calling with standard tools difficult, possibly canceling out any such gains in detection of rare variants. Linkage-disequilibrium (LD) aware variant callers may provide calling rate and accuracy to make LCWGS viable. We examined the performance of LCWGS in 708 whole genome sequences from a population sample of Native Americans. Most of the sequences have coverage between 3X and 12X. We called variants with the LD-aware variant caller Thunder, as well as the single-sample and multi-sample GATK Unified Genotyper. We assessed variant calling through a comparison of the sequencing results to genotypes measured in 641 of the same subjects using a fixed content exome array, which contains variants across the frequency spectrum. The median concordance for Thunder is 97.5%, compared to 85.5% and 90.4% for single-sample and multi-sample calls with the non-LD aware caller. The median improvement of LD-aware calling over the other methods is dependent on depth; for samples with less than 5X coverage, the median improvement is approximately 30%, decreasing to only 2% for samples with greater than 10X coverage. LD-aware variant callers are expected to perform less well for rare variants; for variants at which a single minor allele appears in the data, the multi-sample Unified Genotyper identifies approximately 57% of the variant sites, while Thunder identifies only 41%. Our results indicate viability of the low-coverage strategy for WGS studies can be achieved through use of LD-aware callers.

1550F

Finding the Clinical Answer in Genomic Sequence: Narrowing the Search Space for Disease-Causing Mutations. *S. Garcia[1], G. Chandratillake[1], M. Clark[1], A. Patwardhan[1], S. Chervitz[1], R. Chen[1,2], E. Ashley[1,3,4], R. Altman[1,4,5], J. West[1], R. Chen[1].* 1) Personalis, Inc., Menlo Park, CA; 2) Icahn School of Medicine, Mount Sinai, New York, NY; 3) Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 4) Department of Genetics, Stanford University, Stanford, CA; 5) Department of Bioengineering, Stanford University, Stanford, CA.

Diagnosis via genomic sequencing relies on identification of causative disease variants from the tens of thousands of variants present in an exome. Protocols typically apply hard filters to exclude variants by applying expectations concerning which variant types could be disease-causing. Sensitivity suffers when criteria are too strict, missing cases that expand the phenotypic or genotypic spectrum of known diseases. Conversely, relaxation of filtering criteria may result in overwhelming numbers of candidates, delaying or preventing diagnosis. We investigated whether a knowledge-based ranking system linking reported clinical information with curated phenotype information combined with ranked genotype expectations based on family history could facilitate the identification of causally related variants without sacrificing the ability to detect novel candidates. Samples with clinical features and known causative variants representing a broad range of conditions and variant types (including structural variants of various sizes) were sequenced, and had variants aligned and called. A newly developed database linking clinical features to genes was used to identify and rank candidate genes for each sample. Identification of all possible inheritance patterns, allowing for *de novo* events and non-penetrance, yielded genotype expectations ranked by likelihood. We compared a standard approach using variant pathogenicity predictions combined with simple genotype predictions to our novel approach. Our approach reduced the number of candidate variants requiring manual review. For example, in a trio segregating an autosomal dominant trait, 597 candidate variants were identified using the standard approach, while our approach ranked 41 candidates. Similarly, in a trio segregating an autosomal recessive trait, our approach ranked four candidates; the standard approach identified 58. In these, and all cases tested, the known causative variant was ranked first by our approach. The use of our novel, knowledge-based ranking successfully identifies the most likely causative variants in genomic data, reducing manual review time. With current estimates in the range of 20-60 minutes required for review of each variant, this approach has potential to dramatically improve turnaround time for exome/genome sequencing without sacrificing the potential for novel discovery.

1551W

Detecting disease-causing alleles with the human gene connectome.
Y. Itan[1], J-L. Casanova[1,2,3], L. Abel[1,2], S-Y. Zhang[1,2], L. Quintana-Murci[4], G. Vogt[2], D. Fried[5], P. Nitschke[3], M. Herman[1], A. Abhyankar[1]. 1) The Rockefeller University, New York, NY; 2) Necker Hospital for Sick Children, Paris, France; 3) Paris Descartes University, Paris, France; 4) Pasteur Institute, Paris, France; 5) Ben-Gurion University of the Negev, Beer-Sheva, Israel.

Infectious diseases have historically been the greatest killer of mankind. They still account for about 25 percent of all human mortality worldwide. At any time, new epidemics or pandemics can significantly increase this proportion. Only a small proportion of infected individuals develop clinical disease; the collective burden of infection is high because there are many infectious agents. It has become increasingly clear that human genetic background is a key determinant of infectious diseases. To determine the disease-causing allele(s), high-throughput genomic methods are applied and provide thousands of gene variants per patient. However, current methods for validating the true disease-causing single allele at the single patient level are inefficient and extremely time consuming. Moreover, there is no available method for automating the selection of candidate disease-causing alleles at the cohort level, posing a major bottleneck in the field in high-throughput clinical genomics. Resolving this problem will revolutionize the field. We recently developed a novel approach, the 'human gene connectome' (HGC) - a concept, method and database that describes the set of all in silico-predicted biologically plausible routes and distances between all pairs of human genes (Itan Y et al. PNAS, 2013). With the HGC, we generated a 'gene-specific connectome' for each human gene - the set of all human genes ranked by their predicted biological proximity to the core gene of interest. We demonstrated that the HGC is currently the most powerful approach for prioritizing high-throughput genetic variants in Mendelian disease studies, by effectively identifying novel herpes simplex encephalitis (HSE) morbid alleles in whole exome sequencing (WES) patients data. We hypothesize that within a cohort of patients with the same Mendelian (or nearly Mendelian) disease, the cluster that contains the true disease-causing gene for each patient is the HGC-predicted biologically smallest cluster. Investigating a cohort of 108 exome sequenced HSE patients, we developed and applied a Mendelian clustering algorithm, which approximates the biologically smallest HGC-predicted cluster that contains one allele per patient. By that we approximate a solution for an NP-complete algorithmic problem and estimate the disease causing allele for each individual in the cohort.

1552T

Incorporating phenotypic information to improve Mendelian disease-gene predictions.
A. Javed, S. Agrawal, P.C. Ng. Genome Institute of Singapore, A*STAR, 60 Biopolis Street, Genome, #02-01 Singapore 138672.

We introduce a method which combines phenotypic and genotypic information, within Bayesian framework, in a holistic prediction implicating the gene(s) involved. The patient symptomology is mapped to a list of known disorders using Human Phenotype Ontology. The significance of the match is translated into probability and assigned to the genes implicated for the disorders. The putative role of similar genes is included using random walk with restart of a probabilistic gene-gene interaction network. This network was constructed by combining multiple data sources and exhibits strong enrichment of known genetic heterogeneity; pvalue<1e-16 using HGMD and OMIM. Next the patient's sequencing data is analyzed for genotypic predictions. Co-segregation of the disease within pedigree, and allelic frequency in public data, helps prune the list to candidate rare mutations. Among coding variants, different predictors are used to estimate damaging role of nonsynonymous, splice-site, and indel mutations. The performance is quantified using HGMD as positive set and common variants in dbSNP as neutral set. For noncoding mutations positive set was bolstered by known GWAS hits and neutral set by common mutations in CGI public data. In each category a threshold can be set to achieve 80% (or more) true positives, with 3% (or less) false positives. The variant predictions are pooled within genes for coding and noncoding mutations separately. 1000 Genomes samples were used to define the null distribution of loss of functionality of each gene. The performance was evaluated in in-silico patients. 39 diseases from HGMD were selected and patient's symptom and genomes simulated; the former by randomly sampling the symptoms of the disease, and the later by adding an implicated variant in a healthy sample from 1000 Genomes. Causal gene was selected as top ranked in 79% simulations. Next the method was compared against state of the art algorithm VAAST restricting to nonsynonymous causal variants only. Our results compared favorably, the causal gene was lone top ranked in 68% (and top-5 in 73%) of simulations. VAAST, working only with the genotypic information, was able to identify the causal gene as top ranked in 67% cases; however the top rank was assigned to 14 genes (on average) in each simulation. Our method is the first comprehensive effort which incorporates disease symptoms and estimates the deleterious role of both coding and regulatory mutations in the prediction.

1553F

VHB: a web-based Visualization tool to compare the Haplotype Blocks of a study sample with the 1000 genome project data.
C.C. Lai[1], Y.T. Chen[2], W.P. Hsieh[3], C.Y. Tang[1,4]. 1) Department of Computer Science, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C; 2) Supercomputing research center, National Cheng Kung University, 1 University road, Tainan, Taiwan 701; 3) Institute of Statistics, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C; 4) Department of Computer Science and Information Engineering, Providence University, No. 200, Sec. 7, Taiwan Boulevard, Shalu Dist., Taichung City 43301 Taiwan.

The haplotype blocks constructed with variations detected from the 1000 genomes project help us to improve the association analysis. The linkage between the point mutations (SNPs, Indels) and traits can be further refined with information of haplotype blocks. It is useful to compare the haplotype blocks of a new study sample to the ones of the same ethnic group in the 1000 genome project. We developed a web tool, Visualization of Haplotype Blocks (VHB), to provide better visualization of the big data with the plots of pairwise linkage disequilibrium (LD). Researchers could upload the variation data of one or several samples in a Variant Call Format (VCF). The statistics of pairwise LD in the selected region are calculated from the variations of study samples and data the 1000 genome project which will be stored in the database of the VHB. Then all variation data of the samples will be parsed and imported into the database schema for temporary storage (30 days). Genome regions and ethnic groups of the 1000 genome project are selected in the web interface and the results can be displayed in the Generic Genome Browser locally, or exported as the general feature format (GFF) file, which is the standard output for the genome browser and one type of annotation data format of ANNOVAR software. The VHB web-based tool is available at http://safe.cs.nthu.edu.tw/VHB/. It is implemented in Perl programming language and MySQL relational database, and an Intel Core i7 CPU and 8GB memory computer is used under Linux platform.

1554W

Inferring HIV Quasispecies from Deep Sequencing Data.
S. Mangul[1], N. Wu[1], N. Mancuso[2], A. Zelikovsky[2], R. Sun[1], E. Eskin[1]. 1) University of California, Los Angeles; 2) Georgia State University.

Human immunodeficiency virus (HIV) exhibit high genomic diversity within infected hosts, which may explain resistance to existing drugs. Monitoring and quantifying the HIV population requires inferring set of closely related viral variants , referred to as a quasispecies, and estimating their relatives frequencies. Next-generation sequencing (NGS) is a promising technology for characterizing viral diversity due to its ability to generate large numbers of reads at low cost. However, there is a trade-off among throughput, read length and accuracy for different high throughput sequencing platforms. Error rates in next-generation sequencing platforms, such as Illumina, make it infeasible to use this technology for reconstructing HIV quasispecies and inferring their relative abundances. We suggest to use barcode technique during library preparation allowing to distinguish sequencing error from true variants. Using this technique we are able to obtain high accuracy sequencing reads. Here, we introduce a method for inferring HIV quasispecies from high accurate paired-end reads obtained using the barcoding technique described above. Method consists of the following key steps: (a) build consensus from reads; (b) map reads to consensus; (c) build conflict graph and assemble quasispecies; (d) infer relative abundances of quasispecies using EM algorithm. To build a consensus from paired-end Illumina reads Vicuna tool is used. Vicuna is de novo population consensus assembly software able to produce single consensus from deep viral sequencing data. We use Mosaik tool with default parameters to map reads to assembled consensus. Based on the read mapping we construct a conflict graph which is used to represent HIV quasispicies in a sample. A conflict graph is a directed acyclic graph where each vertex represents unique read. Two reads are connected by an edge if they overlap but disagree on the overlap. To solve the quasispicies assembly problem it is necessary to partition the conflict graph into a minimal number of independent sets. This problem also known as graph coloring problem which is NP-hard. To solve this problem we iteratively partition the vertices of the graph into two disjoint subsets using greedy algorithm that maximizes cut between this subsets. Preliminary experimental results on synthetic datasets show that our method is able to reconstruct HIV quasispecies from deep sequencing Illumina paired-end data.

1555T

Detection of sample-level contamination in next generation sequencing experiments. *T.E. Scheetz[1,2], A.P. DeLuca[1,2], E.M. Stone[1,2], T.A. Braun[1,2].* 1) Dept Ophthalmology, Univ Iowa, Iowa City, IA; 2) Institue for Vision Research, Univ Iowa, Iowa City, IA.

The goal of the this research project is to detect contamination of genomic DNA samples used in next-generation sequencing applications. When performing massively parallel sequencing in a clinical setting, it is critical to ensure that any variations detected result from the correct patient sample - and *only* from that sample. Known genotype fingerprints can help validate sample identity. But additional quantitative measures are required to ensure sample integrity.

We used the Genome Analysis Toolkit (GATK) to call variations. The relative number of supporting reads (supporting / cover) was calculated for each variation. The distribution of the relative number of reads supporting each variant was compared to a distribution derived from a cohort of control samples. Contamination was detected as an increase in variations with a relative number of supporting reads below 35%.

The system we have developed is a systematic approach for identifying contamination in samples used in next-generation sequencing experiments. Contamination presents as a substantial and distinctive increase in the fraction of variations for which the relative fraction of supporting reads is below 50%. We have evaluated several dozen of our own exomes, as well as thousands of publicly available exomes, for evidence of sample-level contamination. Our analysis correctly identified two positive controls (known/intentional contaminated samples) within our own dataset. In addition, we identified several exomes from publically available datasets that show all of the hallmarks of sample contamination. Together with our collaborators we will be validating samples that appear contaminated to evaluate our algorithm's specificity and sensitivity.

We have developed a simple method for identifying contaminated samples in exome sequencing experiments. Further research in this area is needed to determine the power of this method in identifying and quantifying the extent of contamination, and the amount of contamination that can be tolerated without compromising accuracy.

1556F

New Mixed Model Estimates Drug-Effect or Disease Onset in Association with Covariates. *M. Xu, Y.Y. Shugart.* Unit on Statistical Genomics, Intramural Research Program, National Mental Health, Bethesda, MD.

Many statistical models have been developed to associate human diseases with genetic or environmental factors. Previously, the disease is often described by only one measurement. Until recently, studies seeking to associate genetic and environmental factors with human disease have relied on disease phenotypes characterized by only one measurement. The task of determining how longitudinal measurements are associated with a disease's driving factor and with other covariates requires more sophisticated type of analysis. To meet these new needs, we have developed a 'mixed model' approach, that not only pinpoints the turning-point when disease becomes manifest or therapeutic drugs begin to take effect but also delineate associations over time with such potentially confounding factors as genotype and disease subtype. Our new model therefore describes covariance for repeated measurements over time and incorporates a correlation structure for a stationary time series. The phenotypic data used here include sequenced measurements taken during the period before a drug effect or disease symptom becomes manifest, and also after the first effects or symptoms appear. It should be noted that for most diseases and drugs, the precise time of onset marking the change between these two sequences is not known. Inasmuch as sequenced data covering the whole time course from naïve to full-blown effects is more readily available for drug effects than for the development of diseases, we demonstrated our new model by estimating the distribution of drug effect onset and analyzing these estimated time points' association with key covariate factors. Using a pilot data set, we conducted data-driven computational experiments and evaluative simulations to test our new framework's ability to estimate the onset of a drug's effects. The working example we will present demonstrates that our new statistical approach successfully predicted the onset of a drug's effects on baseline anxiety in a subgroup of STAR*D participants. Our results also confirm that the new approach can calculate the required depth for a targeted power with only a nominal rate of type I errors. Finally, our demonstration shows that with our new statistic, researchers including those working with publicly available data~to pinpoint the onset of a specific disease or a specific drug and to determine how onset may be affected by such potentially influential covariates such as genotypes or sub-phenotypes.

1557W

Fast and Accurate Diploid Genotype Imputation via Segmental Hidden Markov Model. *L. Zhang[1], Y.F. Pei[1], H.W. Deng[2], Y.P. Wang[2].* 1) University of Shanghai for Science and Technology, Shanghai, Shanghai, China; 2) Biostatistics and bioinformatics, Tulane university, New orleans, LA.

Fast and accurate genotype imputation is necessary for facilitating gene mapping studies, especially with the ever increasing numbers of both common and rare variants generated by high throughput sequencing experiments. However, most of the existing imputation approaches suffer from either inaccurate results or heavy computational demand. In this study, aiming to perform fast and accurate imputation analysis, we propose a novel, fast and yet accurate method to impute diploid genotypes. Specifically, we extend a hidden Markov model that is widely used to describe haplotype structures. But we model hidden states onto single reference haplotypes rather than onto pairs of haplotypes. Consequently the computational complexity is linear to size of reference haplotypes. We further develop an algorithm to speed up the calculation. Working on compact representation of segmental reference haplotypes, the algorithm always calculates an exact form of transition probabilities regardless of partition of segments. Both simulation studies and real data analyses demonstrated that our proposed method was comparable to most of the existing popular methods in terms of imputation accuracy, but was much more efficient in terms of computation. The developed algorithm can further speed up the calculation by several folds without loss of accuracy. The proposed method will be useful in large scale imputation studies with a large number of reference subjects. The implemented multi-threading software FISH is publicly available.

1558T

CeRNA interactions cooperate with genomic variability to modulate drivers of tumorigenesis. *H.S. Chiu[1,2,3], X. Yang[1,2,3], M.R. Martínez[1,2,3], P. Sumazin[1,2,3], A. Califano[1,2,3].* 1) Columbia Initiative in Systems Biology, Columbia University, 630 West 168th Street, New York, NY 10032; 2) Center for Computational Biology and Bioinformatics, Columbia University, 630 West 168th Street, New York, NY 10032; 3) Department of Biomedical Informatics, Columbia University, 630 West 168th Street, New York, NY 10032.

Recent evidence suggests that RNAs compete for binding and regulation by a finite pool of microRNAs (miRNAs), thus regulating each other through a competitive endogenous RNA (ceRNA) mechanism. Using a kinetic model, we show that pathophysiologically relevant regulation occurs when ceRNA interactions are mediated by multiple miRNAs, while interactions mediated by a single miRNA may have negligible magnitude. Furthermore, our model predicts that ceRNA interactions mediated by multiple miRNAs are largely independent of the individual miRNA's expression and thus highly conserved across distinct cellular states. Indeed, we predict and validate an ultra-conserved network that includes more than 160,000 ceRNA interactions, which are conserved across tumor and non-tumor related cellular contexts. We show that this network integrates genetic and epigenetic alterations of cognate ceRNA regulators to dysregulate established oncogenes and tumor suppressors, accounting for a large fraction of the missing genomic variability in tumors.

1559F

MITOCHONDRIAL DISEASE SEQUENCE DATA RESOURCE (MSeqDR) CONSORTIUM: A Global Grass-Roots Effort to Compile, Organize, Annotate, and Analyze Whole Exome and/or Genome Datasets from Individuals with Suspected Mitochondrial Disease. *X. Gai[1], D. Krotoski[2], M. Gonzalez[3], D.C. Wallace[4], M. Parisi[2], S. Zuchner[3], M.J. Falk[5].* 1) Center for Biomedical Informatics, Loyola University Stritch School of Medicine, Maywood, IL 60153; 2) NICHD, NIH, Bethesda, MD 20892; 3) Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136; 4) Center for Mitochondrial Medicine, The Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104; 5) Division of Human Genetics, The Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104.

The success of whole exome sequencing (WES) for highly heterogeneous disorders, such as mitochondrial disease, is limited by the substantial technical and bioinformatic challenges of identifying and prioritizing the extensive sequence variants present in each patient. Success rates can be greatly improved if a large cohort of patient exome data from both clinical and research laboratories are assembled to permit systemiatic sequence variant analysis, annotatation, and informed interpretation. Since June 2012, we have established a grass-roots effort facilitated by the United Mitochondrial Disease Foundation (UMDF) and NIH called the 'Mitochondrial Disease Sequence Data Resource (MSeqDR) Consortium'. We have engaged more than 100 mitochondrial disease experts to identify and prioritize specific WES data analysis needs of the global mitochondrial disease community. Currently, we are moving forward toward our common goal of establishing a central data resource for the coherent compilation, organization, annotation, and analysis of WES data in dual genomes from all patients with suspected mitochondrial disease. Exome and phenotype data is united in a publicly accessible, secure, and user-friendly web-based tool that is accessible from a central website. MSeqDR prototype development is now underway with support from the UMDF and North American Mitochondrial Disease Consortium involving investigators from 9 countries in North America, Europe, Australia, and Asia. A global MSeqDR will fill the existing void in bioinformatics tools and knowledge necessary for efficient WES data analysis in the mitochondrial disease community. This resource will potentially support reliable genetic diagnosis in individual cases, facilitate identification of additional cases of rare genetic disorders, provide evidence to exclude from consideration sequence variants of unknown significance that are unlikely to be disease-causing, permit identification of modifier genetic factors underlying disease variability, and even identify genetic variants that are likely to alter response to emerging therapies.

1560W

Detection and characterization of mutations due to DNA damage and PCR amplification errors in targeted ultra-deep next-generation sequencing data. *J. Eboreime, N. Arnheim, P. Calabrese.* Molecular and Computational Biology, University of Southern California, Los Angeles, CA.

Next-generation sequencing (NGS) has been used to estimate the genome average mutation frequency. The coverage of such NGS studies is insufficient to provide a mutation frequency estimate at individual nucleotide sites so variation among sites is unknown. Deep NGS sequencing has the capability to accurately estimate mutation frequencies at hundreds of different nucleotides simultaneously in hundreds of thousands of copies of a small relevant gene DNA segment. This could allow for the detection of germline mutation hot spots. One problem is that NGS deep sequencing suffers an error rate higher than the expected mutation rate. To improve NGS fidelity we adapted a new protocol called the Safe-Sequencing System (SSS), developed in the Vogelstein laboratory, and applied it to a 150 bp region in FGFR3 that includes the site of a common de novo disease mutation transmitted in an autosomal dominant fashion. SSS utilizes a unique identifier (UID) to distinctly tag each original genomic DNA molecule. During testis library preparation, PCR of each distinctly tagged template molecule creates a UID family. NGS reads with the same UID and disease mutation are candidates for being true mutants. We divided the testis of an unaffected 68 year old man into 192 pieces, and measured the somatic mutation frequency in 32 of the pieces. Using SSS we detected an unexpectedly high average mutation frequency (4.6×10^{-5}). We observed that among the sample of 32 pieces there were cases where some pieces had significantly higher mutation rates at certain nucleotides compared to other sites. We also observed many C>A/G>T transversions consistent with oxidation of guanine to 8-oxo-guanine and deamination of cytosine to uracil (C>T/G>A) but other mutations were also unexpectedly frequent. Not correcting for these high frequencies, presumably due to DNA damage and other artifacts, can greatly affect the accurate detection of rare in vivo mutants. The Duplex Sequencing method, developed in the Loeb laboratory, improves the SSS method by not only uniquely tagging each molecule with a UID but also tagging each strand of the molecule with the same sequence (or its complement). Comparing both original strands of a DNA duplex can exclude premutagenic lesions that arose before library preparation. In addition, our data allows us to estimate, and correct for, errors that arise during the PCR amplifications required for NGS library formation.

1561T

Comparison of the performance of read mappers and assembly methods on indel calling. *H. Lin, Y. Shen.* Columbia University, New York, NY.

The primary technique for detecting genomic variations from next generation sequencing data is to map reads to a reference genome and call potential variants based on the aligned reads. Although this approach has worked well for detecting SNP's, the problem of detecting insertions and deletions from short read data remains challenging, especially if the insertions and deletions are large. A number of new techniques have been developed for mapping reads more effectively in the presence of insertions and deletions. In this work, we compare the performance of a number of new read mappers and their ability to detect indels (i.e. insertions and deletions) from real sequencing data. Our data consists of sequencing from trios consisting of a proband and his or her parents. This trio data provides a unique opportunity to evaluate the quality of the indels called from different read mappers, in addition to measuring the overall number of indels called. In general, measuring the accuracy of indels called from real sequencing data can be a challenge, but with sequencing from trios, we can measure the mendelian concordance and the transmission ratio of the indels called, which provides a reasonable measure of precision. In addition, for each read mapper, we also measure the number of novel indels called, which are not found in existing databases, such as dbSNP, and also measure their mendelian concordance and transmission ratios. Lastly, we also explore the performance of some newly developed assembly based approaches for indel calling, and the impact of error correction on the methods.

1562F

A Bayesian framework for de novo mutations calling in trios. *Q. Wei[1], Y. Liu[2], Y. Han[2], X. Zhan[3], W. Chen[4], B. Li[1].* 1) Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 2) Center for Human Genetic Variation, Duke University, Durham, NC; 3) Department of Biostatistics, University of Michigan, MI; 4) Department of Pediatric, University of Pittsburgh, PA.

For most human complex diseases, although numerous loci have been identified, the genetic basis remains largely unknown. For sporadic cases without family history, increasing evidence shows that de novo mutations play an important role in the genetics of thesediseases such as neurodevelopment disorders. Identifying de novo mutations in such cases will not only help directly find genes implicated in disease etiology but also provided candidate genes or pathways for mappinginherited variants associated with the disease. High-throughput next-generation sequencing enables a genome- or exome-wide detection of de novo mutations by sequencing proband-parents trios.Since de novo mutations are extremely rare, e.g. 50-100 per generation, it is challenging to sifttrue mutations throughsequencing error and alignment artifacts. Traditional approaches infer individual genotypes and identify putative de novo mutations by comparing proband's and parental genotypes. Recently an efficient likelihood-based framework (polymutt) was proposed and shown to outperform standard approaches dramatically. A limitation of polymutt is that the pre-specified prior mutation rate has a significant impact on the de novomutation calling,resulting inreduction of sensitivity and specificity when improper priors are assigned. To address this problem, we developedtriodenovo, a Bayesian framework that separatesde novo calling from the mutation rate for trio data so that the prior mutation rate can be adjust post-hoc.Through extensively simulations we showed that this new method has higher sensitivity and specificity than polymutt and standard approaches, especially when the depth of coverage is low (e.g. 20X). Coupled with machine learning approaches to filtering false positive candidates due to alignment artifacts, we showed that triodenovoachieves increased specificity without scarifying sensitivity on real data.We hope that this new framework is useful to the research community toefficiently identify de novo mutations to facilitate the association mapping of genes for human diseases.

1563W

PhenoExplorer: a tool to help researchers identify relevant studies and phenotypic variables in dbGaP. *J. Ambite[1], L. Lange[2], S. Sharma[1], S. Voinea[1], C. Hsu[1], Y. Arens[1].* 1) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 2) Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC.

The National Institutes of Health (NIH) created the dbGaP resource as a repository for genetic and phenotypic data. dbGaP provides the scientific community an unprecedented opportunity to gain access to data from multiple studies, but identifying the pertinent subset of studies and phenotype variables in dbGaP is often challenging. The true utility of this resource is thus limited by a user's ability to identify the specific data relevant to their own research. Some studies include thousands of phenotype variables, many being study-derived variables measuring similar traits, with descriptions varying widely across studies. For example, hypertension status is described variously as 'HTN', 'high blood pressure', and 'high BP', causing many exact text searches to fail to identify all relevant variables. Moreover, a search for hormone replacement therapy should also return studies with variables about estrogen intake, which requires a semantic search. In addition, users often wish to limit phenotype searches to studies with certain features, such as those with African Americans and/or with sequence data. To address these issues, we developed PhenoExplorer.org, a free online tool that allows researchers to identify dbGaP studies containing phenotypes of interest and the corresponding phenotypic variables. Specifically, a researcher can search for studies along a set of dimensions, including race/ethnicity, sex, study design, type of genetic data, genotype platform, and diseases studied. Crucially, the researcher can also enter free text describing the phenotype of interest. The tool returns all studies satisfying the metadata constraints and containing phenotypic variables relevant to the user-described phenotype. Our core novel contribution is the semantic similarity metric used to identify phenotypic variables related to the user's search string. This metric uses an ensemble of similarity predictors based on information retrieval and machine learning techniques. One predictor is a semantic text similarity function automatically learned from manual phenotype harmonization done for the NHLBI CARe consortium. Since variable descriptions in dbGaP are typically short (1-10 words), the tool expands them with definitions and synonyms from the Unified Medical Language System. Other predictors use variations of information retrieval text similarity techniques (TF-IDF, cosine similarity). In summary, PhenoExplorer enhances access to the wealth of data at dbGaP.

1564T

How to reduce false positives: A quality assurance pipeline for phenotype data. *A. Matchan[1], NW. Rayner[2], AE. Farmaki[3], K. Hatzikotoulas[1], E. Tsafantakis[4], M. Karaleftheri[5], L. Southam[1,2], K. Panoutsopoulou[1], G. Dedoussis[3], E. Zeggini[1].* 1) Applied Statistical Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 3) Harokopio University Athens, Athens, Greece; 4) Anogia Medical Centre, Anogia, Crete, Greece; 5) Echinos Health Centre, Echinos, Greece.

High quality phenotype data are an essential component of studies examining the association between rare variants and complex traits. Population isolates are being increasingly used to empower these studies due to the reduced amount of genetic and environmental variability within these groups. We are studying two isolated populations in Greece (HELlenic Isolated Cohorts study, www.helic.org). The main aims of the project are to link phenotype to genotype for traits of medical relevance and to enable population genetics studies. An extended set of phenotype data (~350 traits) has been collected for all samples (n~3000). This includes anthropometric, cardiometabolic, biochemical, haematological and diet-related traits. Collection of these data relies, primarily, on interview-based questionnaire methods followed by manual double blind entry of the data. This process can introduce errors, including keying errors, misinterpretation of questionnaire answers, language translation problems, duplicate data collection and the issue of how missing data (as opposed to unanswered questions) should be interpreted in downstream analysis to avoid misleading results. To address these issues we have developed a phenotype cleaning pipeline to remove errors and inconsistencies in the data as much as possible. The double-blind data entry procedure enables automated checking for discrepancies and captures the majority of user input errors. Outliers in quantitative traits are programmatically identified and data values are assessed against normal ranges as published in the literature. Cross-trait checking is additionally employed to assess the integrity of the data e.g. checking height and BMI versus gender. Genotype quality control information is also used to identify duplicates (4.4% of samples in the HELIC cohort) and assess the robustness of the genotype-phenotype link based on gender. We find that ~10% of entries in the HELIC dataset contained an error. The clean phenotype data are centrally stored and changes are tracked using a version control system. As a result of this pipeline we have a broad range of high quality phenotype data available for 99.8% of the HELIC samples, enabling powerful genotype-phenotype association studies. The effect of phenotype errors on false positive and false negative associations can be substantial. We have put in place a data cleaning pipeline which minimises error and therefore empowers locus discovery in downstream analyses.

1565F

Reconstructing Pedigrees from Estimates of Genomic Sharing in Admixed Populations. *J.E. Below[1], J. Staples[2], A. Reiner[3], L. Ekunwe[4], E.L. Akylbekova[5], S.K. Musani[5], J.G. Wilson[5], C.R. Hanis[1], D. Nickerson[2].* 1) Human Genetics, University of Texas Health Science Center, Houston, TX; 2) University of Washington, Genome Sciences, Seattle, WA; 3) University of Washington, School of Public Health, Seattle, WA; 4) Jackson State University, Jackson, MS; 5) University of Mississippi Medical Center, Jackson, MS.

Understanding and correctly utilizing relatedness among samples is essential for all genetic analysis. However, records of sample relatedness are often incorrect, incomplete, or unavailable. PRIMUS is an algorithm that utilizes genome-wide estimates of identity by descent (IBD) to assign relationship categories and leverages these pairwise relationships to identify all possible pedigrees consistent with the observed genetic sharing. Reconstructing pedigrees in admixed populations is complicated by the fact that many algorithms for estimating IBD from genome-wide SNP data assume sampling from a single homogeneous ancestral population. The presence of ancestry informative markers (AIMs) in estimating IBD using a method of moments can strongly bias relationship classification, resulting in inaccurate or failed pedigree reconstruction. We have implemented a principal component analysis within PRIMUS to identify signatures of admixture, appropriate reference population minor allele frequencies, and identify and remove AIMs prior to IBD estimation. Controlling for multiple ancestral groups, we reconstructed pedigrees for all 1,985 Mexican Americans from the Starr County Health Study (SC) as well as 3030 African Americans from the Jackson Heart Study (JHS). PRIMUS unambiguously identified 197 previously undescribed pedigrees in SC and reconstructed pedigrees for 338 families within JHS. Our method provides accurate reconstruction in genetically heterogeneous samples. We present data for the resulting pedigrees, and show that PRIMUS is powerful for both identifying novel pedigrees in large admixed genetic cohorts and for validating known pedigrees.

1566W

Detecting Differentially Expressed Genes in RNA-Seq Data with Unknown Conditions. *G. Klambauer, T. Unterthiner, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Upper Austria, Austria.

Methods that identify differential expression in RNA-Seq data are currently limited to study designs in which two or more sample conditions are known a priori. However, these biological conditions like activated regulatory and metabolic pathways are typically unknown in genetic studies such as the HapMap or the 1000 Genomes project. We suggest DEXUS for detecting differential expression in RNA-Seq data for which the sample conditions are unknown. In a Bayesian framework DEXUS models read counts as a finite mixture of negative binomial distributions in which each mixture component corresponds to a condition. Evidence of differential expression is measured by the informative/non-informative (I/NI) value, which allows differentially expressed transcripts to be extracted at a desired specificity (significance level) or sensitivity (power). DEXUS performed excellently in identifying differentially expressed transcripts in data with unknown conditions. On 2,400 simulated data sets, I/NI value thresholds of 0.025, 0.05, and 0.1 yielded average specificities of 92%, 97%, and 99% at sensitivities of 76%, 61%, and 38% respectively. In cohorts with genetic and RNA-Seq data, DEXUS was able to detect differentially expressed transcripts that could be related to genetic variants via the identified conditions. These genetic variants can be classified into structural variants like copy number variations and single nucleotide variants, that is, eQTLs.

1567T

Development of quality control processes for next-generation sequencing technologies. *P. Ebert, J. Calley, R. Higgs, X. Ma, T. Barber.* Eli Lilly and Company, Indianapolis, IN.

The use of next-generation sequencing (NGS) as a tool for interrogation of somatic/germline variation in DNA and differential expression in RNA pervades many areas of biology. With the adoption of this technology has come the ability to detect many types of experimental variation that were essentially invisible with previous technologies. The sampling process for RNASeq is very different than that for DNASeq, which leads to different challenges. We have developed rapid and sensitive methods to detect and quantitate the sources of variation which occur during the 'wet-lab' stage of NGS. These include methods to detect and remove contaminating sequences, detect 3' bias often observed in RNA sequencing data, and resolve sample identity issues. These processes are providing critical information about the quality of samples generated from internal projects (as well as data acquired from external sources) which is providing significant value to our research efforts.

1568F

MAPRSeq - Mayo Analysis Pipeline for RNA Seq: A comprehensive workflow for RNA-Sequencing data analysis. *A. NAIR, K. KALARI, J. BHAVASAR, X. TANG, J. DAVILA, J. NIE, D. O'BRIEN, JP. KOCHER.* Mayo Clinic, Rochester, MN.

RNA-Sequencing (RNA-Seq) technology is information-rich; the breadth of information gained spans from large structural changes to single nucleotide variants (SNVs). By efficiently analyzing RNA-Seq data, we can query and obtain a variety of genomic features, such as gene expression, novel and fusion transcripts, alternative splice sites, long non-coding and circular RNAs, SNVs etc. Majority of the RNA-Seq bioinformatics tools output one or two genomic features for downstream analysis. There is no extensive workflow that can be used to obtain a number of features from RNA-Seq data. Therefore, at Mayo Clinic, we have developed MAPRSeq - a computational workflow that takes the utmost advantage from an RNA-Seq experiment and provides comprehensive reports on genomic features for secondary data analysis. MAPRSeq workflow integrates a suite of open source bioinformatics tools along with in-house developed methods to analyze paired-end RNA-Seq data. Read alignment is performed with Tophat which uses Bowtie - a fast, memory efficient, short sequence aligner. Tophat aligns reads to the transcriptome and further to the genome to report both existing and novel junctions. Along with the alignment (BAM) and junction (BED) files, Tophat also provides a list of expressed fusion transcripts using the TopHat-Fusion algorithm. The BAM file is processed using HTSeq to summarize expression at gene level. Exon quantification is obtained with in-house methods that leverage BEDTools. In addition to raw gene and exon expression counts, MAPRSeq also provides normalized values (RPKM). For accurate variant detection, GATK is used to call SNVs that are further annotated with quality score, coverage and additional criteria using VQSR. MAPRSeq workflow reports several analytical functions, including alignment statistics, in-depth quality control metrics, gene and exon expression levels, fusion transcripts and SNVs for each sample. Circos plots are also provided to visualize fusion transcripts. MAPRSeq incorporates Integrated Genomics Viewer (IGV) to visualize alignment and coverage along the transcriptome as well as exon-exon junctions. MAPRSeq incorporates UCSC tracks on transcription and regulation in IGV as well to facilitate user interpretation. MAPRSeq is available at https://code.google.com/p/maprseq/.The workflow is well optimized to run on a single Linux machine as well as in a cluster environment to fully leverage multiple processors.

1569W

Assessment of the Impact of Read Length on RNA-seq Results: An ABRF Consortium Study. *J.A. Rosenfeld[1,2], G. Rudy[3], S. Chhangawala[4,5], P. Wu[6], S. Tighe[7], M.D Wang[6], D.A. Baldwin[8], G. Grills[9], C.E. Mason[4,5], The ABRF-NGS Consortium.* 1) Medicine, Rutgers - New Jersey Medical School, Newark, NJ; 2) American Museum of Natural History, New York, NY; 3) Golden Helix, Bozeman, MT 59718; 4) The Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY; 5) Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY; 6) The Wallace H. Coulter Dept. of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332; 7) Vermont Cancer Center, University of Vermont, Burlington, VT 05405; 8) Pathonomics LLC, Philadelphia, PA 19104; 9) Institute of Biotechnology, Cornell University, Ithaca, NY 14853.

RNA sequencing is a rich assay for delineating the information of the transcriptome. However, few RNA-seq standard data sets exist to help evaluate methods for the quantification of gene or splice form expression. Moreover, each next-generation sequencing (NGS) platform has unique aspects of library synthesis, sequencing, alignment, and data processing. Little is known about the cross-site reproducibility, technical variance and interoperability of NGS platforms for RNA-seq. The goals of the Association of Biomolecular Resource Facilities - Next Generation Sequencing (ABRF-NGS) Study are to evaluate the performance of NGS platforms and to identify optimal methods and best practices. As part of this evaluation, we analyzed the utility of short vs. long reads, and single vs. paired reads from each end of a template, for identifying differentially expressed genes and novel isoforms. Read length and pairing are critical parameters affecting experimental design, analysis and cost, and longest available reads are often used in paired-end mode by default, whether or not these most expensive choices are necessary for the accurate RNA-seq profiling. To test this question, we used two standard ABRF RNA-seq libraries that were used throughout the ABRF-NGS project and the preceding MAQC project. To confirm our results on an independent dataset, we utilized data from ENCODE on 2 cell lines. In order to mitigate the influence of a particular computational algorithm, we utilized 3 separate pipelines to analyze the data. In order to further control the data, we used paired-end 100bp reads and computationally produced synthetic shorter and single-end reads from these real reads by trimming their length. The results were extremely consistent, but not in a positive way. We found that read-length and paired or single-end status are important factors in determining the results of an RNA-seq experiment. We found that the number of differentially expressed (DE) genes varied greatly as the read length changed. For many of the genes, there was a slight shift in FPKM, which was sufficient to move a hit from being DE to being insignificant This study indicates that while RNA-Seq is a powerful method which can produce useful results, the current technique yields a substantial variability in results that cannot be dismissed. Critical improvements are needed in RNA-seq sample preparation, data generation, and bioinformatics analysis tools.

1570T

Enlight: a web-based tool for integrating GWAS results with biological annotations. *Y. Guo[1,2], D. Conti[1,2], K. Wang[1,2,3].* 1) Preventive Medicine, University of Southern California, Los Angeles, CA; 2) Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA; 3) Psychiatry & the Behavioral Sciences, University of Southern California, Los Angeles, CA.

Genome-wide association studies (GWAS) have revealed thousands of variants associated with diverse human diseases and traits, butmost of the top hits are believed to beproxy markers for the true causal variants. Identifying causal variants remains a key challenge in post-GWAS era. Many identified regions fall into non-coding regions, making it especially hard to associate statistical significance with predicted functionality. Results from the ENCODE project, Epigenome Roadmap project, and other large-scale projectsmay help provide biological insights and generate hypotheses, but individual researchers do not have strong informatics skills or familiarity with these databases toleverage the massive amounts of annotation data in their own studies. To overcome this problem, we created a web-based tool, Enlight, for investigating and visualizing the relationships between all variants within a GWAS loci and available annotation information. It is ideal for non-coding variants, and is equally applicable for coding variants.Its 'GWAS plotting' function was modified from the LocusZoom software, which outputs publication-ready graphs with GWAS-specific information such as linkage disequilibrium, effect sizes and/or p-values, for genotyped and imputed variants. Additionally, its 'annotation plotting' function overlays functional annotation information to GWAS variants and adjacent variants (such as those from 1000 Genomes Project and NHLBI-ESP project), to help identify functional elements such as enhancers, promoters, methylation patterns, HiC-Seq region, ChIP-Seq peaks, RNA-Seq peaks and other relevant epigenetic information. Users have the flexibility to select a specific tissue or cell line of their own interest, and then load the tissue-specific annotation information into the graph. Users can choose from a number of pre-made trackscompiled by the UCSC genome browser, but can also upload custom data tracks in BED format. Enlight is freely available at http://enlight.usc.edu.

1571F

RNA-Seq Analysis of Alternative Splicing Events in Drosophila melanogaster. *Y. Li[1,2], X. Rao[1], C. Amos[2], B. Liu[1].* 1) The Center for Genetics and Genomics, University of Texas MD Anderson Cancer Center; 2) Center for Genomic Medicine, Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College.

Alternative splicing is an important biological process in the generation of derivative functional transcripts with same genomic sequences. RNA-seq technology has been widely used for researches in transcriptomics. The profile of gene differential expression and alternative splicing can be analyzed by the various statistical analytical methods. In this article, we are reporting a protocol for differential analysis of splicing junctions and intron retentions. The DEXSeq, an R bioconductor originally designed for gene exons expression analysis (Anders S, Reyes A, Huber W, Genome Res, 2012 Oct; 22(10):2008-2017), was adopted on splicing junctions and intron retentions analysis. The design of reference files for splicing junctions and intron retentions was introduced to fulfill the DEXSeq analysis. To evaluate the protocol, the public RNA-seq datasets generated from the Drosophila melanogaster S2-DRSC cells with RNAi depletion of RNA binding proteins (Brooks AN et al., Genome Res 2011 Feb; 21(2):193-202) have been analyzed. With the further study on the differential analysis of splicing junctions and intron retentions, sequence motifs were identified at the flanking sequence of significant splicing junctions and intron retentions. Ingenuity Pathway Analysis (IPA) of the vertebrate homologs of the genes with significant splicing events may provide functional information. The results from this study show that DEXSeq package can be applied on alternative splicing analysis on RNA-seq data; conserved DNA sequence motifs may imply the possible roles of the RNA binding proteins during the splicing events.

1572W

Modeling complex autoimmune disease susceptibility in the skin with regulatory and genetic-genomic interface networks. *J.C. Chen[1,2], A.M. Christiano[2,3].* 1) Joint Centers for Systems Biology, Columbia University, New York, NY; 2) Department of Genetics and Development, Columbia University, New York, NY; 3) Department of Dermatology, Columbia University, New York, NY.

Autoimmune diseases that present in the skin and hair, such as Alopecia areata and psoriasis, are complex genetic diseases whose pathogenesis resides at the interface between two distinct organs: the immune system and an affected 'end organ'. While research has shown that certain genetic variants affecting immune response genes can lead to disease in patients and that these variants can be are shared across different autoimmune diseases, the symptomatic organs vary from patient to patient. This indicates that there can exist genetic factors in the end organs that render them susceptible to aberrant autoimmune response. Since multiple autoimmune diseases can present in the skin, and can sometimes present concomitantly in a single patient, the skin serves as an ideal model to study the molecular regulatory processes that contribute to the onset of autoimmune disease in a specific organ. Historically, human diseases have been modeled by large, complex gene expression panels or physiological traits. While informative, this approach does not provide direct insight into the regulation and induction of these signatures. In order to understand the molecular regulation of autoimmune disease susceptibility in end organs such as the skin, we have created a molecular regulatory model for the skin based on reverse-engineered transcriptional and genetic-genomic interface networks. We used the systems biology algorithms ARACNe and DIGGIn to generate integrative networks of gene regulation in human skin using gene expression and genomic microarrays. We have implemented an integrative approach that allows us to leverage both genetic and genomic data to model autoimmune disease sensitivity in the skin as a set of key molecular regulators inferred from emergent networks and the mutations that affect these key regulators. Importantly, our method provides a way to model autoimmune disease as a function of its molecular regulators, rather than its biomarker panels or physiological traits. This perspective allows us to parse causal genes from their effects, and how they directly link to autoimmune disease susceptibility. This systems biology perspective invites both novel methodologies in studying complex, polygenic gene expression patterns, and the development of novel therapeutic strategies informed by regulatory networks.

1573T

Network Communicability: An Effective Alternative Metric for Genome Analysis. *C. Shaw, I.M. Campbell.* Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX.

Network similarity metrics are increasingly utilized to productively analyze genome-wide data. Conventional approaches such as shortest path and clique-based techniques have been useful but are not well suited to all applications. Recently, computational scientists in other disciplines have developed network communicability as a complementary method to analyze network relationships. Network communicability considers all paths between two network members. Given the success of previous analyses of the human protein-protein interaction network, we sought to apply network communicability to this complex data. As a challenging test case, we sought to partition human protein products into disease classes using communicability together with a set of training genes with known disease annotations based on OMIM. Our analysis reveals that communicability has several key advantages over shortest path and other previous methods. Additionally, we used copy number variation data from our diagnostic lab to show that the disease classes of genes deleted in patients correlate with their respective phenotypes. Our data suggest that metrics based on network communicability have considerable utility in the analysis of large-scale biological networks and may be fruitfully applied to a number of computational problems.

1574F

Quantifying gene expression and allele specific expression simultaneously using personal human genomes. *N. Raghupathy, K. Choi, S.C. Munger, G.A. Churchill.* The Jackson Laboratory, Bar Harbor, ME.

Large-scale genome and transcriptome sequencing efforts have characterized millions of common genetic variants across human populations. However development of tools that can effectively utilize this individual-specific variation to inform quantitation of gene abundance and allele-specific expression (ASE) have lagged behind. Current approaches employ two steps to quantify gene expression abundance and ASE from RNA-seq data; gene-level abundance is estimated from alignment of all reads genome-wide, while ASE is assessed separately by analyzing only reads that overlap known SNP locations. To address limitations of current RNA-seq alignment and quantitation methods, we developed the complimentary tools Seqnature and EMASE. Seqnature incorporates known polymorphisms and short indels from genetically diverse and heterozygous model organisms into reference genomes, and can construct individualized haploid or diploid transcriptomes suitable for read alignment by common aligners. EMASE is a model-based quantitation approach that employs an Expectation Maximization (EM) algorithm to apportion multi-reads at the level of the gene, isoform, and allele. Here we extend Seqnature to construct personalized diploid human genomes and transcriptomes from phased genetic variation data, and apply EMASE to estimate allele-specific and gene expression abundance simultaneously in human samples. We show that accounting for individual genetic variation by modeling as diploids at the alignment step and using an EM approach for abundance estimation improves RNA-seq alignment, gene expression and allele specific expression estimates.

1575W

A Comprehensive Resequence-Analysis of 250kb Region of 8q24.21 in Men of African Ancestry. *C.C. Chung[1,3], A.W. Hsing[2], E. Yeboah[4], R. Biritwum[5], Y. Tettey[6], A. Adjei[6], M.B. Cook[1], A. De Marzo[7], G. Netto[7], J.F. Boland[1,3], M. Yeager[1,3], S.J. Chanock[1,3].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD; 2) Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD; 3) Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, Advanced Technology Program, SAIC-Frederick Inc., NCI-Frederick, MD; 4) Department of Urology, University of Ghana, Ghana; 5) Department of Epidemiology, University of Ghana, Ghana; 6) Departmnet of Pathology, University of Ghana, Ghana; 7) Department of Pathology, Johns Hopkins University, Baltimore, MD.

Genome-wide association studies (GWAS) have identified that a ~1M region centromeric to the *MYC* oncogene on chromosome 8q24.21 harbors at least 5 independent loci associated with prostate cancer risk as well as additional loci associated with cancers of the breast, colon, bladder, and chronic lymphocytic leukemia (CLL). Because GWAS identify genetic markers that may be indirectly associated with disease, fine-mapping based on sequence analysis in distinct populations provides important insights into the differences in patterns of linkage disequilibrium (LD) and is critical in defining the optimal variants to nominate for biological follow-up. In order to catalog the genetic variation in individuals of African ancestry, we resequenced a region (250kb; chr8:128,050,768-128,300,801, hg19) containing several prostate cancer susceptibility loci as well as a locus that is associated with CLL. Our sample set included a total of 125 samples, including 78 samples from a population-based study in Ghana and 47 samples of African-Americans from Johns Hopkins University. We have identified 1,838 SNPs of which 285 were novel. Using genotypes derived from sequencing, we refined the LD and recombination hotspots within the region and determined a set of tag SNPs to be used in individuals of African ancestry in future fine-mapping studies. Based on LD, we annotated putative risk loci and their surrogates using ENCODE data, which could be important in establishing priorities for future functional work designed to explain the biological basis of associations between SNPs and both prostate cancer and chronic lymphocytic leukemia.

1576T

Integrating multiple reference sequences, known variation and de novo assembly for personal genome inference. *A.T. Dilthey[1], Z. Iqbal[1], C. Cox[2], M.R. Nelson[3], G. McVean[1].* 1) The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Quantitative Sciences, GlaxoSmithKline, Stevenage, United Kingdom; 3) Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, North Carolina, United States of America.

The genome of a species typically encompasses great diversity, reflecting the mutation processes and selective forces experienced within its evolutionary history. A single sample may be used as a canonical exemplar, leading to a reference genome as the basis for analysis. This is a good first approximation and is the basis of sequence assembly for most high throughput sequencing experiments. However, this approach typically fails in regions where the sample is highly diverged from the reference and does not take advantage of known variants. We have developed an approach that combines multiple reference sequences, previously known variations and de novo assembly for inferring genome sequence. The first step is to infer a diploid reference using a Hidden Markov Model on an ordered graph representing a multiple sequence alignment of known references. These may be entire genomes, long haplotypes, divergent exons or short variants and are augmented with sequence from the sample under study. A diploid genome corresponds to a paired path through the graph. The second step is to map reads to the pair of personal references and infer variants by standard approaches. By doing so, we improve sensitivity to novel variant detection, particularly in regions that diverge from the standard single reference. We demonstrate the value of the personal reference approach in the analysis of the human Major Histocompatibility Complex , particularly in its ability to characterise structural variation and regions of extreme polymorphism frequently found in this region.

1577F

Genotyping of Exotic Structural Variants Using BWA and Pindel in Whole Genome and Exome Sequence Data. *D.S. Hanna[1], J.D. Smith[1], D.A. Nickerson[1], J.M. Swanson[2], E.M. Faust[3].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Child Development Center, University of California, Irvine, CA; 3) Department of Environmental & Occupational Health Sciences, University of Washington, Seattle, WA.

Insertions, deletions and other structural variants, such as inversions and segmental duplications, leave specific signature patterns within the sequence alignments that usually go undetected, or are often genotyped incorrectly as clumps of single nucleotide variants. However, these signatures are often discernible and unique, and can, therefore, be detected using sophisticated algorithms. These variants are not well assessed because of the difficulty in their detection, and have already been shown to be associated with disease and contribute to unique phenotypes. Using a combination of the latest BWA-MEM aligner, GATK recalibration tools and Pindel, we show improvements in the identification of novel variants in the population by pinpointing the exact breakpoints of structural variants, and thus finding the length of the variation. The genotype of each sample can then be ascertained by examining the sequence allele balance at the breakpoints. Using a population consisting of 50 trios of mixed ethnicities obtained for the National Children's Study,, we have identified and traced exotic variation in the families and, when possible, we validated our results with genotyping arrays. By robustly identifying these variants, we were able to increase our power to discover putatively causal variation in family-based studies and to identify new associations in population studies. An added benefit is to gain insight into how to increase true positives while decreasing false negatives, and how variables such as sequence quality, allelic depth and sequence read length affect these metrics.

1578W

HIGH-COVERAGE WHOLE GENOME SEQUENCING OF THIRTY BRAZILIAN ADMIXED INDIVIDUALS. *M. Machado[1], R. Moreira[1], E. Tarazona-Santos[1], A.C. Pereira[5], M.L. Barreto[4], B.L. Horta[3], M.F. Lima-Costa[2], A. Horimoto[5], N. Esteban[5], F. Kehdy[1], M. Rodrigues[1], W.C.S. Magalhaes[1], Brazilian Epigen Consortium.* 1) Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 2) Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil; 3) Universidade Federal de Pelotas, Brasil; 4) Instituto de Saúde Coletiva, Federal University of Bahia, Brazil; 5) Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of São Paulo.

The Brazilian Epigen Initiative is a genetic epidemiology project which main aims are to infer population structure and genomic ancestrality of the three largest Brazilian cohorts as well as investigate the genome-wide association among a set of different clinical outcomes. In order to reach these objectives, the project focused on three different approaches: Genotyping of 4.3 M SNPs (HumanOmni5) for 270 samples; Genotyping of 2.3 M SNPs (HumanOmni2.5) for 6496 samples and Whole genome sequencing for 30 Brazilian admixed samples (Pelotas from Southern Brazil (n=10), Bambuí from the South East (n = 10) and Salvador from North East (n = 10)), which were obtained by paired-end reads strategy using Illumina technology. Sequencing whole genomes across multiple samples in a population provides an unprecedent opportunity for comprehensively characterize polymorphic variants in a population. Even though the 1KGP has offered brief insights into the importance of population-level sequencing, the low coverage of the pilot project has impaired the ability to confidently detect rare and low frequency variants. Studying private variants, in particular in admixed populations, are necessary in order to catalogue the human genome diversity, especially those responsible to complex diseases. Here we report preliminary results regarding the sequencing section of the project: Data quality: on average, each genome was sequenced 42 X (mean coverage), which 128 GB (on average) of bases successfully passed filter and aligned to the reference genome (NCBI37); 82% of bases that showed quality (QScore) >= 30 and 96% of Non-N reference bases with a coverage >= 10X; Polymorphisms: we identified 15033927 SNPs, out of which 13333782 have been previously reported by dbSNP (build 137) and 1700145 have not; Summary of identified SNPs annotated using ANNOVAR: UTR3(10625), ncRNA_exonic (3759), intergenic (924842), splicing (60), ncRNA_intronic (51526), ncRNA_UTR3 (216), upstream (11897), upstream;downstream (426), ncRNA_splicing (17), exonic (7467), ncRNA_UTR5 (79), UTR5 (3668) ,downstream (9379), intronic (543793)and UTR5;UTR3 (5). The high-coverage strategy of the EPIGEN-Brazil initiative complements the information that is being obtained from the 1KGP, increasing the number of sequenced individuals with African and Native American ancestry.

1579T

Multi-platform and cross-methodological reproducibility of transcriptome profiling by RNA-seq in the ABRF Next-Generation Sequencing Study (ABRF-NGS). *C.E. Mason[1,2], S. Li[1,2], S.W. Tighe[3], C.M. Nicolet[4], D. Grove[5], S. Levy[6], W. Farmerie[7], A. Viale[8], C. Wright[9], P.A. Schweitzer[10], Y. Gao[11], D. Kim[11], J. Boland[12], B. Hicks[12], R. Kim[13], S. Chhangawala[1,2], N. Jafari[14], N. Raghavachari[15], C. Hendrickson[6], D. Roberson[12], J. Rosenfeld[16], T. Smith[17], J. Underwood[18], M. Wang[19], P. Zumbo[1,2], D. Baldwin[20], G. Grills[10], ABRF-NGS Consortium.* 1) Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York, USA; 2) The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York, USA; 3) Vermont Cancer Center, University of Vermont, Burlington, Vermont, USA; 4) Keck Medical Center, University of Southern California, Los Angeles, California, USA; 5) The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA; 6) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA; 7) Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, Florida, USA; 8) Memorial Sloan-Kettering Cancer Institute, New York, New York, USA; 9) Roy J. Carver Biotechnology Center, University of Illinois, Urbana, Illinois, USA; 10) Biotechnology Resource Center, Institute of Biotechnology, Cornell University, Ithaca, New York, USA; 11) Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA; 12) NIH/NCI/SAIC-Frederick, Gaithersburg, Maryland, USA; 13) Genome Center, University of California, Davis, Davis, California, USA; 14) Center for Genetic Medicine, Northwestern University, Chicago, Illinois, USA; 15) NIH/NHLBI, Bethesda, Maryland, USA; 16) Division of High Performance and Research Computing, University of Medicine and Dentistry of New Jersey, Newark, New Jersey, USA; 17) PerkinElmer Inc., Seattle, Washington, USA; 18) Pacific Biosciences, Menlo Park, California, USA; 19) Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA; 20) Pathonomics LLC, Philadelphia, Pennsylvania, USA.

Next-generation sequencing (NGS) technology applications such as RNA-sequencing (RNA-seq) have dramatically expanded the potential for novel genomics discoveries, but the proliferation of platforms and protocols has created a need for reference data sets to help gauge their performance. Here we describe the ABRF-NGS Study on RNA-seq, which leverages replicate experiments across multiple sites using two reference RNA standards tested with four protocols (polyA selected, ribo-depleted, size selected, and degraded RNA), and examined across five NGS platforms (Illumina HiSeq 2000/2500, Life Technologies PGM/Proton, Roche 454 GS FLX+, and PacBio). These results show high intra-platform consistency ($R2 > 0.9$), high inter-platform concordance ($R2 > 0.8$), and a large set of novel splice junctions observed across all platforms. Also, we observe that ribosomal RNA depletion can both salvage degraded RNA samples and also be readily compared to polyA-enriched fractions. These data provide a broad foundation for standardization, evaluation and improvement of RNA-seq methods.

1580F

Identifying Genomic Copy Number Alteration and Loss of Heterozygosity in Next-Generation Sequence Data. *S. Rozen[1], J.R. McPherson[1], Y. Wu[1,2], P. Tan[1,3,4].* 1) Duke-NUS Graduate Med Sch, Singapore, Singapore; 2) Singapore-MIT Alliance in Research and Technology; 3) National Cancer Centre, Singapore; 4) Genome Institute of Singapore.

Losses and duplications of large genomic regions resulting in copy number alterations or loss of heterozygosity (LOH) are common drivers of cancer development. Until recently, these aberrations have been characterized primarily by means of SNP and copy-number oligonucleotide arrays. However, next-generation sequencing has become a dominant approach for detecting mutations in cancer genomes, and the read data generated in these studies can also be used for characterizing genomic aberrations in cancers. We developed an analytical approach and software that considers read depth and read counts for each allele to identify genomic regions with copy number alteration or LOH. The approach, 'RDAAC' (Read Depth and Allele Counts) requires aligned sequencing reads from matched malignant and non-malignant DNA. RDAAC simultaneously estimates the proportion of non-tumor genomes in the sample and the number of copies of each SNP allele in the tumor sample. We assessed RDAAC's performance in two ways. First, we compared its results to results from analyzing SNP arrays. Second, we assessed its performance on simulated data, for which true copy number alterations, regions of LOH, and admixture of non-malignant DNA were known. RDAAC functioned well on whole-exome sequencing data even in the presence of generalized polyploidy and >50% admixture of DNA from non-malignant cells. We developed an R package that implements (1) RDAAC on top of SAMTtools(Li, et al. 2009) and ASCAT (which was designed for microarray data, Van Loo et al., 2010) and (2) the simulator for synthetic data.

1581W

Reducing platform bias in next-generation sequencing. *L. Saag[1,2], U. Gerst Talas[1], M. Mitt[1], R. Villems[1,2], M. Metspalu[2].* 1) University of Tartu, Tartu, Estonia; 2) Estonian Biocentre, Tartu, Estonia.

Next-generation sequencing platforms enable re-sequencing of the human genome at high speed and relatively low cost. At the same time, there are discrepancies in base calls from different sources - the platform bias. The relatively low concordance between the results from different sequencing platforms has been reported (Lam et al. Nature Biotechnology 30, 78-82; Ratan et al. PLOS One 8: e55089). This can be a serious problem when data from various sources need to be pooled, especially, for example, in cases where different populations (or cases and controls) are sequenced with different technologies. We compared the concordance of the base calls of four human genomes, produced on the widely utilized Illumina and Complete Genomics (CG) platforms, sequenced at average coverage of 26x and 40x, and processed with BWA plus GATK and CG proprietary pipeline, respectively. The effect of various filters on the platform bias was evaluated. As expected, the bias was much larger concerning indels (at least 20% of the variants after initial QC) than compared to single nucleotide polymorphism (SNP) calls, however, significant discrepancy was detected also for the latter (at least 5% of the variants after initial QC). It is known that the next-generation sequencing errors are mostly due to mismapping of the sequence reads. Recently Lee and Schatz (Bioinformatics 28, 2097-2105) and Derrien and others (PLOS One 7: e30377) proposed tools that can help to avoid such errors by identifying and excluding regions of the genome with low mappability. We built on similar in-house efforts. Applying the resulting poor-mapping windows together with adjusted quality and coverage filters, we were able to minimize the platform bias significantly. We work towards further reducing the discordance including evaluating the effect of other sequence read mappers (Bowtie2, GATK2, etc).

1582T

Customized and Personalized Next Generation Genomics. *A.N. Singh.* Computer Science, Virginia Tech, Blacksburg, VA.

Structural variations, SVs, with size 1 base-pair to several 1000s of base-pairs with their precise breakpoints and single-nucleotide polymorphisms, SNPs, were determined for members of a family of four. It is also discovered that the mitochondrial DNA is less prone to SVs re-arrangements than SNPs and can have paternal leakage of inheritance which proposes better standards for determining ancestry and divergence between races and species. Sex determination of an individual is found to be strongly confirmed by means of calls of nucleotide bases of SVs to the Y chromosome. SVs would serve as fingerprint of an individual contributing to his traits and drug responses. These in silico techniques for analysis would become such a widespread application that a total transformation of the bio and medical industry would go through.

1583F

Identifying Mendelian Disease Genes; an Analysis Tool of PhenoDB.
N. Sobreira[1], F. Schiettecatte[2], D. Valle[1], A. Hamosh[1]. 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) FS Consulting, Salem, MA.

WES has been the main method used to search for Mendelian disease genes in the last 3 years. But identifying the pathogenic mutation among thousands to millions of genomic variants is a challenge, and variant prioritization strategies are required. The choice of these strategies depends on the availability of well-phenotyped patients and family members, the mode of inheritance, the severity of the disease and its population frequency. Here we describe the PhenoDB analysis tool developed to rapidly identify a list of candidate variants by applying various strategies for disease variant prioritization. PhenoDB is a Web-based portal for managing and analyzing phenotypic/clinical information and WES/WGS data. The VCF and ANNOVAR files, genotyping array data, QC report, and final results file are available in the Analysis module, in which the user can choose the family members to include, the inheritance mode, what kind of variants to filter on (missense, nonsense, splicing, 3'UTR, 5'UTR, synonymous), optionally exclude variants found in the dbSNP 126, 129 and/or 131, the MAF value for exclusion of variants in the 1000 genome and EVS databases, and exclusion of chrX variants. The analysis runs in less than 1 minute and generates a list of candidate genes plus a log of the prioritization strategy. We also add information from external databases such as OMIM, MGI, GeneCards, etc. Multiple analyses with different prioritization strategies can be saved. Any saved ANNOVAR file or analysis result can be filtered using the 57 genes listed in the ACMG incidental findings list or compared against the genes in an OMIM phenotypic series. The user can also compare analysis results across families for overlapping mutated genes and cohorts (with or without locus heterogeneity) can be analyzed. A Sandbox is provided where ANNOVAR files from families not in PhenoDB can be uploaded and analyzed. Genes and variants can be searched across families in the analysis results. The phenoDB analysis tool has been proved to be easy, fast and efficient. In the Baylor-Hopkins Center for Mendelian Genomics we have now analyzed 83 families using this tool. We have identified the causative gene in 14 families and suggested candidate genes that are being further investigated in 39 other families. PhenoDB including the analysis tool is freely available through the Johns Hopkins McKusick-Nathans Institute of Genetic Medicine (http://phenodb.net).

1584W

The European Genome-phenome Archive (EGA).
V. Kumanduri, A. Datta, J. Almeida-King, L. Clarke, I. Lappalainen, P. Flicek, J. Paschall. EMBL-EBI, EMBL-EBI European Bioinformatics Institute, Hinxton, United Kingdom.

The European Genome-phenome Archive (EGA) V. Kumanduri, A. Datta, J. Almeida-King, L. Clarke, I. Lappalainen, P. Flicek and J. Paschall EMBL-EBI Wellcome Trust, Cambridge, United Kingdom The European Genome-phenome Archive (EGA) provides a permanent archive for genetic and phenotypic data and a secure means to share subject level data. EGA contains data collected from individuals for the purpose of medical or genetic research and whose consent agreements require security and request management. Data access decisions are not made by the EGA but by a submitter defined Data Access Committee (DAC). Once DAC authorization to the data has been granted, the EGA provides secure tools for downloading and decrypting the authorized files and general support for questions related to the data content. The EGA is available at http://www.ebi.ac.uk/ega/. Data from more than 350 studies, comprising 500 datasets and 200,000 individuals have been submitted to the EGA and is made available to authorized researchers. These studies include data from array-based genotyping experiments as well as raw DNA sequence data from re-sequencing or transcriptomics projects. While continuing to perform the main function of archiving datasets exactly as submitted, EGA is now developing an option for submitters to choose value-added processing steps. Specifically, we aim to apply a defined analysis pipeline to provide standardized variant call sets that can aid in comparisons of methods and in the combining of data from multiple sources. In order to achieve this, the EGA is building a data analysis workflow that allows us to align reads against a reference, provide high quality variant calling and distribute phased genotype information for the analysed samples. For those submitters who request this standardized processing, these features will be run in addition to archiving of the primary data files and analysis products. Anyone interested in further information about the EGA in general should contact ega-helpdesk@ebi.ac.uk.

1585T

THE MEDICAL EXOME PROJECT: From concept to implementation.
A. Santani[1], S. Gowrishankar[3], C. da Silva[2], D. Mandelkar[1], A. Sasson[1], M. Sarmady[1], R. Shakhbatyan[3], S. Tinker[2], D. Church[5], B. Funke[3,4], M. Hegde[2]. 1) Department of Pathology and Laboratory Medicine, Childrens Hospital of Philadelphia, Philadelphia, PA; 2) Emory Genetics Laboratory, Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 3) Laboratory for Molecular Medicine, Cambridge, MA; 4) Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 5) National Center for Biotechnology Information, NLM, NIH, Bethesda, MA.

Next Generation Sequencing (NGS) is being rapidly adopted by clinical laboratories and has allowed development of disease targeted gene panels and more recently exome and genome sequencing. One drawback of gene panels is their suboptimal clinical sensitivity, which is often only marginally increased by adding novel genes, leaving the need for additional testing if no mutation is identified. In contrast, exome and genome sequencing (ES/GS) interrogates all genes, but is flawed by incomplete coverage and interrogates a significant number of genes with no clinical relevance. THE MEDICAL EXOME project was launched to bridge the gap between gene panels and ES/GS by curating the genes of clinical relevance and improving the technical performance of exome capture assays by enhancing their coverage. Candidate medically relevant genes were drawn from three primary sources: HGMD, OMIM (morbid map) and ClinVar. Additional genes were chosen from Cosmic, Jackson lab's MGI database, Gene Tests, Orphanet and locus specific databases (LOVDs). This draft medical exome was further enhanced by adding disease risk loci (NHGRI GWAS catalog) and pharmacogenetic targets (PharmGKB). A pilot curation project removed genes that had no evidence for a role in disease by developing a scoring grid from 0 (no evidence) to 3 (definitely associated with disease). In addition, exons with conserved paralogous sequences were excluded to avoid technical and analytical challenges. Over 5,000 genes were extracted from the above databases, 877 of which were removed due to lack of evidence for a role in disease. A pilot analysis of an enhanced version of the Agilent V5 exome indicated optimal coverage (all bases >20x) for ~2/3 of the exons, insufficient coverage (<40% of bases >20x) for ~10% and various gradations of completeness for the remaining ~20%. Alternate approaches are being employed to enhance the coverage of suboptimal genes. We will present the development of a curated gene list using an evidence-based gene scoring approach and preliminary results of the enhanced medical exome. This list will be made available to the community through public resources such as the ICCG consortium and resources maintained at NCBI. The availability of a curated medical exome will improve sensitivity and specificity of NGS based exome and genome sequencing and will provide a stepping stone for standardizing interpretation of genetic test results by clinical testing laboratories.

1586F

Accurate identification of allele imbalances in samples with limited genotype information available. *M.L. Buchkovich[1], K.E. Eklund[1], Q. Duan[1], Y. Li[1,2,3], K.L. Mohlke[1], T.S. Furey[1,4].* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 3) Department of Computer Science, University of North Carolina, Chapel Hill, NC; 4) Biology Department, University of North Carolina, Chapel Hill, NC.

Quantitative sequence data that detect epigenetic modifications, transcription factor binding and nucleosome depletion identify regions of transcriptional regulation. At heterozygous sites within regulatory regions, enrichment of sequence reads containing one allele, or allele imbalance, can indicate allelic differences in regulatory activity. Allele-aware aligners map sequences containing either allele with equal probability to a reference genome. We have shown previously that when sequencing-based genotypes are available, allele-aware aligners map sequences to heterozygous sites with increased accuracy compared to traditional aligners. We now have expanded our allele-aware alignment methods to identify allele imbalance in additional samples with more limited genotype information available. We aligned 20 bp DNase-seq sequences from GM12878, a human lymphoblastoid cell line, including complete (genomic sequencing-based), and limited (Human1M-Duo genotyping + imputation) genotype information. To increase alignment accuracy, we created custom reference genomes using the complete or limited genotype data by changing hg19 to match the allele present in GM12878 at ~1.2M-1.3M homozygous sites. We aligned sequences to these references using the allele-aware aligner GSNAP and heterozygous sites from each genotype set. Requiring at least two sequences containing each allele, we identified (binomial test, P<.001) 191 sites of allele imbalance when using complete genotype data and 70 sites when using limited genotype data. Identifying imbalances in GM12878 data using limited genotype information had a positive predictive value of 84% (62/70), a sensitivity of 32% (62/191), and a specificity of 99.99% (69,505/69,513). Only 3 of 129 imbalanced sites identified solely with complete genotypes were predicted by imputation to be heterozygous, suggesting that imputation missed the other variants. Testing for allele imbalance at additional heterozygous sites detected in the DNase-seq alignment could increase our ability to identify imbalances at the 126 missed sites and improve sensitivity. In conclusion, this approach accurately identifies sites of allele imbalance without using complete, sequence-based genotype information. The ability to identify these imbalances in a broader range of datasets from multiple tissues will be particularly useful in understanding how variants with allelic differences in gene transcription regulation influence complex traits and diseases.

1587W

Scalable variant identification and imputation across large multigenerational pedigrees from high-throughput sequencing data by joint Bayesian variant calling. *F.M. De La Vega[1], S. Malakshah[1], R. Littin[2], L. Trigg[2], A. Jackson[2], D. Ware[2], J.G. Cleary[2].* 1) Real Time Genomics, Inc., San Bruno, CA, USA; 2) Real Time Genomics, Ltd., Hamilton, New Zealand.

Whole-genome sequencing (WGS) is being applied in research and clinical studies to identify variants for rare diseases and complex traits. Often these studies sequence related individuals to varying sequencing depths across a pedigree. However, variant calling methods for high-throughput sequencing (HTS) ignore such relationships or are not scalable to multigenerational pedigrees. We present a Bayesian framework that jointly calls variants across a pedigree implicitly leveraging shared haplotypes. When calling variants at a given position for a nuclear family, we look-up the alignment data of all family members simultaneously, scoring genotypes across pedigree members with priors based on Mendel's laws of variant segregation, and handling sex chromosomes as special cases. As validation, we analyzed data from a 17 individual, 3-generation CEPH pedigree sequenced to 40X average depth. We focus our analysis on NA12878, a female in the second generation and for which extensive orthogonal validation data is available from the 1000 Genomes Project and other efforts. Compared to singleton calling, our family caller produced more high quality variants and eliminated spurious calls as judged by standard quality metrics such as Ti/Tv, Het/Hom ratios, and dbSNP/OMNI array data concordance. Through the analysis of the segregation of variants to the 11 offspring of the third generation called independently, we estimate less than 1%, 2.5% and 2.75% of false positive calls for SNV, indels, and MNPs, respectively. Family designs should allow identifying de novo mutations, which are of great interest in neurodevelopmental and other disorders. By allowing a small prior for de novo mutations and scoring specifically for this type of mutations, we observed 100% and 96% sensitivity in identifying previously validated germline and somatic de novo mutations in NA12878, respectively, concurrent with a 10X reduction in Mendelian errors. Simple Mendelian scoring would scale exponentially across multi-generational pedigrees. Instead, we implemented a method for acyclic prior propagation beyond nuclear families, approximating the full pedigree Bayesian solution in linear time. Our results show that our joint pedigree calling method outperforms singleton and population variant calling in pedigrees, allows for the identification of de novo mutations with greater specificity, and is scalable to large genomics and human disease studies.

1588T

Advantages of Single-Molecule, Solid-State Nanodetectors for Genome Mapping. *B. Galvin, B. Bready, D. Dederich, A. Forget, J. Frietas, H. Geiser, P. Goldstein, D. Goryunov, H. Heaton, D. Hevroni, M. Jouzi, M. Kaiser, H. Lee, T. Leffert, D. Lloyd, S. Marappan, E. Olejnik, J. Oliver, L. Seward, A. Snider, J. Tang, J. Thompson, M. Zhang.* Nabsys, Inc., Providence, RI.

Next-generation sequencing systems have revolutionized the use of sequencing for answering many biological questions. However, the limited read lengths and requirement for DNA amplification in order to detect signals cause the data quality and completeness to suffer significantly. Complete de novo sequencing of even relatively short prokaryotic genomes is not possible with current NGS systems which must be supplemented with other data or reference genomes to even make an attempt at producing a finished genome. Optical mapping technologies have arisen to fill the long-range information gap but methods introduced thus far have limited resolution and low throughput relative to what is needed for routine mapping of human genomes. In contrast, solid-state, silicon-based nanodetectors have been developed that use electronic measurement to distinguish DNA features and thus are not limited to the resolution limits imposed by the wavelength of light. This, coupled with the high throughput that can be attained using detectors that translocate DNA at greater than 1 million base pairs per second, provide a path to routinely generating de novo assembled maps of human and similar size genomes. The ability to generate high resolution maps for single DNAs 10s to 100s of kilobases long will be shown. Performance characteristics of our detectors demonstrating high signal-to-noise characteristics, sub-diffraction limit resolution, and massive throughput will be discussed.

1589F

Validation of the second and the third generation sequencers by *de Novo* assembling using AT rich, repetitive, homopolymeric Human BAC DNA. *T. Hirano[1,2], Y. Terabayashi[3], K. Teruya[3], M. Teruya[4], M. Shimoji[3], H. Tamotsu[3], A. Arasaki[3], K. Nakano[3], A. Shiroma[3], K. Satou[3], Y. Yamaoka[5,6], A. Sekine[7].* 1) Fellow, Okinawa Science and Technoogy Promotion Center, Okinawa, Japan; 2) Founder, Okinawa Institute of Advanced Sciences, Okinawa, Japan; 3) Research and Development Division, Okinawa Institute of Advanced Sciences, Okinawa, Japan; 4) Okinawa Industrial Technology Center, Okinawa, Japan; 5) Division of International Health, Oita University, Oita, Japan; 6) Department of Medicine, Baylor College of Medicine, Houston, TX; 7) EBM Research Center, Kyoto University, Kyoto, Japan.

The sequencing technology is indispensable for any field of genomic analysis. Human Genome Project was drastically accelerated by sequencing and assembling technology. These technologies have been continuously advancing with the appearance of the second and the third generation sequencers. With increase of the number and the kinds of sequencers, the standardization of genome sequencing is coming to be important. The reference materials, reference data, and reference methods for the assessment of performance of genome sequencing are developed in the "Genome in a bottle Consortium" organized by NIST. Culture cells were often used as the Standard Reference Material in typical experiments. But in genome sequencing, the stability of DNA is the most required characteristics. Genomic mutations are often produced during its cultivation. In this study, the second and the third generation sequencers were validated using Human BAC clone of RP-11 library which is used in Human Genome Project and mapped on CYP gene coding locus. By using BAC clone, the performance of each sequencer was properly accessed without effects of ploidy. The selected clone involves AT rich, repetitive, and homopolymeric locus. The clone was sequenced by Capillary, 454, Illumina, Ion Torrent, and PacBio. The read coverage bias, and raw read accuracy, including specific errors on homopolymeric locus were evaluated by using the result of mapping analysis on RefSeq. The sequence produced by each sequencer was de Novo assembled using Arachne, Newbler, Velvet, Mira, and HGAp, respectively. The unique contig which covers whole length without gap was constructed by using only the sequence of PacBioRS, because of long reads and equal coverage. The established method of de Novo assembling was applied for other clones or stomach cancer causing microbial, Helicobacter pylori. Additionally, our recent result of de Novo assembling of using PacBioRS II which represents the enhancement of read length and the improvement of read quality will be shown.

1590W

Developing 400-base Sequencing for the Ion PGM® System. *A.I. Kraltcheva, D.J. Mazur, G. Luo, X. Peng, T.L. Lincecum, G. Lowman, M.A. Landes, B. Strohecker, K.N. Aguinaldo, T. Nikiforov, E. Tozer, J.J. Shirley, P. Vander Horn.* ION Torrent, LIFE Technologies, Carlsbad, CA.

The utility of long sequencing reads for enhanced genomic annotation and assembly has been a key differentiator for successful next generation sequencing applications. Through optimization of both the sequencing and amplification biochemistries, we have significantly improved Ion Torrent's semiconductor chip-based sequencing system. Along with improvements in engineering and software, these innovations result in robust high quality 400-base sequencing reads on the Ion Torrent PGM™ platform. Using this improved 400 bp biochemistry on a 318 chip, we have produced >3G aligned Q20 bases with even coverage on the *Escherichia coli* genome. Additionally, with optimized emPCR conditions we have rapidly improved coverage uniformity of the more challenging *Rhodobacter sphaeroides* genome from 88% to 95%. Compared to the previous PGM™ sequencing protocols, the Ion Torrent 400bp Template/Sequencing kits give nearly double the read length and throughput. Such improvements in the PGM® system enable a broader range of applications, such as enhanced de novo genome assemblies, Human Leukocyte Antigen (HLA) sequencing, bacterial identification, and meta-genomic analysis. *For Research Use Only. Not for use in diagnostic procedures.

1591T

GenomeBrowse: Visual analytics and false-positive discovery for DNA and RNA-seq NGS data. *G. Linse Peterson, G. Rudy, S. Gardner, M. Thiesen, A. Laughbaum.* Golden Helix, Bozeman, MT.

High-throughput sequencing (HTS) has recently provided price competitive alternatives to microarrays for both RNA expression profiling with the RNA-seq protocol and DNA genotyping with whole genome and whole exome sequencing. Although the bioinformatics tools have matured for secondary analysis of sequence data, including alignment, variant calling, and gene and transcript level quantification, the outputs of these tools often require inspecting the 'raw read alignments' for putative variants and genes with interesting expression profiles. Investigating these variants in their VCF format and the alignments in BAM format allows for detection of false-positives as well as aiding the interpretation process by providing a rich genomic context.

We introduce Golden Helix GenomeBrowse™, a free visualization tool for DNA and RNA sequence alignment and variant calls along with annotations tracks from a rich catalog of pre-curated public data. GenomeBrowse is built from the ground up with the guiding principles of (1) working seamlessly with cloud-hosted data as fluidly and quickly as local files, (2) being intuitive to use for non-bioinformaticians to utilize in their research, and (3) having a multi-threaded architecture to make big-data visualization and analysis accessible to anybody capable of navigating Google Earth. By including integration with a rich repository of public data, users have no barriers to the process of interpreting their sequencing results. In particular, we demonstrate the ability of GenomeBrowse to stream exome sequencing of a trio from the Amazon Cloud from whole genome views down to the gene level with annotation tracks ranging from 1000 Genomes, dbSNP, genes, and miRNAs. We show how GenomeBrowse can highlight false-positive Single Nucleotide Variants and small Insertion/Deletions, confirm the inheritance pattern of putative functional variants, and aid in the interpretation of a variant's impact.

1592F

Improving the robustness of personal genome variant discovery: the impact of technical replicates, sample source and analysis method. *D. Mittelman[1], A. Del Duca[2], R.M. Iwasiow[2], N. Leibovich[3], J. Wang[3], M. Tayeb[2].* 1) Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA; 2) DNA Genotek Inc. (A wholly owned subsidiary of OraSure Technologies Inc.), Ottawa, ON, Canada K2K 1L1; 3) Arpeggi, Inc., Austin, TX, USA 78749.

Background: Whole genome sequencing (WGS) promises to disrupt prognostic and diagnostic genetic testing in the clinic. Although the adoption of WGS is accelerating, the standards and acceptance criteria for clinical WGS are not fully defined. Variant calling robustness and reproducibility of WGS analysis was assessed across routinely used biological sample types, among technical replicates, and between sequencing library preparations. Consideration was also given to sample-to-result costs. **Methods:** Two multigenerational families comprising a total of 7 individuals were enrolled in the study based on low (<5%), medium (6-20%) and high (>21%) levels of non-human DNA content in their saliva as measured by qPCR. Matched sample pairs of blood (EDTA, Becton Dickinson, Inc.) and saliva (Oragene, DNA Genotek Inc.) were prepared from each participant and sequenced through the Illumina Genome Network (IGN), generating a total of 20 whole genome datasets. Multiple variant discovery pipelines, including CASAVA and GATK were used to interrogate reproducibility of biological and technical replicates and Mendelian consistency in parent-offspring trios. The computational burden (i.e. cost from DNA-to-result) between sample types was also compared. **Results:** A direct linear correlation (R^2= 0.88) between non-human DNA content in a saliva sample and the proportion of unmapped reads after read alignment was found, suggesting that reads derived from non-human DNA are often not mapped. The average difference in unmapped reads between paired blood and saliva samples from the same individual was 8.7%. On average, 4% of variant calls (SNPs and small indels) were discordant between technical replicates of blood samples. Likewise, discordant variant calls between saliva and blood averaged 4%. Finally, QC parameters, such as minimum read depth, are suggested, which help maximize variant calling consistency among technical replicates and across different sample types. **Conclusion:** Variant concordance between matched pairs of saliva and blood is equivalent to concordance between technical replicates of blood. This strongly suggests that saliva sampling is an effective alternative to blood for whole genome sequencing analysis.

1593W

Genome in a Bottle Consortium: Update on a Public-Private-Academic Consortium Developing a Standards Infrastructure for Human Genome Sequencing. *S.A. Munro, J. Zook, M. Salit, Genome in a Bottle Consortium.* National Institute of Standards and Technology, Gaithersburg, MD.

In 2012, NIST convened the Genome in a Bottle Consortium (www.genomeinabottle.org) to develop the reference materials, reference methods, and reference data needed to assess confidence in human whole genome variant calls. We will report on progress of the four consortium working groups. The Reference Material (RM) Selection and Design group has selected the first 4 genome families from the Personal Genome Project for whole genome RMs, a candidate cancer/normal cell line pair, and artificial structures for point mutations. The Measurements for Reference Material Characterization group has initiated experiments to characterize a pilot RM (NA12878) using multiple sequencing methods, other methods, and validation of selected variants using orthogonal technologies. The Bioinformatics, Data Integration, and Data Representation working group has developed data integration and analysis methods, data sharing protocols, and representation formats. The Performance Metrics and Figures of Merit group has initiated collation of proposed performance metrics and figures of merit from consortium members, leveraging existing resources. The products of these working groups will be a set of well-characterized whole genome and synthetic DNA RMs along with the methods (documentary standards) and reference data necessary for use of the RMs. These products will be designed to help enable translation of whole genome sequencing to clinical applications by providing widely accepted materials, methods, and data for performance assessment.

1594T

Change can be good: updating the human reference genome assembly. *V.A. Schneider[1], P. Flicek[2], T. Graves[3], T. Hubbard[4], D.M. Church[1] for the Genome Reference Consortium.* 1) NCBI/NIH, Bethesda, MD 28092; 2) EBI, Hinxton, Cambridge, CB10 1SD, U.K; 3) The Genome Institute at Washington University, St. Louis, MO 63108; 4) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, U.K.

Since its publication in 2003, the human reference genome assembly has become an integral reagent in biological research. In today's era of next generation (NGS) and personal genome sequencing, it is the framework upon which individual genome analysis is based as well as a common coordinate system for genomic annotation. While the human reference retains its distinction as the highest quality mammalian genome, studies in the past decade have demonstrated that the linear chromosome model initially used for the assembly is insufficient in its ability to represent population variation. Within this model, the insertion of sequences from multiple haplotypes at complex variant locations led to non-existent allele combinations and artificial gaps. To provide a more accurate genome representation and better represent variation, the Genome Reference Consortium (GRC), the group responsible for the upkeep of the human reference assembly, developed a new assembly model. First implemented for GRCh37 (hg19), the new model retained the linear chromosomes but introduced alternate loci, scaffold sequences placed into chromosome context via alignment, as a mechanism for providing alternate sequence representation at complex or variant genomic regions. The new assembly model also introduced patches, operationally similar to the alternate loci, as a means for providing users access to assembly corrections and new variant sequence representations without coordinate-disrupting full assembly updates. Since the production of GRCh37, the GRC has released patches to 170 genomic regions, representing 3% of the chromosome sequence and adding more than 6 Mb of novel sequence and over 180 gene models not found on the chromosome assembly. This model necessitates an update to current analysis tools that cannot yet accommodate allelic duplication. We will present data demonstrating the added benefit of using the full assembly, including alternate loci and patches, with respect to gene annotation, NGS read alignment and variation analyses. Nonetheless, coordinate-changing major assembly releases remain vital as underlying mis-assemblies in the chromosome assembly can still confound analysis. The GRC is planning to release GRCh38 in the fall of 2013. We will present our progress on the new assembly and illustrate several of its features, including the large-scale correction of erroneous bases, addition of missing sequences and retiling of misassembled regions.

1595F

A Fast Solution to NGS Strand-specific RNA Library Prep. *C. Sumner, D. Munafo, P. Liu, L. Apone, B. Langhorst, E. Yigit, F. Stewart, E. Dimalanta, T. Davis, J. Bybee, L. Mazzola, D. Rodriguez, V. Panchapakesa.* New England Biolabs, Ipswich, MA.

RNA-Seq is a powerful technique that allows for sensitive digital quantification of transcript levels, and enables the detection of noncanonical transcription start sites as well as termination sites, alternative splice isoforms and transcript mutation and editing. Standard 'next-generation' RNA-sequencing approaches generally require double-stranded cDNA Synthesis, which erases RNA strand information. Strand information is important for correct annotation of novel genes, identification of antisense transcripts with potential regulatory roles, and for correct determination of gene expression levels in the presence of antisense transcripts. To obtain high quality stranded sequence information, RNA sequencing libraries must be highly-strand-specific and highly diverse. Ideally, they should be produced at high yield and by a fast protocol. To address these challenges, we have developed the NEBNext® Ultra Directional RNA Library Prep Kit for Illumina, which is based on labeling second strand cDNA with dUTP, allowing for subsequent excision to retain RNA strand information. This approach is incorporated into a streamlined workflow, enabling directional RNA library construction in ~5 hrs with minimal hands on time. Here, we provide an analysis of library quality (library complexity, continuity of gene coverage, strand specificity, and 3' and 5' bias), as well as demonstrate compatibility with NEBNext®96 barcode strategy and ribosomal RNA removal.

1596W

Targeted Sequencing for Preterm Birth Associated Genes. *A. Uzun[1,2], I. Kurihara[1], B. McGonnigal[1], J.F. Padbury[1,2].* 1) Pediatrics, Women and Infants Hospital, Providence, RI; 2) Pediatrics, Brown Medical School, Brown University, Providence, RI.

Despite significant advances in the care of pregnant women and low birth weight infants, preterm birth remains the leading cause of newborn morbidity and mortality as well as the main cause of hospitalization in the first year of life in the United States. We hypothesize that rare variants in the genes that contribute to the risk of PTB can be identified using new bioinformatics approaches coupled with high-throughput technologies applied to appropriate cohorts of patients. Understanding of the complex pattern of gene expressions, which enhance or attenuate the risk of Preterm Birth (PTB) will require new strategies. We have developed an alternative approach to identify a more manageable set of genes for preterm birth, which nonetheless incorporates some elements from the discovery in genome wide investigations. We analyzed genotype information from a large GWAS which included >660,000 SNPs from 1,000 mother and infant controls and 1,000 mother and infant preterm birth cases, We used gene set enrichment analysis (GSEA) to investigate the individual contribution(s) of biological pathways to the genetic architecture of preterm birth. We used both our curated genes and a genome-wide approach in the pathway analysis. We identified 329 genes nested within 69 pathways that are highly associated with preterm birth. For a proof of concept, we used long-range PCR to generate a DNA library from the top 11 genes from this gene set. We sequenced gDNA from 4 women with 3 generations of preterm birth and 8 term controls. High quality sequence data from well balanced pools was observed. There were an average of 22,000,000 reads from each patient, with an average of 99 perfect index reads and a Q30 of 91%. The mean Phred score for each patient was 36. A total of 2669 initial variants were reduced to 1015 following filtering for the quality indices described above. 760 were already identified in one of several database archives, including 1000 Genomes, the Exome Sequencing Project or dbSNP. 255 were unique of which 5 were exonic and 4 predicted to be deleterious in genomic regions encompassing PTGS2 (1 deleterious), AKT1 (2 deleterious), and TP53 (1 deleterious). There have been too few samples sequenced to date for valid statistical comparisons but these results nonetheless confirm the importance of not restricting the library construction to exonic sequences alone.

1597T

Well-characterized genomes for understanding genome sequencing performance: integrating datasets from multiple sequencing platforms to form highly confident snp and indel calls. *J.M. Zook[1], B. Chapman[2], O. Hofmann[2], W. Hide[2], D. Mittelman[3], J. Wang[4], M. Salit[1], Genome in a Bottle Consortium.* 1) Biosystems and Biomaterials Division, National Institute of Standards and Technology, Gaithersburg, MD; 2) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 3) Virginia Bioinformatics Institute, Blacksburg, VA; 4) Arpeggi, Inc., Austin, TX.

Clinical translation of genome sequencing requires methods to understand accuracy of genotype calls at millions or billions of positions across a genome. Previous work showing high discordance between sequencing methods and algorithms has highlighted the need for a highly confident set of genotypes across a whole genome that could be used as 'truth' for understanding accuracy. Therefore, we have developed methods to make highly confident SNP, indel, and homozygous reference genotype calls for NA12878, the pilot genome for the Genome in a Bottle Consortium (www.genomeinabottle.org). To minimize bias towards any sequencing method, we integrate 9 whole genome and 3 exome datasets from 5 different sequencing platforms (Illumina, Complete Genomics, SOLiD, 454, and Ion Torrent). We also integrate variant calls from variant callers using only mapping, mapping with local de novo assembly, and global de novo assembly. The resulting genotype calls are more accurate than calls from any individual dataset, and allow performance assessment of more difficult variants than using microarrays. In addition, regions for which no confident genotype call could be made are identified, so that both false positive and false negative error rates can be assessed. We are integrating our highly confident variants into the GCAT website (www.bioplanet.com/gcat) so that anyone can interactively generate performance metrics for different combinations of sequencing datasets, mapping algorithms, and variant callers. In addition, as part of the Genome in a Bottle Consortium, we plan to continue to improve the characterization of the pilot genome and additional genomes, expanding the characterization to additional variant types and more difficult regions of the genome.

1598F

Exome Sequencing Facilitated by A Fast Library Preparation Method with Low Nanogram DNA Input. *P. Liu[1], C.L. Hendrickson[2], B. Boone[2], B. Langhorst[1], L. Apone[1], D. Munafo[1], Y. Yigit[1], C. Sumner[1], V. Panchapakesa[1], D. Rodriguez[1], F. Stewart[1], T. Evans Jr[1], N. Nichols[1], S. Levy[2], E. Dimalanta[1], T. Davis[1].* 1) New England Biolabs, Ipswich, MA; 2) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama.

Even as the cost of next generation sequencing decreases, exome sequencing continues to have substantial value as an effective, lower-cost alternative to whole genome sequencing. Indeed, the use of exome sequencing in clinical diagnostic applications is rapidly increasing, and recent reports of the diagnostic use of exome sequencing have shown its ability in identifying disease-causing mutations for conditions such as congenital deafness and intellectual disabilities, which were formerly undiagnosable. The hope is that such sequencing will enable guidance of physicians on the proper course of treatment, determine personal response to a drug, assess risks for siblings, and possibly predict disease before it starts. A major challenge of exome sequencing-based clinical diagnosis is the limited quantities of sample DNA available. To address this limitation, we have developed a fast library preparation method using novel reagents and adaptors, including a DNA polymerase that has been optimized to minimize GC bias. This method enables exome sequencing of libraries constructed from low nanogram quantities of DNA amplified with limited PCR cycles, and can be used for both intact and fragmented DNA including FFPE and circulating plasma DNA. Moreover, the workflow is compatible with multiple NGS platforms.

1599W

Power and Limitations of RNA-Sequencing. *F. Staedtler, E.J. Oakeley.* Biomarker Development, Novartis Institutes for BioMedical Research (NIBR), Basel, Switzerland.

On behalf of the Sequencing Quality Control Consortium (SEQC). In the FDA-led SEQC (i.e., MAQC-III) project, different RNA-Seq platforms were tested across more than ten sites using well-established reference RNA samples with built-in truths in order to assess the discovery and expression-profiling performance of platforms and analysis pipelines. The results demonstrate that novel exon-exon junctions can still be discovered beyond existing comprehensive annotations and sequencing depth. With various investigations encompassing diverse performance metrics (accuracy, precision, reproducibility, mutual information, titration consistency, and mixing ratio recovery) and comparisons to qPCR and microarray platforms, we found high levels of inter-site and cross-platform concordance for differentially expressed genes. However, performance is clearly platform and pipeline dependent, and transcript-level profiling shows larger variation. The SEQC data sets with over 100 billion reads (10 TB) represent the deepest characterization of the transcriptome for any samples and thus provide a unique resource for testing future developments of RNA-Seq in clinical and regulatory settings.

1600T

Evaluation of Ion Torrent-based rapid deep sequencing for mutation discovery and prevalence screening in rare human myeloproliferative neoplasms and brain tumors. *L. Wang[1], S. Yamaguchi[2], L. Lewis[1], M. Holder[1], K. Chang[1], K. Walker[1], H. Dinh[1], H. Doddapaneni[1], D. Muzny[1], R. Gibbs[1], C. Lau[2], D. Wheeler[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, TX.

Tumors usually consist of a mixture of normal and tumor cells and most tumors, if not all, are heterogeneous, containing more than one clonal types. Sensitive and accurate detection of somatic mutations in low-purity and heterogeneous tumors is a challenging problem. Deep sequencing is needed to identify mutations with low allele fraction and to present each clonal type properly. For some rare tumors such as myeloid disease and brain tumors, which only a small amount of DNA can be collected, most sequencing technologies are not applicable. Ion AmpliSeq technology, based on ultra-high-multiplex PCR, requires as little as 10 ng of input DNA to target sets of interesting genes, delivers fast library construction and deep coverage on targeted regions, making it ideal for mutation discovery and prevalence screening in these rare, heterogeneous and polyclonal tumors. Here, we selected 40 and 23 interesting genes, respectively, from the discovery studies of myeloid disease and brain tumors, and designed two Ion AmpliSeq arrays targeting the complete coding sequences of interesting genes. A total of 148 samples were sequenced (16 reaction pools/run) on the PGM sequencer and 30 of them were also sequenced on the Proton Sequencers On average, 98.7 percent of complete coding sequences of targeted genes were successfully covered and a 891x coverage was achieved by PGM run. The Proton run produced about 3 times higher coverage than that of PGM. The variant allelic fraction of mutations identified went down to 0.01. We identified JAK2-V617F mutations in tumors that had a negative result for whole exome sequencing method. Interestingly, we identified novel recurrent somatic mutations in rare brain tumors including CBL, encoding the ubiquitin E3 ligase that function as a negative regulator of receptor protein tyrosine kinases including KIT, and BCORL1, a transcriptional corepressor and tumor suppressor. We have systematically compared the performance of AmpliSeq PGM and Proton Sequencers including the throughput, on-target coverage, sensitivity and accuracy for substitutions and indels detection and will present here.

1601F

De novo assembly mapping with single-molecule detection using electronic solid-state detectors. *W.H. Heaton, B. Bready, B. Galvin, P. Goldstein, A. Snider, J. Thompson, J. Oliver.* Nabsys, Providence, RI.

Genome wide physical maps are ideally suited to a variety of applications including scaffolding sequence contigs, genome finishing, structural variant analysis, and haplotype phasing. To date, genomic physical mapping has relied on optical technologies which suffer from cost, speed, and resolution limitations. Even so, the benefit of long range genomic information has been demonstrated by these technologies. To circumvent these limitations, we map DNA by attaching probes to target motifs and localize the positions of these probes using solid-state nanodetectors. DNA passing through the detector causes a drop in the current. Attached probes cause a further transient drop in current denoting their relative locations. We report the first de novo assembly of single-molecule electronic DNA maps. This represents the first scalable technology for physical mapping applications.

1602W

Comparative analysis of six splice-aware alignment and two differential expression assessment tools for RNA-seq data. *J.H. Kim[1], O. Evgrafov[1,2], J. Knowles[1,2], K. Wang[1,2,3].* 1) Zilkha Neurogenetic Institute, University of Southern California, CA; 2) Psychiatry & the Behavioral Sciences, University of Southern California, CA; 3) Preventive Medicine, University of Southern California, CA.

Many bioinformatics software tools have been developed to perform splice-aware alignments, quantitate gene expression levels, or test for differential expression for RNA-Seq data. Despite the wide application of RNA-Seq technologies, it is not well established what are the relative advantages and disadvantages between different analytical approaches. To address this issue, we compared five commonly used alignment tools including Tophat, RNA-STAR, MapSplice, RUM, GSNAP and PerM, and compared two commonly used tools including DESeq and Cufflinks for detecting differential expression levels. The first data set include ~81 million 100bp paired-end reads from a tumor sample. Although four of the tools align more than 97% of reads, TopHat shows a map ratio of only 74%. Speed-wise, RNA-STAR outperformed others with 12 minutes of runtime, while other tools require at least 4 hours in a machine with 12 Intel Xeon 2.66GHz cores. However, RNA-STAR requires more than 20Gb of memory, whereas other tools can run with less than 8Gb memory. We quantitated the gene expression level from each method, compared them to each other and to Illumina HumanHT-12 microarray expression data. Expression data from all tools show more than 0.94 of Spearman's rank correlation between each other, and all have lower correlation with the microarray data (~0.74). The second data set include ~70 million 91-bp pair-end reads on neural stem cells with and without NRXN1 knockdown. We applied DESeq and Cufflinks to evaluate the ability of 6 aligners to identify differentially expressed genes. Both methods showed that all 6 aligners identified knockdown of NRXN1 (P=5.0E-05~1.5E-03 and 3.5E-04~6.9E-04, respectively), and demonstrated that GAPDH has no change in gene expression (P=0.4~0.6 and 0.5~0.6, respectively). However, DESeq identified 54 differentially expressed genes from alignments generated by GSNAP, and 78~120 genes from alignments generated by other 5 aligners, yet Cuffdiff identified 28~47 genes from alignments generated by all aligners, under the same FDR<5% and |log2(Fold change)|>1 thresholds. In summary, all six aligners produce similar alignments and similar correlation to microarray data, but RNA-STAR has a clear speed advantage. Our results may provide a practical guide to readers for selecting aligners, quantifying expression levels and discovering novel splice forms from RNA-Seq data sets.

1603T

Next Generation Sequencing PCR Primer Design Tool for Sanger Sequencing Confirmation. *S. Berosik, M. Wenz, A. Karger, P. Brzoska, F. Hu, X. You, W. Liao.* Life Technologies, South San Francisco, CA.

High quality and high accuracy are characteristic of Sanger re-sequencing projects and are primary reasons that next generation sequencing projects compliment their results by capillary electrophoresis data validation. We have developed an on-line tool called Primer Designer™ to streamline the NGS-to-Sanger sequencing workflow by taking the laborious task of PCR primer design out of the hands of the researcher by providing pre-designed assays for the human exome. The primer design tool has been created to enable scientists using next generation sequencing to quickly confirm variants discovered in their work by providing the means to quickly search, order and receive suitable pre-designed PCR primers for Sanger sequencing. Using the Primer Designer™ tool to design M13-tailed and non-tailed PCR primers for Sanger sequencing we will demonstrate validation of 28-variants across 24-amplicons and 19-genes using the BDD, BDTv1.1 and BDTv3.1 sequencing chemistries on the 3500xl Genetic Analyzer capillary electrophoresis platform.

1604F

Mobile element scanning (ME-Scan) for Alu insertions in families and populations. *D.J. Witherspoon[1], W.S. Watkins[1], M.A. Batzer[2], L.B. Jorde[1].* 1) Dept of Human Genetics, University of Utah, Salt Lake City, UT; 2) Dept of Biological Sciences, Louisiana State University, Baton Rouge, LA.

The human genome carries more than one million copies of the Alu retrotransposon family, the remains of more than 50 million years of activity in our ancestors. Alu retrotransposition and ectopic recombination between Alu copies still generate novel mutations and cause genetic diseases today. A clear understanding of their genomic impact and evolution requires detailed yet comprehensive data on their activity. This is challenging due to the large background of nearly identical sequences in the genome and the modest rate at which new insertions appear in the germline (estimated at 1 per 20 births.) We have developed a targeted high-throughput sequencing technique to efficiently identify novel Alu insertions (Mobile Element Scanning, ME-Scan; Witherspoon et. al BMC Genomics 2010, Witherspoon et al. Genome Res 2013). We have redesigned it to exploit the longer reads and higher capacity of the Illumina HiSeq 2500 platform using standard sequencing parameters. Up to 53 samples can be pooled in a single sequencing library using pairs of custom six-bp indexes. We have applied this design to 290 individuals from diverse populations (HapMap YRI, JPT, TSI, CHB and CEU; Indian Madiga, Mala, Relli and Brahmin; African Alur, Hema, Nande, Luhya and Pygmy) and 85 individuals from 10 parent-offspring pedigrees. The rate of Mendelian inheritance errors (insertions observed in offspring but absent in parents) is <0.5%; no high-confidence de novo Alu insertions have been detected yet. ME-Scan assays of an additional 300 offspring and their parents are underway. Deeper multiplexing and higher sequencing throughput has decreased cost while increasing the coverage (at least 1 million usable mapped reads per individual at cost of ~$55 for reagents and sequencing). In combination with redesigned computational methods, the longer reads and higher coverage increase sensitivity for heterozygous singleton insertions in non-repetitive regions beyond our previously-reported 91%. They also enable reliable identification of novel insertions inside of existing repeated elements. Alu insertion points can be resolved to the exact nucleotide in more than 99% of cases. Strong evidence supporting 5,250 novel Alu insertions was generated. Among the novel singleton insertions, we find 5 embedded in protein-coding exons of METTL20, TUB, TTN, PEG3, and TSC2 (each supported by 200-2400 read pairs.) Twenty additional novel rare insertions are found in noncoding exons.

1605W

Integrative analysis of metabolomics and transcriptomics data: A unified model framework to identify underlying system pathways. *C. Ekstrom[1], K. Brink-Jensen[2].* 1) Biostatistics, University of Copenhagen, Copenhagen, Denmark; 2) Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark.

The abundance of high-dimensional measurements in the form of gene expression and mass spectroscopy calls for models to elucidate the underlying biological system. We propose a statistical method that is applicable to dataset consisting of Liquid Chromatography-Mass Spectroscopy (LC-MS) and gene expression (DNA microarray) measurements from the same samples, to identify genes controlling the production of metabolites. Due to the high dimensionality of both LC-MS and DNA microarray data, dimension reduction and variable selection are key elements of the analysis. Our proposed approach starts by identifying the basis functions ("building blocks") that constitute the output from a mass spectrometry experiment. Subsequently, the weights of these basis functions are related to the observations from the corresponding gene expression data in order to identify which genes are associated with specific patterns seen in the metabolite data. The modeling framework is extremely flexible as well as computationally fast and can accommodate treatment effects and other variables related to the experimental design. We demonstrate that within the proposed framework, genes regulating the production of specific metabolites can be identified correctly unless the variation in the noise is more than twice that of the signal.

1606T

Tissue-specific functional relationship networks effectively predict replicated GWAS discoveries. *C.S. Greene[1], A.K. Wong[2], A. Krishnan[3], D.S. Himmelstein[3,4], O.G. Troyanskaya[2,3].* 1) Department of Genetics, The Geisel School of Medicine at Dartmouth, Hanover, NH; 2) Department of Computer Science, Princeton University, Princeton NJ; 3) The Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton NJ; 4) Biology & Medical Informatics, University of California San Francisco, San Francisco CA.

Modern large-scale human genetics experimental designs take advantage of new technologies that allow for inexpensive high-throughput genotyping to measure genetic variations in hundreds of thousands of individuals. These strategies aim to identify variants in genes associated with important phenotypes. Such unbiased experimental strategies have both benefits, e.g. the opportunity for discovery, but also drawbacks associated with multiple testing considerations. To address these limitations, we developed a method to integrate thousands of publicly available datasets, including gene-expression, protein-protein interaction, and transcription factor binding experiments into networks of tissue-specific functional relationships. We have generated a globally integrated network, as well as networks specific to functional relationships in individual tissues and cell lineages. Such tissue-specific networks encode pathway and functional information in the context of a single tissue or cell lineage. We evaluate the ability of these networks to predict genes associated with disease by using known disease genes from OMIM as seed genes for a state of the art machine learning algorithm. We assess this strategy's ability to predict the replicating discoveries from 198 genome-wide association studies corresponding to 23 diseases. Our evaluation compares the ability to identify genes containing replicated variants over a set of genes matched by chromosomal region. The state of the art learning algorithm is able to effectively use functional relationship networks to prioritize such genes. Furthermore, we observe that tissue-specific networks are better able to identify replicated discoveries than a global network of functional relationships. Such networks also provide the opportunity to prioritize candidates for epistasis analysis by large-scale molecular experiments. We anticipate that this approach will have broad applications for analysis of both main and epistatic effects, including candidate selection for replication and interpretation of overall findings in a molecular context.

1607F

A Semantic Computing Platform to Enable Translational and Clinical Omics. *J. Hirsch, J. Carroll, T. Loeser.* Syapse, Palo Alto, Ca.

To bring omics data from benchtop to point of care, labs and clinics must be able to handle three types of data with very different properties and requirements. The first, biomedical knowledge, is highly complex, continually evolving and comprises millions of concepts and relationships. The second, medical data including clinical health and outcomes records, is temporal, unstructured, and hard to access. The third, omics data such as whole-genome sequence, is structured but voluminous. Attempts to bridge the three have had limited success. No single data architecture allows efficient querying of these types of data. The lack of scalable infrastructure that can integrate complex biomedical knowledge, temporal medical data, and omics data is a barrier to widespread use of omics data in clinical decisionmaking. Syapse has built a data platform that enables capture, modeling, and query of all three data types, along with applications and programming interfaces for seamless integration with lab and clinical workflows. Using a proprietary, semantic layer based on Resource Description Framework (RDF) and related standards, the Syapse platform enables assembly of a dynamic knowledgebase from biomedical ontologies such as SNOMED CT and OMIM as well as customers' internally curated knowledge. Similarly, HL7-formatted medical data can be imported and represented as RDF objects. Lastly, Syapse enables federated queries that associate RDF-represented knowledge with omics data while retaining the benefits of scalable storage and indexing techniques. Biologists and clinicians can access the platform through a rich web application layer that enables role-specific customization at any point in the clinical omics workflow. We will describe how biologists and clinicians use Syapse as the infrastructure of an omics learning healthcare system. Clinical R&D performs data mining queries, e.g. selecting patients who share disease and treatment characteristics to identify associations between omics profiles and clinical outcomes. Organizations update clinical evidence such as variant interpretation or pharmacogenetic associations in a knowledgebase that triggers alerts in affected patient records. At point of care, clinical decision support interfaces present internal or external treatment guidelines based on patients' omics profiles. The Syapse platform is a cloud-based solution that allows labs and clinics to deliver translational and clinical omics with minimal IT resources.

1608W

Establishing a clinical next-generation sequencing information infrastructure. *A. Hoover, M. Kubal.* GenoLogics Life Science Software, Seattle, WA.

In a relatively short period of time next-generation sequencing has established itself as a clinically informative means to pinpoint otherwise difficult to diagnose disease. While the price of sequencing itself has plummeted and many of the challenges surrounding data analysis and storage volumes are actively being managed and/or solved, the issues associated with integrating and interpreting upfront patient medical data, patient sample processing and analysis information, and downstream diagnostics result delivery are still widely unsolved. In this presentation we discuss the issues of data and patient privacy, multisystem integration touch points, data integration and exchange standards, and provide examples of solutions from a variety of laboratories that may be useful to organizations looking to implement a NGS data ecosystem to support diagnostic NGS.

1609T

From big data to smart data: an open-source solution for genome-scale variant data warehousing and discovery. *M.J. Italia[1], B. Ruth[1], M. Sarmady[1], JC. Perin[1], D. Naegely[1], A. Santani[2], M. Dulik[2], NB. Spinner[2,3], ID. Krantz[4], JW. Pennington[1], PS. White[1,4].* 1) Center for Biomedical Informatics, The Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pediatrics, University of Pennsylvania, Philadelphia, PA.

NGS-enabled genome-scale diagnostics and research are generating millions of genomic variants on a routine basis. The growth of this data provides the opportunity to derive information content such as local population allele frequencies and variant differentials between samples, patients, and cohorts. However, NGS technologies also present data management and use challenges that overwhelm traditional data and laboratory workflows. We sought to mitigate these challenges and to advance opportunities by developing Varify, a freely-available, open-source integrated variant data warehouse, knowledge base, and analysis suite. We constructed Varify in collaboration with the NHGRI Clinical Sequencing Exploratory Research program and our clinical laboratories. The Varify application was built using Harvest (harvest.research.chop.edu), an open-source, web-based biomedical application development framework. The data warehouse was deployed on the open source PostgreSQL relational database. Unique variants are combined from all patients into one table. This reduces storage requirements while securely linking variants to patients and patient cohorts. Annotation data include RefSeq transcripts, SIFT, PolyPhen, HGVS nomenclature, HGMD, Alamut, and population frequencies from 1000 genomes & EVS. A variant loading pipeline reads .vcf files annotated by snpEff and links variants to annotations based on genomic coordinates and genotypes. We have scaled Varify to store and deliver interactive real-time filtering on over 100 million unique variants across several thousand patients from exome and whole genome sequencing. Initial capabilities enabled by Varify include patient-specific query and filtration of variants using complex annotation criteria; calculation of allele frequency for custom cohorts; capture of analyst decisions and evidence on pathogenicity; and ad-hoc query of variants across patients based on gene, phenotype, and clinical characteristics. Varify is currently deployed in a CAP-compliant clinical diagnostic environment, an enterprise research instance, and as a consortium resource for NICHD's Newborn Screening Translational Research Network. Varify has added value in both the clinical diagnostic and research setting by transitioning from stand-alone, partially annotated variant files into an integrated resource that adds significant new value to diagnostic and discovery projects.

1610F

Integration of ENCODE datasets to epigenomics analysis using colored De Bruijn graphs. *C. Joly Beauparlant, J. Corbeil, A. Droit.* CHU de Québec, Université Laval, Quebec City, Quebec, Canada.

Recently, the ENCODE consortium produced a wealth of genomic data in order to start cataloging functional elements in the human genome. The challenge is now to integrate this information in our day-to-day analysis of smaller scale experiments to better understand the biological data. This is not a trivial task considering the difficulty to handle all this information and the sheer size of the datasets produced. We propose to use colored De Bruijn graphs as a mean to quickly identify genomic regions from epigenomics experiments that overlap regions previously described by the ENCODE project. This data structure that has successfully been implemented in multiple de novo assemblers can be easily adapted to work with genomic regions instead of raw sequencing data. The elements in the graph can then be colored using external datasets, like those produced by ENCODE. The contigs can then be sorted based on their colors to quickly spot specific combination of genomic elements that overlap. A prototype for this algorithm has been developed in Python to demonstrate the feasibility of the project. Even using an interpreted programming language, multiples samples can be compared in a relatively short amount of time (in a few hours). The algorithm is general enough to be able to use any list of sequences for the coloring of the graph. The next development step is to implement the algorithm as a plugin of the Ray software, a massively distributed genome assembler. The distributed nature of the project will give the scientific community a tool to manage and integrate the flow of data provided by the ENCODE project and other next-generation datasets.

1611W

The Japanese Reference Genome in Human Variation Database. *A. Koike[1], M. Kawashima[2], Y. Suzuki[3], H. Sawai[2], M. Yoshida[1], N. Nishida[2], I. Inoue[4], S. Tsuji[5], K. Tokunaga[2].* 1) Hitachi, Ltd. Cent. Res. Lab., Japan; 2) Dept. Human Genetics, Univ. Tokyo; 3) Dept. Med. Gen. Sci. Univ. Tokyo; 4) Dept. Mol. Life Sci. and Mol. Med., Nat. Inst. Genet; 5) Dept Neurology, Univ. Tokyo.

The HapMap Project and 1000 genome project have revealed genetic differences among populations. Building a reference genome in each population is a crucial step for exploitation of disease-related variations. We have estimated a Japanese reference genome using whole genome sequencing data of five healthy Japanese samples and Japanese samples of 1000 genome PJ, and Affymetrix 6.0, AXIOM, and Omni-2.5 data of more than 400 healthy Japanese samples. The SNPs and small deletions/insertions, and CNVs are calculated after mapping all reads on the hg19 reference genome. The unmapped reads and clipped reads are assembled and insertion positions are identified using clipped reads whose partial sequences are mapped on the hg19 genome and rest parts of sequences are mapped on the assembled sequences. The variation concordances between different platforms are higher than 99.9%. The built reference genome is provided as one of reference genomes of Human Variation Database in Japanese database project.

The Human Variation Database (http://gwas.biosciencedbc.jp/cgi-bin/hvdb/hv_top.cgi) is a repository database to achieve continuous and intensive management of GWAS/NGS data and data-sharing among researchers. This database also accumulates variations extracted from scientific literatures by manual curation to systematize the disease-related variations and improve our understanding of disease mechanisms. Disease-related 20,000 variation entries are currently registered in the genome level. Furthermore, protein-interaction networks with variations can be drawn by a graphic viewer to facilitate understanding of synergetic effects between multiple variations in the same disease and relationships between similar diseases. In this presentation, we overview the influence of calculation method on the Japanese reference genome and population differences of disease-related variations and phenotype-unrelated variations. Acknowledgement: This work was supported by the contract research fund "Database Integration Program" from the Japan Science and Technology Agency.

1612T

***IS_{core}* and *ALT_{rate}*: Inferring human diversity using genome-wide gene-based patterns of nucleotide substitution, insertion and deletion.** *K.Y. Lo[1], Y.R. Lo[2], F.C. Hsiao[2], W.J. Lin[2], Y.D. Chiu[2], C.K. Liu[2], A. Yao[2], C.J. Chen[1], C.Y. Shen[2], C.H. Chen[2].* 1) Genomic Research Center, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei, Taiwan 115; 2) Institute of Biomedical Sciences, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei, Taiwan 115.

As the technologies of the next-generation sequencing are being developed, large scale of genomic researches, such as the 1000 Genome Project, have carried out using either whole-genome and/or exome sequencing to detect genomic variation. The variants identified include single nucleotide polymorphisms (SNPs), insertions/deletions (indels), copy number variations (CNVs), repeat sequences, and rearrangements in the human genome. These variants may have well preserved genetic evidences of human diversity. Previous studies have demonstrated that the genome-wide patterns of SNPs can successfully infer human diversity at the population level and ancestry information at the individual level. However, such an application with the density of SNPs and the indels has still remained unclear. In this study, we propose two gene-based density ratios being able to result in genome-wide patterns for inferring population diversity and individual ancestry: IS_{core}, the ratio of the number of indels to the number of SNPs in each of the UTR and protein coding regions across the human genome, and ALT_{rate}, an alternative measure of the variant density, the total number of the indels and SNPs in each region. We have calculated IS_{core} and ALT_{rate} on the basis of the sequencing data of the1000 Genome Project. In order to efficiently perform computational analyzing and graphical visualization, we have innovated and extended the functionality of the existing public open resources: cluter and Treeview. Our results show that the distribution of IS_{core} varies among genes across the human genome, but with similar trend among the continental groups at most genes. In particular, some genes show relative higher IS_{core} in all ethnicity groups, which suggests that some segments of the human genome contain more indels than SNPs and some prefer SNPs over Indels. Further clustering analysis identifies genes with different distributions of IS_{core} among ethnicity groups and hence differentiates these ethnicity groups. Similar results were found using the alternative ALT_{rate} scores. We also found that the ratios might be biasedly over-estimated in some regions with much lower depth/coverage of sequencing and they should be corrected before analysis. In summary, these two gene-based density ratios, IS_{core} and ALT_{rate}, can be utilized for inferring human diversity and individual ancestry. The ratios may be useful to screen the polymorphism patterns in both patient populations and normal populations.

1613F

Using ontologies to enhance integration and analyses of ENCODE data. *V.S. Malladi[1], J.S. Strattan[1], D.T. Erickson[1], E.T. Chan[1], E.L. Hong[1], G. Barber[2], G. Binkley[1], J. Garcia[2], B.C Hitz[1], D. Karolchik[2], K. Learned[2], B. Lee[2], S. Miyasato[1], G. Moro[2], G.R. Roe[1], K. Rosenbloom[2], L.D. Rowe[1], N.R. Podduturi[1], M. Simison[1], C.A. Sloan[1], E. Weiler[2], W.J. Kent[2], J.M. Cherry[1].* 1) Stanford University, Department of Genetics, 300 Pasteur Dr., Stanford, CA, 94305; 2) Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA, 95064.

The Encyclopedia of DNA Elements (ENCODE), Roadmap Epigenomics (REMC), and modENCODE projects, are large collaborative efforts that aim to provide public resources for the scientific community. The goal of ENCODE, now in its 8th year, is to create a comprehensive catalog of functional elements in the human and mouse genomes. The modENCODE project shares this goal but focuses its investigation on the model organisms C. elegans and D. melanogaster, creating a comparative resource to provide insight into human processes. REMC shares similar assays and tissues to ENCODE while investigating the human epigenome. Though each project has distinct scientific goals, these projects complement the data generated by the ENCODE project.

The ENCODE Data Coordination Center (DCC), which collects all data and metadata generated by the ENCODE project, is currently integrating metadata from modENCODE and REMC. To further enhance the analysis that can be performed within and across these three projects, the DCC has made use of ontologies to annotate these metadata. The DCC has used The Ontology for Biomedical Investigations (OBI;http://obi-ontology.org) to facilitate data identification from various assays sharing similar biological objectives to be searched (e.g. epigenetic modification), allowing an investigator to retrieve all data across the three projects matching various assays (e.g. RRBS, MRE-seq). Data identification across shared anatomy, morphology, and development are accomplished by using and developing cross-references between Cell Type Ontology (http://cellontology.org/) and UBERON (http://uberon.org/). Researchers querying for a biological system (e.g. digestive system) will retrieve data generated from tissues and cells that comprise that system (e.g. colon and epithelial cell of stomach). Here we present our implementation of ontologies to integrate these three projects and how it may be used to identify experiments that match the interests of a researcher. Data from the ENCODE project can be accessed via the ENCODE portal (http://www.encodeproject.org) and the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

1614W

A web-based framework for querying genomic relational databases using SQL. *S.F. Saccone, P.L. Jones.* Department of Psychiatry, Washington University, St Louis, MO.

Genomic databases are an important resource for the study of human genetics and genomic medicine. These databases are commonly accessed through a wealth of visual tools such as genome browsers and other graphical web-based query interfaces. Our BioQ web application (http://bioq.saclab.net) is an SQL-based query tool that provides a flexible framework for querying a given list of heterogeneous genomic features such as chromosomal segments, genes, transcripts, exons and variants across a number of common genomic databases such as the 1000 Genomes Project, dbSNP, Ensembl, VEGA and the European Nucleotide Archive. We introduce new tools in BioQ that allow users to build SQL queries while utilizing various relationships between genomic features such as genes within genomic regions, transcripts associated with genes, exons associated with transcripts and genetic variants within exons. These relationships and their order of execution can be configured by the user in the BioQ web interface in order to define powerful SQL queries across multiple heterogeneous genomic databases. BioQ provides extensive interactive documentation and visualization tools that help the user understand how schemas from different heterogeneous databases are mapped to one another in order to implement genomic feature interrelationships as well as detailed information on data provenance and experimental process flow for data from individual databases. We will demonstrate how this method can be applied to common data integration tasks such as the annotation of results from whole-genome genetic association studies.

1615T

A Systematic Approach to Large-Cohort Biomarker Discovery. *A. Solomon, S. Sanga, A. Vladimirova, T. Klingler.* Station X, 185 Berry Street, Suite 5503, San Francisco, CA 94107.

Nucleic acid sequencing is becoming ubiquitous, with approximately 2,500 high-throughput (Hi-Seq & Proton) instruments, and 3,800 mid-throughput (PGM and MiSeq) instruments placed in labs around the world. The bottleneck is no longer biochemical; we are now able to rapidly generate large amounts of sequence data that is high-quality and relatively low cost. The bottleneck is analysis. A well-powered study involves hundreds or thousands of individual samples, and requires a multi-dimensional analysis that incorporates transcript levels, mutations, structural changes, as well as clinical information such as time-to-event or magnitude of disease. Robust statistical methods are required that understand concepts such as gene regions, pathways, and functional classifications. Just as important is the ability to access and manage projects, and communicate results. Station X has developed a software environment called GenePool™ for the management, analysis and communication of genomic information for cohort-scale biomarker studies. GenePool has been architected to support thousands of whole genomes, exomes, and transcriptomes in either a cloud environment or stored locally. Best-practice statistical analyses methods are available in a context-specific drop-down menu. GenePool is designed to return results in seconds rather than minutes or hours, which enables users to iterate their analyses quickly. Data can be filtered based on statistical ranges, functional categories, and clinical meta-data. Station X has also developed dynamic visualizations that can communicate large data sets and their results in a single view. Collaboration is facilitated through shared permissions, rather than maintaining multiple copies of the same dataset. One of the most valuable public data resources is the Tumor Cancer Genome Atlas. However, it can be difficult to access and the clinical meta-data is difficult to incorporate. Station X has integrated and curated the entire TCGA dataset and has identified highly significant genomic biomarkers in breast and lung tumors that stratify patients based on disease outcome. Two examples will be shown. GenePool enables rapid genomic biomarker discovery through best-practice analyses accessed via an easy-to-use user interface, management of cohort-scale data sets, and communication of results through visualization and shared data. Highly significant biomarkers of patient outcome have been identified through analysis of the TCGA data.

1616F

Beyond the noise: A case study on optimizing computing performance in the cloud. *J. Yeager, G. Manglik.* CliQr Technologies, Sunnyvale, CA.

Life science researchers are continually using ways to leverage cloud computing, and applications now range from web-based apps to enterprise wide big data/compute. Optimal performance of diverse applications requires a broad set of cloud topologies and developers are often finding themselves locked into clouds that fail to meet their computing, network, or storage demands. Inter-cloud porting would seem the answer and vendors now provide cloud portability services, however most are not specialized to the needs of the life science research or require interim migration steps to run applications on each different cloud. Few provide the researcher with dashboards and a benchmarking system to monitor performance across clouds and researchers again end up in a sub-optimal performance environment. CliQr has developed a new method, which has now been applied to the scientific computing industry, which offers both straightforward app migration and cloud portability. Our system provides dynamic performance monitoring for the applications researchers use in the cloud and minimizes the development timeframe, with many standard components such as Python, R, and Perl already available and optimized. We will present the results of a real-world case study where one such life science company has benefitted from the CliQr deployment of their packages. The two packages in question serve similar components of the medical industry yet have significantly varying infrastructure demands. Both applications required stringent security (HIPAA Compliance) and the ability to deploy as Software as a Service (SaaS). The first package needed to be 24/7 accessible with a fast customer interface at a single institution, but had low bandwidth, storage and computation needs; while the second required scaling to huge storage, high bandwidth, global access and moderate computational needs. We will demonstrate how porting to the cloud using the CliQr platform required less effort than performing the implementation across multiple cloud providers, reducing development costs and time to deployment while ensuring maximum performance for each. Using the CliQr built-in analytics the apps were methodically tested for which cloud provided the best performance for each program within hours. We will summarize the general cases as seen by our customers and present guidelines for choosing the best nodes within a cloud, and the best cloud provider, for the most common types of applications.

1617W

Automatic analysis of personal genomes for clinical advisors. *G.E. Zinman, The Clinical Genomic Expert team.* Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA.

The next generation of medicine is envisioned to specifically tailor treatments for patients based on their unique genetic profile and lifestyle. With the rapidly decreasing cost of DNA sequencing and the large investments of medical institutes in digitizing medical records this vision is almost at reach. This new revolution requires the integration of multiple streams of data from different sources - whole patient genome, condition-specific genomic measurements, clinical data from the medical record, and other information including clinical and medical literature - to diagnose diseases, identify relevant treatment options, and monitor response to therapies. This wealth of information must make use of automated analysis to support physicians in making clinical decisions. In order to address these challenges, we recently launched the 'Clinical Genomic Expert' an initiative to build an open-source framework that will utilize machine learning algorithms that can incorporate genetic and clinical phenotypes to model the relationships between complex diseases and genome variation, identify a patient's susceptibility to disease, and predict which therapies might be most effective or cause the fewest side effects. The framework aims to provide an end-to-end solution for medical centers supporting the processing of variant calls as well as data from the Electronic Medical Records (EMR) or patient genomic data. The Clinical Genomic Expert implements several machine learning algorithms that can efficiently solve regression problems on network or tree structures between multiple phenotypestime as well as time series measurement of traits. These methods account for the inherent structured sparsity in all types of data. The framework also aims to provide a graphical user interface for physicians or clinical researchers to communicate the results of the analysis. Ultimately we hope that the initiative provides a platform for sharing disease models that will allow collaborative work between multiple medical centers without compromising private patient information. URL: http://www.genomicexpert.org/.

1618T

Mining human genetic variation with GEMINI - a novel integrative framework for explorative analysis. *U. Paila[1], B. Chapman[2], R. Kirchner[2], A. Quinlan[1].* 1) Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) School of Public Health, Harvard University, Boston, MA.

Advancements in DNA sequencing technologies have given investigators an unprecedented opportunity to rapidly sequence many genomes for studies of the genetic basis of human disease. High-throughput approaches for human genomes have been instrumental in the identification of new DNA variants linked to disease. However, considering the size and complexity of data generated by such studies, it is challenging to isolate relevant variants without integrating genomic 'knowledge' associated with them (e.g. genes, regulatory elements, protein domains, GERP, 1000 Genomes etc.).

Here we introduce GEMINI (GEnome MINIng), a flexible software tool that addresses these analytical challenges by providing a novel framework for exploratory analysis of human genetic variation in the context of crucial genome annotations (e.g., ENCODE, KEGG, OMIM, dbSNP, 1000 Genomes). GEMINI accommodates studies involving thousands of samples, enables the integration of custom annotations, and supports reproducibility by allowing researchers to 'version' their research. By augmenting many large and heterogeneous genome annotations (both coding & non-coding) into a unified database, it provides a powerful tool for data exploration that addresses the demands of large-scale disease research. It allows users to compose complex queries based on sample genotypes (e.g., 'return all loss of function variants that are homozygous for only sample A'), inheritance patterns, and both pre-installed and custom genome annotations to address their specific research needs.

The GEMINI framework facilitates discovery in a broad range of research including studies of the genetic basis of human diseases, personal genomes, unsolved Mendelian disorders, explorations of rare variants in large pedigrees, and genome-wide case-control studies. The portability and flexibility of the tool will allow other developers to leverage GEMINI to create new tools and develop novel approaches to prioritize genetic variation in diverse contexts. We foresee GEMINI as an outstanding future resource for medical and cancer genomics, given the depth and wealth of annotations it carries, as well as those that we expect to add through future development (e.g. novel methods for identifying significantly mutated genes, collapsing methods, network analysis). GEMINI is a self-contained, open source, evolving tool that will benefit the human genetics community.

1619F

A generalized scalable database model for storing and exploring genetic variations detected using sequencing data. *N. Chennagiri, B. Breton, M. Umbarger, P. Saunders, G. Porreca, C. Kennedy.* Good Start Genetics, Cambridge, MA.

With advances in high throughput sequencing, it is common for multiple sequencing centers to detect variants in the human genome and report them through Variant Call Format (VCF) files. There is a growing need for a unified database to store variant information in VCF files from different sources in a way that allows researchers to perform complex allele-, sample- and population-level queries across centers. We describe a data model that unifies the variant information in VCF file from different samples by storing every unique allele in one universal allele table and storing references to associated samples and sample-level meta data. This not only reduces data redundancy but also makes readily available important information like population level allele frequencies among ethnic groups. An added advantage is that annotation on the allele can be performed once and stored in the database and any sample that gets associated with the allele will have this annotation information already available. The metadata in the VCF files are stored as key-value pairs, enabling import of any VCF file with arbitrary tags without changing the underlying data model. We implemented this data model in an Oracle database and loaded VCF files for 15 genes from the Thousand Genomes Project for 607 samples. We also loaded VCF files from 8456 clinical samples for the same genes and performed allele frequency comparisons between the 1000 genome samples and clinical samples using standard SQL. Loading VCF files containing known disease causing variants to the same database enabled us to annotate the clinical samples with disease related information. Additionally, we added variants detected using Sanger sequencing for a set of 13 samples. This allowed us to perform validation of variants detected using other technologies with Sanger as benchmark. The database model and its Oracle implementation allowed us to perform analysis that is otherwise done by processing numerous VCF files using ad-hoc scripts.

1620W

The European Variation Archive (EVA) at the EBI. *I. Lappalainen, D. Spalding, S. Saha, L. Skipper, J. Ameida-King, V. Kumanduri, P. Flicek, J. Paschall.* EMBL-EBI Wellcome Trust, Cambridge, United Kingdom.

The European Bioinformatics Institute (EBI) has launched a new genetic variation service called the European Variation Archive (EVA). The EVA will consolidate, accession and provide new ways to access genetic variation data. This service has been designed to accommodate data flows based on high throughput sequencing technologies and encompasses variation data at all scales from both germline and somatic sources in all species. The EVA uses the same submission and archiving tools as the European Nucleotide Archive (ENA) and the controlled access provision of the European Genome-phenome Archive (EGA) for data requiring such access. For structural and copy number variants, EVA efforts include continued development of the existing Database of Genomic Variants archive (DGVa) with new tools to access genomic variation of all shapes and sizes. The deep integration between these three services will provide complete chain of provenance for variation data from reads, to alignments, and to variants calls. The EVA service accepts variant calls as indexed and compressed VCF files, a common file format exported by the current next generation variation discovery and genotyping pipelines. Archived VCF files will be made directly available through a defined accession based URL for remote access (such as VCFtools) and they can also be accessed as remote data sources by a variety of applications such as genome browsers. The EVA data warehouse layer is under development and based on the NoSQL document centered MongoDB database. This solution manages queries across submitted datasets as well as global data mining functions such as clustering of variants across studies and allele frequency calculation. The use of a document based NoSQL solution to the data warehouse layer benefits from speed of loading, flexible schema changes to keep up with evolving format standards and dramatically simplifies the data architecture as compared with a relational model. The EVA help-desk assists submitters or provides more information about the service at eva-helpdesk@ebi.ac.uk.

1621T

RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *R. Thompson[1], I. Gut[2], K. Bushby[1], E. Heslop[1], L. Johnston[1], D. Taruscio[3], L. Monaco[4], C. Beroud[5], M. Hansson[6], H. Lochmüller[1] on behalf of RD-Connect.* 1) The MRC Centre for Neuromuscular Diseases at Newcastle, Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, United Kingdom; 2) Centro Nacional de Análisis Genómico, Baldiri Reixac, Barcelona, Spain; 3) Istituto Superiore di Sanità, Viale Regina Elena, Roma, Italy; 4) Fondazione Telethon, Piazza Cavour 1, Milan, Italy; 5) Inserm U491 - Génétique Médicale et Développement, Faculté de Médecine de la Timone 13385 Marseille Cedex 05, France; 6) Centre for Research Ethics & Bioethics, Uppsala University, Box 564, SE-751 22 Uppsala, Sweeden.

Despite many examples of excellent practice, rare disease (RD) research is still frequently fragmented by data type and disease. Individual efforts often have little interoperability and almost no systematic connection of detailed clinical information with genetic information, biomaterial availability or research/trial datasets. Linking data at both an individual-patient and whole-cohort level enables researchers to gain a better overview of their disease of interest, while providing access to data from other research groups in a secure fashion allows researchers in multiple institutions to compare results and gain new insights. Funded by the EU Seventh Framework Programme under the International Rare Diseases Research Consortium (IRDiRC), RD-Connect is a global infrastructure project which links databases, registries, biobanks and clinical bioinformatics data used in RD research into a central research resource. RD-Connect's primary objectives are: / Harmonisation and development of common standards for RD patient registries by developing a common registry infrastructure and data elements / Harmonisation and development of common standards and catalogue for RD biobanks that collect and provide standardised, quality-controlled biomaterials for translational research / Development of clinical bioinformatics tools for analysis and integration of molecular and clinical data to discover new disease genes, pathways and therapeutic targets / Development of an integrated platform to host and analyse data from omics research projects / Development of mechanisms for incorporating patient interests and engaging with stakeholders / Development of best ethical practices and a proposal for a regulatory framework for linking medical and personal data related to RD RD-Connect will accept data generated by IRDiRC projects such as EURenOmics, which focuses on causes, diagnostics, biomarkers and disease models for rare kidney disorders, and Neuromics, which uses next generation whole exome sequencing to increase genetic knowledge of rare neurodegenerative and neuromuscular disorders. The 'siloed' nature of individual research efforts is a continued bottleneck for cutting-edge research towards diagnosis and therapy development in RD. RD-Connect aims to unite existing infrastructures and integrate the latest tools in order to create a comprehensive combined omics data, biobanking, data analysis and patient registry platform for RD used by researchers across the world.

1622F

A comparison of imputation quality: combining different GWAS platforms. *E.P.A van Iperen[1,2], G.K Hovingh[3], F.W Asselbergs[2,4], A.H Zwinderman[1].* 1) Department of Clinical Epidemiology, Biostatistics and Bio-informatics, Academic Medical Center, Amsterdam, The Netherlands; 2) Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands; 3) Department of Vascular Medicine, Academic Medical Centre, The Netherlands; 4) Division of Heart and Lungs, Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands.

Background: In the past decade many Genome-wide Association Studies (GWAS) were performed which have discovered many new associations between single-nucleotide polymorphisms (SNPS) and different phenotypes. Imputation methods are widely used in GWAS. They facilitate the association of ungenotyped variants with phenotypes. Imputation methods can also be used to combine and analyse data genotyped on different genotyping arrays. In this study we investigated if there is a difference in imputation quality between two different approaches of combining GWAS data from different genotyping platforms. We will investigate if combining data from different platforms before the actual imputation performs better than combining the data from different platforms after imputation.

Methods: Existing genotype data from the AMC-PAS cohort were used for this study. The samples were genotyped on three different platforms. A total of 706 individuals were genotyped on the MetaboChip (220K SNPs), a total of 757 individuals were genotyped on the 50K gene-centric Human CVD BeadChip (50K SNPs) and a total of 955 individuals was genotyped on the HumanExome chip (240K SNVs). After pre-imputation quality control (QC), Minimac in combination with MaCH was used for the imputation of all samples with the 1000genomes reference panel. All markers with an r2 value of <0.3 were excluded in our post-imputation QC.

Results: All three datasets were carefully matched on strand, SNP ID and genomic coordinates. This resulted in a dataset of 979 unique individuals and a total of unique 258925 markers. Our results suggest that the imputation of the different platforms independently performs slightly better than combining the different platforms before imputation. As an example: In the analysis were the three platforms were imputed separately we imputed successfully 55552 unique SNPs on chromosome 22. In the analysis where the 3 platforms were combined before imputation we successfully imputed 50500 SNPs. We observed similar findings for the other chromosomes.

Conclusions: Based on the preliminary results we concluded that combining the data from three different platforms together after imputation performs better than combining the data of the 3 platforms before imputation.

1623W

RNA-Seq optimization with eQTL gold standards. *S.E. Ellis[1], S. Gupta[1], F.N. Ashar[1], J.S. Bader[1], A.B. West[2], D.E. Arking[1].* 1) Johns Hopkins University School of Medicine, Baltimore, Maryland; 2) University of Alabama School of Medicine, Birmingham, Alabama.

Over the past decade, there has been intense interest in garnering a more complete understanding of genome-wide gene expression, with RNA-Sequencing (RNA-Seq) having emerged as the gold standard approach to studying the human transcriptome. Accordingly, RNA-Seq experiments have been optimized for library preparation, mapping, and gene expression estimation. These methods, however, have revealed weaknesses in the next stages of analysis of differential expression, with results sensitive to systematic sample stratification or, in more extreme cases, to outliers. To address these issues, we turn to expression quantitative trait loci (eQTLs), or biologically meaningful loci at which gene expression is modified by genotype, to optimize RNA-Seq data analysis. Here, we propose a method in which we utilize previously published eQTLs at the center of our data analysis framework in combination with DNA genotypes and RNA-Seq expression data to demonstrate our ability to assess not only the integrity of sequencing data but also the appropriateness of data handling procedures. Using postmortem brain samples from autism-affected cases and controls, we demonstrate that our method successfully identifies sample outliers and low levels of sample contamination that, if not otherwise identified, would interfere with accurate gene expression quantification. After outlier detection, we assess data normalization procedures, demonstrate the biologic validity of covariate inclusion in RNA-Seq analyses, support the inclusion of UTRs in gene annotation, and highlight the importance of removing outlier samples on a gene-by-gene basis. These steps allow for more accurate approximation of gene expression values that can be confidently used in downstream disease-based comparisons ultimately contributing to understanding of genetic variation, gene expression, and disease.

1624T

A graph-based integration with multi-omics data and genomic knowledge for cancer clinical outcome prediction. *D. Kim[1, 2], J. Joung[2,3], K. Sohn[2,4], H. Shin[5], M.D. Ritchie[1], J.H. Kim[2,6].* 1) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA; 2) Seoul National University Biomedical Informatics (SNUBI), Div. of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea; 3) Translational Bioinformatics Lab (TBL), Samsung Genome Institute (SGI), Samsung Medical Center, Seoul, Korea; 4) Department of Information and Computer Engineering, Ajou University, Suwon, Korea; 5) Department of Industrial & Information Systems Engineering, Ajou University, San 5, Wonchun-dong, Yeongtong-gu, 443-749, Suwon, Korea; 6) Systems Biomedical Informatics Research Center, Seoul National University, Seoul 110799, Korea.

Cancer is a complex disease, which can be dysregulated through multiple mechanisms. Therefore, no single level of genomic data fully elucidates tumor behavior since there are many genomic variations within or between levels in a biological system such as copy number variants, DNA methylation, alternative splicing, miRNA regulation, post translational modification, etc. Nowadays, a number of heterogeneous types of data have become more available from the Cancer Genome Atlas (TCGA), generating multiple molecular levels of omics dimensions from genome to phenome. Given multi-levels of genomic data, information from a one level to another may lead to some clues that help to uncover an unknown biological knowledge. Thus, integration of different levels of data can aid in extracting new knowledge by drawing an integrative conclusion from many pieces of information collected from diverse types of genomic data. Previously, we have proposed a graph-based framework that integrates with multi-layers of genomic data, copy number alteration, DNA methylation, gene expression, and miRNA expression, for the cancer clinical outcome prediction. Here, we propose a new graph-based framework for integrating different levels of genomic data and genomic knowledge in order to overcome variability of diagnostic or prognostic predictors and to increase their performances. As a pilot task, we used an ovarian cancer dataset from TCGA for the stage, grade, and survival outcomes. Integrating multi-omics data with genomic knowledge to construct pre-defined features results in higher performance in clinical outcome prediction and higher stability. For the grade outcome as an example, the model with gene expression data performed with an AUC of 0.7866. However, the models of the integration with pathway, Gene Ontology (GO), chromosomal gene-set, and motif gene-set consistently outperformed the model with genomic data only, attaining AUCs of 0.7873, 0.8433, 0.8254, and 0.8179 respectively. Furthermore, incorporation of genomic knowledge offers more interpretable results from the obtained signatures of pre-defined genomic knowledge and thus provides more insight into the complex molecular mechanisms in cancer. With integration of multi-omics data and genomic knowledge, understanding the molecular pathogenesis and underlying biology in cancer is expected to provide better guidance for improved prognostic indicators and effective therapies.

1625F

Phenotype to Genome: Quantitative Trait Loci (QTL) in the Mouse Genome Informatics (MGI) Database. *Y.S. Zhu, D.J. Reed, P. Hale, C.J. Bult, Mouse Genome Informatics database group.* Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME.

The laboratory mouse is an ideal experimental model organism for identifying the genetic basis of complex phenotypes and disease traits. QTL studies in the laboratory mouse have a high degree of concordance with mapped phenotype regions in the human genome but require far fewer resources and time than do mapping studies in human populations. The Mouse Genome Informatics database (MGI; http://www.informatics.jax.org) includes a compendium of over 4,700 published QTL mapping data for the mouse from the peer-reviewed published literature. The database includes descriptions of the QTL mapping experiments along with links to the published literature and, in some cases, links to the underlying genotype and phenotype data maintained by the QTL Archive project (http://www.qtlarchive.org/). The genome coordinates for the markers used to map QTL regions are used to show the genome context of QTLs using MGI's Mouse Genome Browser. The Genome Browser allows users to explore the all of the genome features annotated in the mapped regions. Newly implemented visualization tools at MGI for navigating regions of conserved synteny between mouse and human facilitate comparative genomics approaches to QTL mapping studies. MGI's integration of genome features with their biological attributes (phenotype, expression, function) facilitates the use of the database system to support in silico candidate gene analysis.

1626W

Comprehensive analyses of the functional roles of KAO-NASHI genes in the vertebrate organogenesis using medaka model. *Y. Tonoyama[1,3], A. Shimizu[2], N. Iwakura[3], Y. Shimizu[1], N. Shimizu[1,3].* 1) Advanced Research for Genome Super Power Center, Keio university; 2) Iwate Tohoku Medical Megabank Organization, Iwate Medical University; 3) Nagahama Institute for Bio-Science and Technology.

The human genome project has provided a computer-estimation on 23,000 protein-coding genes in the human genome. However, many of these protein-coding genes are not fully proven for their existence by experimental evidence. In general, proteins with known motifs are readily classified, but substantial numbers of protein have no obvious motifs in their deduced amino acid sequences. We previously designated these genes/proteins without obvious motifs as KAO-NASHI (Face-less) and initiated a project to unveil their face (kao) by employing comparative genomics and knockdown analysis. We extracted 1,000 KAO-NASHI genes from human genome sequence by step-wise filtration with InterPro motif analysis, BLAST homology search and PubMed document search. A small fish medaka (Oryzias latipes) was chosen as an experimental system to knockdown medaka orthologs of human KAO-NASHI genes with morpholino-antisense oligos (MO). As an initial study, we designed MOs to target translation initiation sites of 100 medaka KAO-NASHI genes. When these MOs were injected into fertilized medaka embryos, their morphogenesis at early developmental stages was disturbed and morphological changes were observed at significant rates. Here, we present the data on a few KAO-NASHI genes in which the MO injected-embryos exhibited significant defective organogenesis involved in brain formation and/or early embryonic vascular development, whereas almost all 5-base mismatch MO-injected embryos exhibited normal phenotype throughout developmental process. Immunoblot analyses using appropriate antibodies showed protein bands consistent with the estimated molecular weight. These results suggest that these KAO-NASHI genes definitely encode the functional proteins act in vertebrate organogenesis and they are not non-coding RNAs. Additionally, we performed differential expression analyses between Rx3 homeobox gene-depleted medaka embryos and control embryos using medaka DNA-microarray to seek for KAO-NASHI genes involved in the eye development. As a result, the expression of three KAO-NASHI genes was consistently affected after knockdown of Rx3, which generates Anophthalmia medaka (without eye), implicating that these genes play important roles during development of vertebrate eye. Thus, our approach using medaka model provides the functional information on the human KAO-NASHI genes/proteins during vertebrate organogenesis.

1627T

Rapid and Uniform Whole Exome Libraries from 50 ng of DNA. *M. Andersen[1], K. Rhodes[1], S. Roman[1], A. Broomer[1], C. Van Loy[1], D. Topacio[1], M. Allen[1], S. Rozenzhak[1], G. Liu[2].* 1) Life Technologies, Carlsbad, CA; 2) Life Technologies, Foster City, CA.

Exome sequencing is a powerful and cost-effective way to research variants linked to genetic disorders. Pairing simple fast exome selection and library generation with the Ion Proton™ sequencer allows two exome libraries to be prepared, sequenced, analyzed, and annotated in two days. This new method allows for the production of whole exome selection and libraries from as little as 50 ng of DNA in 6 hours with only 30 minutes hands-on time. Eight barcoded libraries can be prepared simultaneously on a single plate from 2.4 million uniformly amplified target regions. Genomic DNA of HapMap samples NA12878 (a female genome of European ancestry) and NA19240 (a female genome of Yoruba ancestry) that have been deeply sequenced by the 1000 Genomes Project were used to assess the performance of the Ion Proton™ System with the new exome selection and library method. Rapid read mapping, determination of exome coverage metrics, and variant calling were automatically performed using the Torrent Suite™ Software and Torrent Variant Caller. Variant annotations were completed with Ion Reporter™ software. Greater than 90% of 58 million bases in each exome were covered at 20X and >99% covered at 1X, with >98% concordance with known SNPs.

1628F

ZoomMiR, a computational method to predict and screen for variants that disrupt microRNA binding and activity. *A.N. Dubinsky[3,4], L.E. Edsall[2,5], A.R La Spada[1,3,4], T. Gaasterland[1,2], The NEIGHBOR Consortium.* 1) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 2) Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA; 3) Department of Pediatrics, University of California San Diego, La Jolla, CA; 4) Sanford Consortium for Regenerative Medicine, La Jolla, CA; 5) Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, CA.

Purpose: We developed a computer program and assay procedure to detect and validate genome variants that alter microRNA binding sites on spliced RNA transcripts through high-specificity computational and scalable functional assay screens. Background: Whole genome and exome sequencing identify many variants in untranslated regions (UTR) and introns near exons, i.e., on the order of 25,000 - 30,000 per exome. MicroRNAs bind RNA transcripts predominantly in the 3'UTR and, with some frequency, to coding sequences and 5'UTR. With the abundance of human exomes and complete genomes sequenced in recent years, tools and technologies with low false positive rates are needed for accurate, efficient, reliable screening of microRNA binding site disruption by genome variants. Methods and Results: Available algorithms and databases for computationally predicted microRNA binding sites have limitations that hinder their application to assess binding site differences due to genome variation. Existing binding site predictions require substantial computation and, in some methods, simulation of binding to generate predictive scores. For example, many databases that serve computed sites neglect the second arm of microRNA hairpins, which have been observed to be loaded preferentially into the microRNA RISC complex under tissue- or time-specific conditions. To address this shortcoming we developed a novel, efficient process called ZoomMiR which allows detection of microRNA::mRNA seed sites and ranks potential microRNAs for experimental screens. To validate the computational output of ZoomMiR, we developed a novel method to assay either normal or disrupted miRNA seed sites in a reporter gene system. This dual luciferase-based reporter system permits the quantitation of binding of, and hence regulation by, a computationally ranked miRNA. It is sensitive to the effect of the observed alleles and compares it with a completely disrupted seed site and with a perfect seed site. We applied this method to UTR variations found in 420 exomes from patients with primary open angle glaucoma. Importantly, we analyzed a 3'UTR glaucoma risk SNP in CDKN2B and demonstrated that it disrupted binding of a specific miRNA. Conclusion: ZoomMiR provides critical screening functionality to identify genome variants that disrupt microRNA binding sites and can reveal molecular mechanisms for at risk-SNPs found through GWAS.

1629W

Characterization of Human Betacoronavirus 2c EMC/2012 Linked to Acute Respiratory Distress Syndrome in Humans. *C. Olsen, K. Qaadri, P. Meintjes.* Biomatters, 185 Santa Clara St., Suite 101A, San Francisco, CA.

Human betacoronavirus 2c EMC/2012 was isolated from a subject with acute pneumonia in June 2012. This talk will focus on the characterization of the genome sequence, arrangement, and phylogenetic relation with known coronaviruses sampled from around the world including the Middle East respiratory symptom coronavirus (MERS-CoV). The genome contains 30,119 nucleotides and contains at least 10 predicted open reading frames, 9 of which are predicted to be expressed from a nested set of seven subgenomic mRNAs. Phylogenetic analysis of virulence factors of sequenced genomes indicated that HCoV-EMC/2012 is most closely related to species within the genus Betacoronavirus.

1630T

Comparative analysis of whole exome and whole genome DNA sequencing. *M.M. Parker[1], M.A. Taub[2], K.N. Hetrick[3], H. Ling[3], R.A. Mathias[4], J.B. Hetmanski[1], H. Albacha-Hejazi[5], A.F. Scott[6], I. Ruczinski[2], J.E. Bailey-Wilson[7], T.H. Beaty[1].* 1) Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Biostatistics, Johns Hopkins University, Baltimore, MD; 3) Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD; 4) Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, MD; 5) Ibn Al-Nafees Hospital, Damascus; 6) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 7) National Human Genome Research Institute (NHGRI), Baltimore MD 21224.

Background: Four subjects were sequenced with both whole exome (WES) and whole genome sequencing (WGS) as part of larger studies to identify causal variants for oral clefts. We assess the effect of exome capture technology on sequencing coverage and discovery of single nucleotide variants (SNV) using these data. Sequencing: WES was performed at the Center for Inherited Disease Research (Baltimore, MD) using Illumina's HiSeq 2000 sequencer, and exome capture was carried out using Agilent's Human All Exon V3 50Mb kit following manufacturer's guidelines. WGS was performed by Illumina (San Diego, CA) using their HiSeq 2000 sequencer. Sequence reads were aligned to the human reference genome hg19, and we performed local re-alignment around indels and base quality score recalibration using GATK (v.2.4). Variant calling was done using GATK's UnifiedGenotyper. Analyses presented here were restricted to regions on chromosome 20 targeted by Agilent's capture kit including 5,203 regions totaling 1,246,339 bases. Results: Mean target coverage of the 4 WES samples (93.2x) was higher than the mean coverage of the 4 WGS samples (40.6x), however the variance in coverage of WGS was over 3 times smaller than that of WES coverage (standard deviations: WES =90.8,WGS=27.2). With WES, 91.5% of targeted bases were covered at 10x, 78.5% at 30x and 65.5% at 50x. With WGS, 99.8% of targeted bases were covered at 10x, 86.8% at 30x, and 15.24 % at 50x. We assessed the effect of percent GC content in a targeted region on mean read depth. Higher GC content in a targeted region is correlated with lower read depth in both WGS and WES data, however the decline in read depth as a function of percent GC content is steeper in the WES data than in the WGS data. A total of 3,014 SNVs were detected in the targeted regions of chromosome 20. The concordance rate between SNV detection in WES and WGS was 76.6%. On average, concordant variant calls had higher genotype quality and higher sequencing depth than discordant variant calls (median depth: concordant: 41, discordant: 37; median quality: concordant: 99, discordant: 96.5). The majority (94.1%) of all discordant SNVs were called as variants in the WGS data, but not in the WES data. Variants detected only by WES or WGS had greater average depth and coverage in their respective detection platform. Overall, WES achieved higher mean coverage than WGS in our 4 samples, but WGS detected additional variants missed by WES.

1631F

A Practical Evaluation of Next Generation Sequencing & Molecular Cloning Software. *K. Qaadri, C. Olsen, P. Meintjes.* Biomatters, Inc, 185 Clara Street, Suite 101A San Francisco, CA 94107.

Biologists increasingly face the arduous process of assessing and implementing a combination of freeware, commercial software and web services for the management and analysis of high-throughput experiments. Laboratories can spend a remarkable amount of their research budgets on software, data analysis, and data management systems. The National Institutes of Health (NIH) and the National Science Foundation (NSF) have emphasized the need for contemporary software to be well-documented, interoperable, and extensible. Laboratories often invest significant resources for personnel to build bespoke bioinformatics tools or they purchase of commercial software. This can have a marked impact on productivity and ROI because software tools often do not perform the way needed or hidden costs arise unexpectedly because of inefficiencies in the software. Below is a framework that Biomatters developed as a practical evaluation process to assist core facility managers and principal investigators to determine the best tools for DNA/RNA/protein sequence analysis and molecular cloning. The evaluation performed on commercial software packages used six criteria: 1) user interface, 2) data management, 3) data analysis, 4) feature availability, 5) extensibility, and 6) support.

1632W

Detection of rare variants in degraded FFPE samples using HaloPlex PCR. *F. Roos, E. Åström, M. Danielsson, F. Karlsson, M. Isaksson, L. Forsmark, P. Eriksson, H. Johansson.* Agilent Technologies, Uppsala, Sweden.

Detection of rare DNA mutations from heterogeneous formalin-fixed paraffin embedded (FFPE) tissue samples involves several challenges. The sample amounts retrieved from biopsies are often scarce and the DNA is usually degraded and chemically modified as an effect of the formalin fixation. To maximize the chance of success when performing targeted sequencing studies on such samples, it is imperative to have a well-balanced experimental design as well as high efficiency throughout the enrichment and library preparation in order to ensure the mutations are both sampled and sequenced. The quality of the sample and the expected mutation frequency sets requirements on the minimum amount of input DNA that is needed to ensure that all mutations are sampled. Agilent's FFPE-Derived DNA Quality Assessment protocol is based on a simple multiplex PCR of different amplicon lengths that provide a fast and reliable estimation of the sample quality. Based on the outcome of this assay, measures can be taken to optimize the experimental design to ensure maximum probability of finding all variants. These measures include determining the minimum input DNA amount and required sequencing depth. Increased enrichment and library preparation efficiency will further help minimizing the required DNA input amount while maximizing the number of DNA molecules available for sequencing thus increasing the limit of detection for rare variants. HaloPlex PCR can be used to enrich for customized target regions and to prepare libraries for sequencing on Illumina or Ion Torrent in less than 8 hours with target regions ranging from a single gene up to the complete exome. A standard HaloPlex assay captures and PCR amplifies thousands of amplicons of sizes between 100 and 400 bp. For some degraded DNA samples, the availability of the longer fragments will be significantly lower compared to the shorter fragments leading to non-uniform sequencing depths over the targeted regions. We demonstrate that by focusing on targeting shorter amplicons by design as well as targeting both polarities of the genomic DNA molecules, the efficiency of capture is significantly increased leading to higher and more consistent sequencing coverage between samples of different quality. We demonstrate that improved experimental design combined with an increased capture efficiency enable lower DNA input amounts while keeping a high sensitivity for alleles down to 1% in frequency.

1633T

Globus Genomics: Enabling high-throughput analysis and management of NGS data for neurodevelopmental disorders. *D. Sulakhe[1], A. Paciorkowski[3], G. Mirzaa[2], R. Madduri[1], Q. Zhang[2], K. Aldinger[2], J. Bennett[2], L. Lacinski[1], P. Dave[1], W. Dobyns[2].* 1) Computation Institute, University of Chicago, Chicago, IL; 2) Center for Integrative Brain Research, Seattle Children's Research Institute, University of Washington, Seattle, WA; 3) Center for Neural Development and Disease, University of Rochester Medical Center, Rochester, NY.

The availability of low-cost high-throughput sequencing methods in the form of next-generation sequencing is revolutionizing translational research. However, handling such large volumes of sequencing datasets and their analyses introduce great challenges including secure and reliable data transfers, availability of scalable computational resources for analysis, and defining reproducible and reusable analytic workflows. To address these challenges, we have developed a translational research platform called 'Globus Genomics'. Globus Genomics provides an end-to-end solution by integrating state-of-the-art technologies and infrastructures such as Globus Online for data-transfer, Galaxy for workflow management, and AWS for on-demand computational resources. Globus Genomics has established data endpoints at various widely used sequencing centers including the Broad Institute and Perkin-Elmer, to enable electronic transfer of data directly into or between research labs for immediate analysis or collaborations. The platform hosts an enhanced Galaxy instance with hundreds (539 tools) of widely used next-gen sequence analysis tools and many pre-defined best practices pipelines for whole exome, RNA-Seq, or Chip-seq data analysis. Unlimited scalability, enabling analysis of numerous exomes simultaneously, is possible due to the platform's ability to provision on-demand compute clusters on Amazon and submit workflows to that cluster from Galaxy. Globus Genomics allows dynamic tool specific provisioning of Amazon EC2 nodes, thus accommodating a wide range of CPU and memory intensive analytical tools requiring varying compute capabilities that helps in dramatically reducing execution times. This platform has been used successfully at the Dobyns laboratory at the University of Washington to transfer hundreds of exomes amounting to tens of terabytes from Perkin-Elmer sequencing center and local servers to Amazon AWS, allowing data to be ready for analysis a few hours after sequence generation. The use of Globus Genomics successfully cut back on data transfer time from a few weeks to a few hours. Furthermore, the platform has provided a 20X performance improvement in the upstream analysis by allowing the simultaneous analysis of 20 exomes in parallel. The Globus Genomics platform offers a powerful and efficient tool for the transfer and analysis of nextgen data for clinical and research purposes.

1634F

Detecting Contamination in Next Generation DNA Sequencing Libraries. *M.A. Umbarger, M.J. Coyne, E.D. Boyden, G.J. Porreca.* Research and Development, Good Start Genetics, Cambridge, MA.

The massive amount of data generated by next generation DNA sequencing (NGS) systems frequently necessitates sample multiplexing to reduce cost. As a result, sample library construction is often performed in parallel in high-density plate format, thereby increasing the probability of sample cross-contamination. In clinical settings, such contamination can cause important disease-causing mutations to be missed. It is therefore imperative that clinical labs employing sample multiplexing develop strategies that allow contamination to be detected.

We have developed an approach that enables the detection of contaminated samples run through targeted sequencing assays. This approach is based upon the assumption that contamination will skew the allele fractions measured at genomic positions where the contaminated and contaminating samples differ in genotype, and therefore that skewed allele fractions can be used to detect contaminated samples. To implement this approach we have supplemented a set of molecular inversion probes that capture the coding regions of genes associated with recessive genetic disorders with probes designed to capture single nucleotide polymorphisms that have high minor allele frequencies. Thus, for each assayed sample, we measure the allele fractions at a set of commonly variant positions. These measured values are subsequently compared to a set of reference distributions generated from non-contaminated samples to derive a score that indicates the degree to which the observed fractions deviate from the reference distributions. Simulation and spike-in experiments illustrate that these scores are correlated with contamination levels and therefore can be utilized to flag samples that are contaminated at levels relevant to clinical sequencing. Thus, we conclude that our allele-fraction based system represents a promising new tool that can be utilized to discern contamination and therefore improve the accuracy of targeted sequencing assays.

1635W

Cluster solutions to the analysis of large genetic datasets. *D.J. Van Booven[1], A. Mehta[1], R.F. Acosta Lebrigio[1], E.R. Martin[1], R.H. Ulloa[1], J. Zysman[2], G.W. Beecham[1].* 1) John P Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Center for Computational Sciences, University of Miami, Miami, FL.

Recent technological advancements in genomic sequencing have resulted in vast amounts of genetic data being produced at a rate that outgrows traditional computational methods. Even the smallest sequencing projects now produce data that cannot easily be analyzed on desktop machines. To combat the large data problem, several high performance computing (HPC) methods have been developed; these range from 'embarrassingly parallel' implementations of simple algorithms to more current platforms and frameworks - Hadoop/MapReduce, MongoDB, etc. For many applications Hadoop framework has become the model for parallel processing of large datasets. To investigate the utility of these platforms to perform the analysis of next-generation sequencing data (NGS), we are implementing and comparing a variety of these methods to NGS problems. Specifically, we are comparing a traditional HPC approach with the Hadoop framework and with a MongoDB framework. These approaches are applicable to a variety of NGS methods, and can be implemented on local HPC clusters, or in commercial 'cloud' architectures such as Amazon's Web Services. We are focusing on two primary NGS analysis areas: sequence assembly and post-bioinformatics pipeline operations (i.e., variant annotation and genotype-level analyses). Sequence assembly is a multiple node/file comparison that may benefit from this type of technology; group genotype calling may also benefit from this approach. The more repetitive post-pipeline processing jobs such as annotation and statistical analyses likely will benefit from a NoSQL Databases like MongoDB. We do not focus on sequence alignment and basic quality control because they are largely one-off, 'embarrassingly parallel' tasks that would not greatly benefit from existing big data tools. We investigate these approaches on real-world NGS data, including both exome and whole genome samples. The approaches are evaluated based on processing and real time taken to perform tasks, memory usage, long term disk space considerations, as well as an overall cost, and ease of use and implementation. This analysis provides guidance on the implementation of various NGS approaches, and aids in cutting time and computational costs for NGS analysis.

1636T

High-throughput CNV Analysis for Translational Research. *G. Geiss, B. Birditt, J. Gerlach, J. Beechem.* NanoString Technologies, Seattle, WA.

The nCounter® Analysis System from NanoString Technologies is a proven tool for precise, highly multiplexed analysis of DNA, RNA, and miRNA. The system is ideal for large translation studies since it requires no amplification (except for single cell analysis) or cloning, works well on clinically relevant sample types (e.g. archival FFPE tissue), and has an automated workflow. Recent updates and improvements have been focused on delivering higher sample throughput while minimizing the amount of input required. Here, we describe the ability to measure up to 200 CNV regions on 4 DNA samples per imaging lane using unique sets of color-coded molecular barcodes. This increases sample throughput by 4 while maintaining assay performance. In addition, input requirements for all sample types and nCounter assays have been reduced via optimization of the post-hybridization processing protocol. We demonstrate the performance of the enhanced protocols by measuring gene expression and CNV in matched FFPE and frozen tissue from tumor and normal samples. Results demonstrate that a 4-fold increase in sample throughput with lower sample input requirements can be realized without diminishing CNV assay performance.

1637F

Genome Analysis of Individual Cells. *C. Korfhage, E. Fisch, E. Fricke, S. Baedker, U. Deutsch, D. Loeffert.* R&D, QIAGEN GmbH, Hilden, Germany.

DNA sequence analysis and genotyping of biological samples using next-generation sequencing (NGS), microarrays, or real-time PCR is often limited by the small amount of sample available. A single cell comprises only one to four copies of genomic DNA, depending on the organism (haploid or diploid organism) and the cell cycle phase. The DNA amount of a single cell ranges from a few femtograms in bacteria to picograms in mammals. However, a thorough analysis of the genome requires a few hundred nanograms up to micrograms of genomic DNA. Consequently, accurate whole genome amplification (WGA) is required for reliable genetic analysis (e.g., NGS) when genomic DNA is limited, as in the case of single cell DNA. Usually single-cell genomic analysis, and in particular single-cell sequencing suffers from incomplete or biased genome amplification with missing or underestimated sequence information. In order to overcome these typical drawbacks, we developed an easy-to-apply single cell WGA method. This method is based on isothermal multiple displacement amplification (MDA) and consists of an innovative lysis and use of an optimized form of the Phi 29 Polymerase. To prove the method's robustness for single cell amplification, we amplified a variety of human and bacterial single cells, and checked the resulting genome coverage with NGS and qPCR methods. Discussed are experiments on cell-to-cell variations, GC content in comparison to genomic DNA, percentage of genome coverage with respective error rates, and genome-wide real-time PCR analysis. Overall the new method results in the effective lysis of cells, complete DNA denaturation, and reliable amplification of the whole genome of a single cell with high accuracy and minimal amplification bias.

1638W

Subcellular Fractionation Proteomics Is An Indispensable Tool for Polypharmacology Studies for the Identification of Molecular Targets and Transcription Modulator - A Case Study on the Anticancer Auranofin. *S. Tian, F.-M. Siu, Y.M.E. Fung, C.-N. Lok, C.-M. Che.* Department of Chemistry, State Key Laboratory on Synthetic Chemistry and Open Laboratory of Chemical Biology of the Institute of Molecular Technology for Drug Discovery and Synthesis, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China.

In the era of polypharmacology, identify mechanisms of action (MOAs) and molecular targets of drug is highly important. Typical proteomics study determines the protein expression changes in the whole cell lysate of the cancer cells upon drug treatments. This method suffers from problems of low coverage and lack of subcellular location information, and renders identification of molecular targets or MOAs difficult. In the present study, we have developed a subcellular fractionation (SCF) based proteomics approach, and applied to study the MOAs of the anticancer Auranofin (AuRF). Compare to that of the typical proteomics study, SCF approach increase the proteome coverage by 4.8-folds: the number of non-redundant proteins identified using LTQ-Orbitrap Mass Spectrometer by this SCF approach is over 2,700, compared to that of 567 proteins in the typical approach. The enriched proteome revealed HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A Reductase) as the potential inhibitory target for AuRF. This finding was verified by enzyme inhibitory assay with IC_{50} values of 3.5 μM. Using the nuclear sub-proteome, our analysis revealed that $p14^{ARF}$ (gene as *CDKN2A*) was enriched in 2.5 fold as a transcription modulator upon the AuRF treatment. We hypothesized that the up-regulation of $p14^{ARF}$ was associated with E2F-dependent transcription and p53 pathway. This hypothesis was further verified by luciferase reporter system and western blotting assay. Collectively, the subcellular fractionation proteomics is an indispensable tool for polypharmacology studies.

1639T

A Complete System for Next-Generation Sequencing Workflow: from Automated Assay Design and Library Preparation to Superior Sequencing Results. *J. Walker, C. Egidio, X. Wang, K. Datta, R. Ramakrishnan.* Fluidigm, South San Francisco, CA.

Objective: To match the increased throughput and decreased turnaround time associated with next-generation sequencing (NGS) platforms, it is essential to have an efficient method for library preparation. The Access Array System is capable of generating amplified material from up to 480 genomic regions across 48 unique samples per IFC. The unique design of the Access Array IFC permits efficient DNA amplification in multiplex reactions and enables successful amplification of challenging DNA samples, such as those from FFPE tissues. In this study, we demonstrate the flexibility and performance of the Access Array System in producing libraries from both high quality and FFPE-derived DNA samples, with consistent results on multiple Illumina and Ion Torrent NGS platforms. Methods: 301 primer pairs were designed covering all exons of the genes for BRCA1, BRCA2, TP53, EGFR and MET. These primer pairs were used in multiplex format with the Access Array system to target-enrich samples for NGS. Samples used included high quality genomic cell line DNA (50 ng/sample) and FFPE-preserved lung tissue DNA (with an input of 100 copies of DNA). Results: The libraries prepared were tested on the Illumina (MiSeq, HiSeq) and Ion Torrent PGM platforms. Differences in performance were observed for each of the platforms tested. Results across Illumina platforms were very consistent, with between 95% to 97% of all reads mapping to the human genome, as compared to up to 85% on the PGM. Of the reads mapping to the human genome, 99.8% of the reads consistently mapped to the specific target region in all platforms. On the MiSeq, 85-96% of all assays tested generated a number of reads within five-fold from the average (median 95%, variance 5%) compared to 88-98% (median 96%, variance 4%) on the HiSeq. On the PGM 314 chip, 85-98% of all assays tested generated a number of reads within five-fold from the average (median 96%, variance 6%), compared to 96-98% (median 97%, variance 1%) on the PGM 316 chip. Comparable results were obtained with FFPE-derived DNA samples. Conclusion: These results demonstrate that the Access Array System is capable of generating excellent data regardless of sequencing platform. This system is also a powerful tool to obtain DNA sequence information from FFPE samples, with over 85% of the target regions detected, from as little as 100 DNA copies, comparing very favorably with published reports.

1640F

Rapid Preparation of Exome Sequencing Libraries Using HaloPlex. *H. Johansson, B. Skarpås, E. Agné, A. Karlgren, K. Zetterman, M. Isaksson, P. Eriksson, L. Forsmark, F. Roos.* GSD, Agilent, Uppsala, Sweden.

The current amount of data generated from benchtop NGS instruments allows for comprehensive sequencing experiments in a single run. HaloPlex can be used to enrich for customized target regions and to prepare libraries for sequencing on Illumina or Ion Torrent in less than 1.5 days with target regions ranging from a single gene up to the complete exome. The HaloPlex protocol consists of four main steps. In the first step the genomic DNA is fragmented using restriction enzymes followed by the second step where HaloPlex probes are hybridized to targeted fragments. The third step involves DNA ligation of a common primer motif to all targeted fragments which is then used in the final step for multiplex amplification all targeted fragments using one single primer pair. Samples are barcoded during ligation and can thereby be pooled to optimize use of the sequencing capacity on the sequencing chip. To further improve the HaloPlex Exome kit we have developed a new protocol that requires only 50 ng input and that provides coverage of 99% of the target region with less than 5 gigabases of sequencing. Furthermore, we demonstrate a novel protocol where target molecules can be prepared for sequencing without the need for PCR amplification. Using this approach we are able to avoid amplification artifacts by sequencing the actual sample molecules. The content of the exome kit is based on relevant regions from CCDS, Refseq gene, GENCODE, Vega, UCSC, TCGA and MirBase. A total 37 Mb region is targeted in the hybridization reaction which contains over 2 million HaloPlex probes complementary to a redundant set of target molecules covering the ROI. To improve uniformity and thereby reduce the amount of sequencing capacity required, we have focused on efficiency of probes in regions with low coverage. In common with other DNA capture technologies, regions with high GC content (>70%) are less efficiently captured than average GC regions. To compensate for this we have optimized the hybridization conditions and developed a new probe design algorithm were this difference in efficiency is taken into account.

1641W

Use of a Targeted Next Generation Sequencing Approach for the Study of a Cardiac Valve Malformation With Complex Polygenic Heritability. *E.M. Bonachea[1], G.A Zender[1], D. Corsmeier[2], S.M. Fitzgerald-Butt[1], D. Newsom[2], P. White[2,3], V. Garg[1,3], K.L. McBride[1,3].* 1) Center Cardiovascular & Pulmonary Research, Nationwide Childrens Hospital, Columbus, OH; 2) Center for Microbial Pathogenesis, Nationwide Childrens Hospital, Columbus, OH; 3) Dept Pediatrics, Ohio State University.

Purpose. Bicuspid aortic valve (BAV) is the most common type of heart defect with a population prevalence of 1-2%. BAV is highly heritable with a complex polygenic etiology, however few causative genes are known. Current methods of elucidating the genetic contributors have limitations. Sanger sequencing is time and labor intensive, while next-generation sequencing (NGS) remains costly for large cohorts and requires extensive bioinformatics processing. Here, we describe a novel approach to targeted multigene sequencing of pooled samples that allows for focused sequencing of a large number of genes relevant to cardiac development in a well-phenotyped cohort in a cost-effective manner. **Methods.** Candidate genes were selected based on known causal role, mouse models and cardiac developmental biology. A custom capture of 97 candidate genes was designed using the Agilent SureSelect system to probe whole gene intervals. The cohort of 82 unrelated individuals with echo confirmed BAV was combined into 19 pools with a unique overlapping design; a given variant could be attributed to a single individual on the basis of which pools contained the variation. Sequencing was performed on the Illumina HiSeq2000, followed by bioinformatic processing for basecalling, alignment, variant identification, and in-silico analysis of pathogenicity of identified variants. The standard GATK algorithm was compared to CRISP, an algorithm designed to identify variants in pooled samples. Sanger sequencing of one of the genes was used to validate results. **Results.** Targeted capture identified 99 rare exonic variants involving 34 of the 97 candidate genes. GATK calling using an allele frequency threshold of 2.5% compared favorably to the CRISP algorithm. CRISP was able to identify all exonic variants found by Sanger sequencing, while 2 variants were not called by GATK. The approach resulted in a cost savings of $78,750 as compared to whole exome sequencing of individuals or $509,000 as compared to whole genome sequencing. **Conclusion.** Targeted capture allowed for decreased bioinformatics processing by focusing only on those gene pathways thought to be pertinent to the disease under investigation. In addition, the pooling design reduced overall sequencing costs. The CRISP algorithm provided a higher sensitivity than GATK in this design. The data suggest a role for targeted NGS of pooled samples when investigating birth defects which are proposed to have a complex and polygenic inheritance.

1642T

An integrated approach for accurate calling and assessment of structural variations for clinical diagnostics. *H. Lam[1], M. Li[1], S. Chervitz[1], D. Newburger[1], S. Garcia[1], G. Chandratillake[1], M. Clark[1], N. Leng[1], J. Harris[1], M. Pratt[1], M. Snyder[1,2], J. West[1], R. Chen[1].* 1) Personalis, Inc., Menlo Park, Ca; 2) Stanford University, Stanford, Ca.

Structural variants (SVs)~inversions, translocations, deletions, and duplications~play a central role in both rare and common diseases. However, the ability to accurately detect and characterize SVs in genomic sequence data remains poor. We have developed a highly sensitive approach to identify SVs and score these variants based on their predicted likelihood of causing disease. Our approach integrates four orthogonal methods with targeted local reassembly to improve SV detection and determines the genomic context, zygosity, and exact breakpoints of the SVs when possible. SVs are quantitatively scored based on their occurrence within genes known to cause disease using public and proprietary gene-to-disease databases. Their similarity to previously reported SVs is assessed utilizing a comprehensive set of >2 million unique SVs annotated with population frequency, pathogenicity, and associated phenotypes, if known. We determined the accuracy of our SV detection approach by analyzing deletions from both simulated and experimental genomic data. Our approach achieved very high sensitivity while maintaining very low false discovery rate (FDR). With simulated data (46X sequencing coverage), its sensitivity and FDR were 96.3% and 1.4% respectively, compared to an average of 55.6% and 27.6% for the four methods used independently. Experimental data (>60X) from a trio was used to construct a 'gold standard' SV call set from three replicates of the child, vetted by heritability of these variants from the parents. The SVs detected in each replicate were compared to this gold standard. The average sensitivity for SV detection in these data was 96.8% and the FDR was 1.4%, consistent with the results from simulation. We demonstrated the utility of this algorithm for medical interpretation by using our method to identify, annotate and prioritize SVs of various sizes in samples known to harbor pathogenic SVs. For example, a <5kb pathogenic deletion in the ATM gene was detected and ranked as the most likely deleterious SV in an ataxia telangiectasia sample. Utilizing our knowledge-based ranking system for disease variant discovery, we integrated the SVs with SNVs and short Indels to correctly detect biallelic mutations in ATM as causative of the observed clinical phenotype. We have used our approach to correctly identify these causative variants underlying other disease samples and are currently extending our approach to support exome analysis.

1643F

Direct selection of microbiome DNA from host DNA. *E. Yigit[1], G. Feehery[1], S. Oyola[2], B. Langhorst[1], L. Apone[1], P. Liu[1], D. Munafo[1], C. Sumner[1], J. Bybee[1], L. Mazzola[1], F. Stewart[1], M. Quail[2], T. Davis[1], E. Dimalanta[1], S. Pradhan[1].* 1) New England Biolabs, Ipswich, MA., USA; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinton, Cambridge, UK CB10 1SA.

Nucleic-acid based techniques such as hybridization, PCR, qPCR and next generation sequencing offer a rapid and highly sensitive option for direct metagenomic detection from specimens when compared with culture-based techniques. Currently, 16S rRNA gene sequencing is the method of choice for metagenomic studies. However, this approach lacks the sensitivity to detect rare members of the microbial community with divergent target sequences, and is not adequate to detect virulence factors in individual strains. Aside from these inherent limitations of amplification and identification of biological samples, the non-microbial host genome itself may interfere with the detection and diagnosis of pathogens due to the higher percentage of host genomic DNA relative to the target microbiome. Therefore, analyses of a metagenome or microbiome directly from host samples by next generation sequencing or PCR are inefficient, difficult and time consuming.

To address this problem, we have developed a unique method for separating large pieces of host DNA from microbial DNA using a methyl-CpG binding domain fused to the Fc portion of a human antibody heavy chain (MBD2a-Fc). This MBD2a-Fc protein is then bound to a protein A magnetic bead and used to separate methylated host DNA from microbial DNA that is unmethylated or methylated at low levels. As a demonstration of the efficacy of this methodology, DNA samples from various sources, such as human saliva, blood, and *Plasmodium falciparum* contaminated human DNA, were enriched and sequenced on different next generation sequencing platforms. Sequencing data showed that non-microbial host DNA reads decreased 50-fold, corresponding to ~90-95% microbiome DNA in the enriched fraction. Importantly, microbiome diversity after the enrichment remained intact. This simple methodology can be used to analyze entire microbiomes in a cost-effective manner utilizing established next generation sequencing platforms, as well as newer single molecule sequencing technologies.

1644W

In solution HLA capture and high-resolution NGS-based typing method and an automated, integrated analysis framework. *M. Wittig[1], J.A. Anmarkrud[2], M. Forster[1], E. Ellinghaus[1], K. Holm[2], L. Wienbrandt[3], S. Sauer[5], M. Schimmler[3], M. Ziemann[4], S. Görg[4], T.H. Karlsen[2], A. Franke[1].* 1) Christian Albrechts University, Institute of Clinical Molecular Biology, Kiel, Germany; 2) Norwegian Primary Sclerosing Cholangitis Research Center, Medical Department, Oslo University Hospital, Rikshospitalet, Oslo, Norway; 3) Christian Albrechts University, Institute of Technical Informatics, Kiel, Germany; 4) University of Lübeck, Institute of Transfusion Medicine, Lübeck, Germany; 5) Max Plank Institute for Molecular Genetics, Berlin, Germany.

The human leukocyte antigen (HLA) locus contains the most polymorphic genes in the human genome. These genes play an important role in immune response and much is already known about their role in autoimmunity and infectious disease. The classical characterization of these genes is based on Sanger- or next-generation sequencing (NGS) of a limited amplicon repertoire or labeled oligonucleotides, which identify allele-specific sequences. Using these traditional methods, the rate of possible ambiguities is high and requires manual evaluation of the results, which is also an error-prone process. Here, we introduce a highly automated method, which employs comprehensive insolution targeted capturing of the complete classical class I and class II loci in combination with NGS. Our implemented fully automated analysis allowed for the accurate characterization of HLA-A (0.99 allele calling rate), HLA-B (0.99), HLA-C (0.99), HLA-DRB1 (0.98), HLA-DQA1 (0.99), HLA-DQB1 (0.99), HLA-DPA1 (0.98) and HLA-DPB1 (0.96). Including possible ambiguities and manual verification allowed for the exact HLA typing of all our reference samples. The reference sample set comprises 261 samples so far and were derived from the International Histocompatibility Working Group (IHWG) biobank and from another German center. The allelic diversity of the reference sample was maximized before enrichment and NGS. For HLA-A we identified 66 different alleles, for HLA-B 106, HLA-C 49, HLA-DRB1 71, HLA-DQA1 20, HLA-DQB1 17, HLA-DPA1 5 and HLA-DPB1 39, respectively. In summary, our method provides a straight-forward workflow, which is mainly due to the use of in-solution targeted capturing rather than traditional amplicon-based techniques. The fully automated allele calling delivers high confident allele calls and the number of possible ambiguities is drastically reduced compared to traditional typing (e.g. class I with on avg. 2.5 possible alternatives per sample). At this very early stage of development, one technician can characterize 182 samples in one week with high confidence and high resolution (6-8 digits).

1645T

The New Sequencer on the Block: Comparison of Life Technology's Proton Sequencer to an Illumina HiSeq for whole-exome sequencing. *J. Boland, M. Yeager, M. Dean, D. Roberson, J. Mitchell, S. Chanock.* Cancer Genomics Research Laboratory, NCI Frederick, Gaithersburg, MD.

We assessed the performance of the new Life Technologies Proton sequencer by comparing whole exome sequence (WES) data in a CEPH trio (family 1463) to the Illumina HiSeq instrument. The Proton identified 96% of single nucleotide polymorphisms (SNPs) detected by the HiSeq but only 40% of small insertion and deletion variants (indels). Further comparison of the trio data with Complete Genomics sequence data and Illumina SNP microarray genotypes documented high concordance and accurate SNP genotyping of both Proton and Illumina platforms. However, our study underscored the problem of accurate detection of indels for both the Proton and HiSeq platforms.

1646F

New automated systems for size-selection in NGS library construction. *C. Boles, S. Singh, T. Barbera, E. Abrams.* Sage Science Inc., Beverly, MA.

Virtually all genomic library construction methods use some kind of size-selection in order to avoid adapter artifacts, and to place boundaries on the library fragment size. The gold standard method for size-selection is preparative agarose gel electrophoresis, a procedure that is laborious, irreproducible, and difficult to automate. In response to this need, Sage Science introduced its Pippin Prep automated preparative electrophoresis system in 2010. Since that time, three trends in library construction have been notable: 1) library chemistry has become much more efficient, enabling a reduction in the amount of genomic DNA needed for library construction, 2) the need for high quality genomic libraries in clinical research and medicine has increased dramatically, 3) high quality sequence data are easier to obtain if multiple libraries with different insert sizes are sequenced. To address these trends, Sage is introducing two new preparative electrophoresis systems. The first is a higher-throughput version of the original Pippin Prep system. Like the Pippin Prep system, the high-throughput system is designed to select one or two size fractions from a genomic DNA sample. The channels of the new system have been dramatically reduced in size, a change enabled by the reduced DNA requirements of new efficient library protocols. The disposable agarose gel cassette will fractionate 12 samples per cassette, and the instrument can process two cassettes (24 samples) per run. Runtimes are typically 0.5 - 1 hour. The second system is designed to fractionate a single genomic sample into 12 contiguous size fractions. The system uses a two-dimensional process to separate DNA through an agarose column in a first direction, and then move the separated DNA fragments sideways into a linear array of elution modules that are positioned alongside the separation column. Each disposable cassette will process a single sample, and the instrument will process two cassettes per run. The instrument can operate in direct current mode for samples up to mid-single kilobases in size, or in pulsed field mode for samples up to 50 kilobases in size. Performance data for the two systems will be described.

1647W

Further improvements in sequencing technologies on Illumina platforms. *H. Duckworth, J. Weir, G. Smith, J. Betley, P. McInerny, P. McCauley, K. Ahmad, D. Bond, S. Robinson, L. Kangas, M. Fabani, A. Iyer, K. Hall.* Illumina Inc., San Diego, CA.

The application of Illumina's sequencing by synthesis (sbs) technology has proved to be very powerful in next generation sequencing and has also been shown to have a high level of headroom for improvements. Illumina has produced platforms that range from systems that: /can generate 600 Gbases of high quality data from libraries in 11 days (HiSeq2000) /produce a 30x coverage of a human genome in less than 27 hours which includes cluster formation (HiSeq2500) /enable short read (36 bases) data in less than 4 hours or data with paired 250 base reads with the MiSeq system The HiSeq 2500 incorporated features that enable (in the rapid run mode) clusters to be produced on the instrument with the flowcell in situ greatly simplifying the workflow. The instrument can be used as either a high output (>600G) or rapid run (120G in <27 hours) instrument. Recently announced future improvements include extending the read length to 2x250 cycles enabling up to 300G of data to be produced in <60 hours. The MiSeq system has recently been extended to enable paired 250 base sequencing with a high numbers of reads (~7M). It is extremely easy to use, the workflow allowing for cluster formation, paired end and data analysis on instrument and can be combined with a very rapid sample preparation using Nextera technology that takes less than 90 minutes. Run lengths can be selected as needed from single 36 base to paired 250 base reads. The system has also been made easier to use by the incorporation of a Cloud platform called BaseSpace that reduces the need for IT infrastructure. In a further extension of MiSeq the number of reads has been increased to 25M and the read length extended to 2x300 bases enabling yields of >15G. Also together with a new chemistry we have been able to demonstrate even longer read lengths. The use of Illumina's sbs technology is ubiquitous and since its launch in 2006 over 3300 peer cited papers have been published. Here we present some recent advances in the use of Illumina's sbs technology and its applications to both whole genome and exome human sequencing.

1648T

Towards the 24 hour medical genome. *K. Hall[1], J. Weir[1], S. Humphray[1], Z. Kingsbury[1], E. Tsogi[1], P. Smith[1], S. Macarthur[1], E.E. Margulies[1], J. Betley[1], J. Peden[1], N. Miller[2], E. Farrow[2], L. Willig[2], J. Petrikin[2], D. Dinwiddie[2], C. Saunders[2], G. Twist[2], L. Smith[2], S. Soden[2], M. Gibson[2], S. Kingsmore[2].* 1) Illumina Cambridge Ltd., Saffron Walden, Essex, United Kingdom; 2) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO, 64108.

Clinical settings exist where a point-of-care genome sequence is needed, such as in a neonatal intensive care unit (NICU). 5% of US newborns are admitted to a NICU, and 30% of these may benefit from a rapidly interpreted genome sequence for differential diagnosis of a single gene disorder. To influence medical decision making in NICU babies, a molecular diagnosis must be made very rapidly. For most genetic diseases, a definitive (molecular) diagnosis is a prerequisite for specific treatment, prognostic assessment and genetic counseling. Theoretically, genome sequencing can allow an acutely ill baby to be tested for all 3,736 known genetic diseases at once. Currently, causative associations have been determined between the genetic disease phenotypes and 2973 genes. In response, we are developing STAT-seq and have previously reported proof-of-concept for 50-hour return of provisional results in 8 newborns. Essential to this is electronic entry of clinical features by the ordering physician, thereby defining a superset of differential diagnoses and genomic regions to be tested. We report that we have now expanded this clinical genome ordering interface to structured vocabularies of over 5,000 diseases and 8,000 clinical terms. We also report improvements in the overall workflow that shorten return of provisional test results to 24 hours. In particular, methods have been developed to shorten the times for sample preparation, sequencing, alignment and variant calling. We will also demonstrate the use of 2 x 250 read lengths for better alignment in repetitive regions and identification of polynucleotide mutations. Finally, we report clinical utility in 25 cases.

1649F

Droplet digital partitioning improves amplicon coverage of multiplexed assays in NGS library construction. *N.J. Heredia, S. Hodges, S. Cooper, S. Tzonev, D. Skvortsov, E. Hefner.* Bio-Rad Laboratories, Digital Biology Center, 7068 Koll Center Pkwy Ste 401, Pleasanton, CA., 94566.

Droplet Digital PCR reduces biases and improves representation of amplicons in next generation sequencing libraries. The amplicons generated by multiplexing assays are improved when partitioned, compared with standard single tube multiplex NGS methods. Partitioning the sample into droplets reduces biases that arise in PCR such as competition between assays. Custom multiplexed assays as well as standard commercial cancer panels were tested for improvements in read coverage when comparing standard workflows and Droplet Digital PCR. Here we present a facile methodology which easily integrates in to current NGS amplicon library workflows for improvement in reducing amplification bias.

1650W

Accelerated sample prep workflow for target enrichment from low input. *H. Hogrefe[1], B. Arezi[1], A. Belyaev[1], M. Borns[1], M. Corioni[2], J. Fox[1], C. Hansen[1], E. Lin[2], B. Novak[2], C. Pabon[2], B. Rogers[1], D. Roberts[2], F. Useche[2].* 1) Agilent Technologies, La Jolla, CA. 92037; 2) Agilent Technologies, Santa Clara, CA 95051.

We describe an accelerated sample prep workflow that supports whole genome and targeted sequencing from low input amounts of DNA. Genomic DNA libraries are prepared from 50ng DNA using Agilent's novel transposition technology and Herculase II DNA polymerase, which together provide robust yields and greater coverage of AT-rich sequences compared to another vendor's transposon sample prep method. Modified for samples prepped by *in vitro* transposition, SureSelect target enrichment was further accelerated by reducing hybridization time (to 16 hours) and number of pipetting steps. In less than 24 hours, researchers can prepare dual-barcoded (post-capture) libraries for targeted sequencing on the Illumina HiSeq or MiSeq platforms. The automation-friendly workflow supports both whole-genome and targeted sequencing using SureSelect catalog and custom (1kb -24Mb) capture sizes. We routinely achieve >60% on-target, >98% coverage at 1x depth, >75% at 20x depth, and <15% duplicates with the Human All Exon V5 bait set. Performance metrics will be presented for additional libraries (bacterial genomes of varying GC content; 96-plex dual-barcode sequencing; human kinome) and competitor kits using the same input material.

1651T

Performance Evaluation of Bench-top Next Generation Sequencers Using Microdroplet PCR-Based Enrichment for Targeted Sequencing in Patients with Autism Spectrum Disorder. *E. Koshimizu[1], S. Miyatake[1], N. Okamoto[2], M. Nakashima[1], Y. Tsurusaki[1], N. Miyake[1], H. Saitsu[1], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan.

Next-generation sequencing (NGS) combined with enrichment of target genes enables highly efficient and low-cost sequencing of multiple genes for genetic diseases. The aim of this study was to validate the accuracy and sensitivity of our method for comprehensive mutation detection in autism spectrum disorder (ASD). We assessed the performance of the bench-top Ion Torrent PGM and Illumina MiSeq platforms as optimized solutions for mutation detection, using microdroplet PCR-based enrichment of 62 ASD associated genes with RainDance technologies. Ten patients with known mutations were sequenced using NGS to validate the sensitivity of our method. The overall read quality was better with MiSeq, largely because of the increased indel-related error associated with PGM. The sensitivity of SNV detection was similar between the two platforms, suggesting they are both suitable for SNV detection in the human genome. Next, we used these methods to analyze 28 patients with ASD. We validated a total of 57 (PGM) and 30 (MiSeq) SNVs. These SNVs were confirmed by Sanger sequencing, with 21 (PGM) and 22 (MiSeq) shown to be true positives. In contrast, after filtering to exclude rare variants, no indel mutations were detected by either PGM or MiSeq. All 21 SNVs detected by PGM were also detected by MiSeq. We analyzed the ability of each platform to detect variants and found that both platform can identify true variants, but PGM produce more false variant calls. The true positive call rates in the entire coding region were 36.8% (PGM) and 73.3% (MiSeq). Based on web-based prediction software, 72.7% of the detected SNVs (16/22) were seemed pathogenic by either PolyPhen-2 (36.3%; 8/22 SNVs), SIFT (50%; 11/22 SNVs), or MutationTaster (13.5%; 3/22 SNVs). Five out of 28 patients had multiple SNVs. Following the multigenic contribution theory in ASD, these could be associated with the onset or the severity of this disease. Thus, our results support the combination of target gene enrichment and NGS as a valuable molecular method for investigating rare variants in ASD.

1652F

Comparison of enzymes, shear time and capture products to improve whole exome sequencing workflow. *B. Marosy, B. Craig, K. Hetrick, H. Ling, A. Robinson, S. Griffith, J. Romm, K.F. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR is continually seeking new ways to improve its workflow and generate high quality data. A major obstacle in next generation sequencing is the ability to amplify through GC/AT rich regions of the genome. Current methodologies for library preparation and targeted capture rely on PCR to enrich for the prepared libraries. In addition, increased sequencing read length capability has also required changes to the workflow to increase insert sizes. Here we tested two different enzymes (Agilent© Herculase II Fusion DNA Polymerase & Kapa Biosystems HiFi DNA Polymerase) and implemented a different shearing strategy to optimize our current workflow. We also applied these improvements to a NimblegenTM workflow to compare exome captures between the Nimblegen SeqEZCapTM v3 and the Agilent SureSelectTM XT Human All Exon v4. Four unrelated HapMap samples with DNA inputs of 500ng were processed in parallel using 1)Agilent SureSelect XT library prep (XT) & Herculase enzyme; 2) XT library prep & HiFi enzyme; 3) XT library prep & HiFi Enzyme w/shearing modifications; 4)Kapa Library prep & HiFi enzyme w/shearing modifications. Library methods 1, 2 and 3 were hybridized for 24hrs using the Agilent v4 exome and library method 4 was hybridized for 72hrs using the Nimblegen v3 exome. All libraries were clustered using the Illumina® cBOTTM Cluster Generation system. One hundred bp paired end sequencing was performed on the Illumina HiSeqTM2000 platform. Sequencing data was downsampled to ensure an equal comparison between methods. In addition, a product-neutral UCSC exon bed file was used as the 'on target' bed file to more accurately compare the capture products. Preliminary sequencing data analysis of methods 1 and 2 yielded 92.4% and 92.2% of targeted bases covered >= 8x; a molecular duplication rate of 5.1% and 4.8%; mean insert size of 175bp and 193bp; respectively. GC plots indicated improvement when using the HiFi enzyme for normalized coverage and base quality at GC/AT rich regions. Preliminary sequencing data analysis comparing the two capture methods (3 and 4) yielded 92.7% and 94.2% of targeted bases covered at >=8x; 98.5% and 98.4% in dbSNP; 3.11 and 3.13 TiTv ratio; 96.3% and 96.3% sensitivity; 49x and 36x mean bait coverage; respectively.

1653W

Simplified and improved methods for preparing high quality genomic libraries for use on Illumina® sequencing systems. *V.P. Smith, S.J. Humphray, R.M. Sanches-Kuiper.* Illumina UK Ltd, Nr Saffron Walden, Essex, United Kingdom.

Simple, affordable and reliable methods for the preparation of high quality genomic libraries are essential for cost-effective whole genome sequencing (WGS). When selecting a library preparation method for WGS, key performance metrics to consider include (i) simplicity and duration of the workflow; (ii) DNA input requirement; (iii) yield of the final library; (iv) uniformity of coverage and (v) library diversity or complexity. We will present results obtained with the latest versions of Illumina's TruSeq® DNA Sample Prep workflows, TruSeq PCR-free and TruSeq Nano, used in combination with the HiSeq® 2500 and MiSeq® sequencing systems. The PCR-free method produces robust yields from moderate DNA input quantities and generates very high quality genomic libraries with sufficient diversity to support deep sequencing of human and other large genomes. Elimination of PCR-induced biases results in libraries that exhibit uniform coverage across the full range of GC contents . The TruSeq Nano workflow produces relatively unbiased, high yield and high diversity libraries from large and complex genomes starting with as little as 100ng of DNA. These workflows are gel-free, straightforward to execute, automation-compatible and can easily be completed in less than a day, making them ideally suited to the fast turnaround times of the HiSeq 2500 and MiSeq systems. The difficulty of meeting the performance requirements described above is increased substantially when working with degraded DNA such as that extracted from FFPE samples. We will also report on the development of methods for obtaining high quality WGS datasets from practical quantities of FFPE DNA.

1654T

Successful whole-exome sequencing of genomic DNA isolated from preserved mixed-placental tissues. *M.K. Veerapen[1, 2], L. Pelaez[3], M.M. Rodriguez[3], J.E. Potter[4], E. Rampersaud[1], O.A. Bodamer[1, 2].* 1) Hussman Institute of Human Genetics, University of Miami, Miami, FL; 2) Department of Biochemistry and Molecular Biology, University of Miami, Miami, FL; 3) Department of Pathology, University of Miami, Miami, FL; 4) Department of Obstetrics and Gynecology, University of Miami, Miami, FL.

Studies on the genetics of obstetrics and gynaecology have previously focused on utilizing fresh placental tissue (PT) for downstream genetic analysis. However, fresh PT would rarely be histopathologically evaluated for deep-phenotyping purposes and cannot be stored for long-term. Therefore, the use of fixed and paraffin embedded PT would be of interest due to the increasing importance for retrospective genetic studies. With the technological advancement of DNA isolation methods and next-generation sequencing, this could increase the efficiency in variant identification from the preserved PT samples. Due to the fact that the placenta contains a mix of maternal and fetal tissue, the experimental and variant identification challenges are significantly increased. We propose to isolate DNA from preserved PT and successfully perform variant identification. We obtained deidentified PT: non-crosslinking UMFix fixed and paraffin embedded (UMFFPE) foetal membrane (FM) and umbilical cord (UC) samples of mother-foetal (son) dyad pairs from Jackson Memorial Hospital, Miami, FL. The UMFFPE-UC 20 µm curls, containing only foetal surface, were processed for DNA isolation. The UMFFPE-FM containing both maternal-foetal (son) surfaces were separated using laser-capture microdissection (LCM) and DNA isolation. Genomic DNA (gDNA) obtained from UMFFPE-LCM-FM and UMFFPE-UC were then processed for library construction, capture and hybridization using the Roche Nimblegen V3 and Agilent SureSelect 50 Mb V4 respectively for sequencing. Paired-end reads were mapped and aligned with the Burrows-Wheeler aligner according to the Genome Analyzer Tool Kit (GATK) v2.5 pipeline. We successfully isolated gDNA from both UMFFPE-LCM-FM and UMFFPE-UC samples of good gDNA yields and quality. The UMFFPE-LCM-FM samples tested negative for foetal contamination using a PCR-based method. Preliminary sequencing results have shown a capture of an average of 78%; and a read of depth of 71%; at 20X. To the best of our knowledge, this is the first report of the successful whole-exome sequencing and gDNA isolation from preserved PT samples. Potentially, this protocol can be utilized for retrospective genetic studies related to obstetric conditions such as preterm births, intrauterine growth restrictions, intrauterine foetal demise, preeclampsia and chorioamnionitis.

1655F

Nextera® Rapid Capture: The Fastest In-Solution Capture Assay for Whole Exome and Custom Targeted Enrichment. *M. Virata, S. Snow, S. Melnyk, L. Galver, E. Allen, S. Kumar, R. Shen, S. de Rozieres.* Illumina, San Diego, CA.

Targeted resequencing provides a cost-effective approach to identify variants by isolating specific genomic regions of interest within a whole genome library. However, high-throughput sample processing and fast time-to-answer can be affected by lengthy library preparation and enrichment procedures. Here we describe the Nextera Rapid Capture Enrichment assay that allows researchers to go from genomic DNA (gDNA) to targeted enrichment data in less than 2.5 days when accompanied by sequencing on a HiSeq2500. By requiring only 18 hours of hybridization time, a single enrichment wash solution, and rapid fluorometric sample quantification procedures, sequencing ready samples can be prepared in 1.5 days with no more than 5 hours of hands-on time. We present data, from a familial CEPH pedigree trio (Mother NA12892, Father NA12891, Child NA12878), showing how the Nextera Rapid Capture Enrichment assay can be used to target ~37 Mb of focused exonic content that covers at least 98% of coding sequences from RefSeq, CCDS, Ensembl, and GENCODE. Despite the low 50 ng gDNA requirement, highly uniform and specific data is generated from only 4 Gb of sequencing (>80% of targeted bases covered by at least 10x). Variants identified in these samples show a high degree of overlap with standard curated databases like dbSNP (98% for SNPs and 70% for Indels). The sensitivity of these variant calls was assessed by comparing these variants to WGS datasets. Similarly, comprehensive coverage of ~62Mb of exons, UTRs, and miRNA targets is shown to require up to 8 Gb of sequencing. Finally, enhancements to Illumina's web-based design tool, DesignStudio, are discussed which enable new design functionality for custom panels utilizing the Nextera Rapid Capture Enrichment Assay.

1656W

Comparison of conventional and PCR-free library preparation methods for next generation sequencing. *P.D. Witmer, B. Marosy, B. Craig, K. Hetrick, A. Robinson, K. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University School of Medicine, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to human disease. As part of our continuing effort to implement emerging NGS technologies, we now offer whole genome sequencing (WGS) as a complement to whole exome sequencing (WES) within the suite of services available at CIDR. Furthering our aim to evaluate improvements in methodology, we have been focusing on DNA library preparation methods for sequencing on the Illumina HiSeq2500 platform. Standard library preparation methods include a PCR enrichment step prior to cluster generation. Biases inherent in PCR amplification result in uneven read coverage and increase the numbers of duplicate fragments present in the library. For the HiSeq instruments in particular, coverage of sequence reads is known to be notably lower in GC rich regions, which can affect mapping quality and variant calling. To compare library preparation methods that employ PCR directly to methods that do not, we constructed libraries for NGS with DNA from a well characterized HapMap Trio (NA12878, NA12891 & NA12892) using a conventional DNA library preparation kit from Kapa Biosystems and the TruSeq DNA PCR-Free Sample Preparation Kit available from Illumina. Our QC metrics indicate that library size increased using the PCR-free kit (2.5 fold), improving the complexity of the sample. We also observed significant decreases in read pair duplicates (2 fold) and chimeric molecules (3 fold) compared with samples enriched by amplification. GC plots demonstrate more uniform coverage for PCR-free libraries albeit with a slight reduction in the total number of bases covered above 20X. For mean coverage, Ti/Tv, homozygous concordance and sensitivity to SNP genotyping arrays, there were no appreciable differences between the two methods. However, for PCR-free libraries, there were notable increases in unmapped reads (3 fold), unpaired read duplicates (2 fold) and the percentage of reads containing adapter sequences, which likely reflect broken mate pairs and a smaller insert size (350 vs. 460 bp). Although the omission of PCR shortens the work flow for library preparation, qPCR is required for quantitation. This added step, coupled with the large amount of DNA starting material needed (1ug), could offset the advantages of PCR-free methods when processing clinical samples for WGS.

1657T

A complete workflow from sample preparation to analysis using SureSelect target enrichment system for Ion Proton semiconductor sequencing. *K. Jeong[1], J. Ong[2], E. Agne[3], F. Karlsson[3], A. Ashutosh[1], C. Cocq[1], F. Useche[1], J. Ghosh[1], H. Johansson[3], S. Happe[2], D. Roberts[1].* 1) Agilent Technologies, Santa Clara, CA; 2) Agilent Technologies, Cedar Creek, TX; 3) Agilent Technologies, Uppsala, Sweden.

Whole exome or targeted sequencing for protein-coding regions has provided a cost effective way to identify common and rare polymorphisms that are associated with Mendelian disorders and complex diseases. With increased capacity of semiconductor sequencing, highly multiplexed samples can be studied in a single sequencing run. However, a complete workflow processing raw DNA samples to identify DNA variants in target regions is not easily accessible. Here we describe an analysis workflow to study multiplexed samples in semiconductor sequencing for several target sizes: 50Mb (Human All Exon), 3.2Mb (Human Kinome) and a 1Mb custom design. The workflow includes library preparation, SureSelect target enrichment, semiconductor sequencing, and variant calling with SureCall software (beta version). Improved and simplified steps for library preparation and target enrichment maximize multiplexing and produce consistent results in the Ion Proton sequencer. Sequencing output can be easily analyzed, visualized and summarized in a report with SureCall which is optimized for use with Agilent's target enrichment system. We demonstrate high capture efficiency, uniformity, and reproducibility of enrichment. The results from different capture sizes show comparable high performance regardless of various targeted regions. The combination of efficient target enrichment system, semiconductor sequencing, and SureCall software provides a fast and convenient tool to assess DNA variants in genomic regions of interest.

1658F

Accurate Modeling of Indel Genotype Likelihoods from Sequencing Data. *A. Tan, HM. Kang, GR. Abecasis.* University of Michigan, Ann Arbor, MI.

Short indels are the second most common class of genetic variants. Next generation sequence data and alignment-based indel calling algorithms allow efficient discovery and genotyping across a large number of samples, while retaining high power to detect rare variants. However, the current state-of-the-art methods for analysis of indels are not as accurate as their counterparts for analysis of SNPs. One of the reasons is that indels vary much in length and sequence complexity and thus require more careful handling to cater for the variety of forms. Here we address the problem of bias towards reference alleles when modeling the likelihood of short indels. We observed that in widely used indel calling algorithms, genotype likelihoods lean preferentially towards reference alleles, and the bias becomes larger as allele length increases for indels. For example, for high quality 5-bp indel sites, the proportion of reference alleles in aligned reads from individuals with heterozygous genotypes ranges from 53% to 73%, compared to the expected value of 50%. In our improved algorithm, we carefully construct appropriate local haplotypes with respect to each candidate allele and use a local hidden Markov model that keeps track of flanking sequences and insertion/deletion fragments to qualify and quantify the evidence for each allele. Our procedure allows us to attain an allele balance close to 50% and reduce genotype likelihood discordance substantially compared to analyses using GATK, samtools, and other callers; particularly for longer alleles. As a result, our model also provides us with improved filtering of variants as it results in a more robust distribution of variant features against the confounding due to allele length. Our indel pipeline performs variant calling, genotyping, and filtering while modeling contamination and systematic alignment artifacts. The complete process is incorporated in the updated version of GotCloud pipeline. Along with other independent efforts to improve the existing indel calling algorithms, we anticipate that our method will contribute to increasing our understanding of the genetic risk factors beyond single nucleotide variants.

1659W

Dissecting Gene Regulation Networks at Single-Cell Resolution. *X. Wang, J. Shuga, P. Chen, J. Wang, S. Weaver, N. Li, L. Szpankowski, B. Fowler, A. Leyrat, G. Sun, M. Unger, J.A. West.* R&D, Fluidigm Inc, 7000 Shoreline Court, Suite 100 South San Francisco, CA 94080.

Gene regulation is central to the life of a cell. Sophisticated programs of gene expression trigger developmental pathways in response to environmental stimuli and cell disorders. Even in nominally homogeneous cell populations, cell-to-cell differences are observed. These differences represent different gene expression patterns driven by modular gene regulation. However, due to technical limitations, biologists traditionally investigate gene regulation by utilizing bulk-scale methods to measure average gene expression values in populations of cells. Therefore, the ensemble gene expression profile of a population may mask the behavior of gene regulation in any individual cell. The Fluidigm C1™ Single-Cell Auto Prep System allows researchers to investigate whole transcriptome profiling of individual cells with a simple, automated workflow. In this study, we used the C1 System to prepare mRNA-Seq libraries from >10 different cell types (e.g. HL60, iPS, NPC, primary neuronal cells), with an average sample size of 24 single cells per cell type, and we then conducted deep sequencing on the Illumina HiSeq platform. Sequence depth of 5-10 million reads per cell was obtained and systematically analyzed for gene expression profile correlation at the single-cell level across diverse cell types. We found that extremely deep sequencing (10-30 million reads) was not required to conduct differential gene expression analysis, and that even with read depths of only 1 to 5 million reads, we were able to develop robust gene expression profiles from single cells. Our results indicate that many highly correlated gene expression profiles can only be observed at the single-cell level and are masked in bulk data. Furthermore, the data revealed that gene correlations in single cells are cell-type dependent while highly correlated genes in single cells enrich for relevant biological pathways. Using the validated transcription factor binding site (TFBS) data from the ENCODE project, we further confirmed that some of these highly correlated genes share common TFBS in their promoter regions. Our study suggests gene co-profiling analysis at single-cell resolution is a powerful tool that can help biologists precisely dissect gene regulation networks and better understand cellular heterogeneity.

1660T

Efficient and Accurate Whole-Genome Human Phasing. *T. Blauwkamp, V. Kuleshov, D. Pushkarev, S. Swamy, A. Halpern, K. Singh, R. Sit, A. Granat, J. Zhang, A. Young, K. Kuhn, R. Shen, M. Kertesz, G. Smith.* Technology Development, Illumina Inc., San Diego, CA.

High throughput DNA sequencing allows whole human genomes to be resequenced rapidly and inexpensively producing a comprehensive list of variants relative to the reference genome. However, short read sequencing technologies are limited in their ability to determine phasing information, thus resulting in heterozygous calls being represented as the average of the maternal and paternal chromosomes. Phasing information is of critical importance to personal medicine as it provides a better linkage between genotype and phenotype, permitting new advances in our understanding of compound heterozygote linked diseases, pharmacogenomics, HLA typing, and prenatal genome sequencing. Here, we describe a new sample prep method that enables whole human genome haplotyping at high accuracy using only 30Gb of sequence data. Genomic DNA was fragmented into ~10Kb fragments, end repaired, and ligated to adapters. Hundreds of aliquots with approximately 50MB of DNA in each were amplified, fragmented and converted into individual shotgun libraries. The pooled libraries were sequenced in a single lane of a HiSeq2500 at 2×100bp to generate ~30Gb of sequence. The resulting sequence information was analyzed to obtain a set of long blocks of ~10Kb, covering multiple heterozygous SNPs, allowing phasing of these SNPs relative to each other. An HMM-based phasing algorithm was used to compute the most likely phase and confidence intervals based on the observed coverage and sequencer quality scores. Phasing of those blocks relative to each other was done by another HMM-based algorithm which uses a panel of previously phased genomes. Comparing our results with phase information inferred by transmission from the parents, we found that over 98% of heterozygous SNPs were phased within long blocks (N50=500kb) at a switch error rate below 1 switch per megabase of phased sequence. We present results obtained from multiple cell lines and human samples. This new library prep method and data analysis pipeline enables whole human genome phasing with only 30Gb of raw sequence, which represents only ~30% more sequencing than current 30x baseline run for human sequencing. Compared to other published reports, this method is capable of phasing a greater fraction of SNPS with ~75% less sequencing. Coupling our higher percentage of SNPs phased with high accuracy and the lowest sequencing requirement, this new technology is the most affordable approach to generating completely phased whole human genomes.

1661F

A Complete Work Flow for Single Cell Next Generation Sequencing: from Flow Sorting, Library Preparation, to Sequencing Analysis. *X.Y. Wang[1], R. Li[2], I. Khrebtukoba[2], J. Crane[1], C. Mason[1], G.P. Schroth[2].* 1) BD Biosciences, San Jose, CA; 2) Illumina Inc, Hayward, CA.

Single cell genome-wide transcriptome analysis is a powerful tool to define cell-to-cell variations, either within a cell population or among different cell populations. However, the accurate delivery of single cells or small numbers of cells with distinct phenotypic profiles to perform next generation sequencing remains a challenge. BD fluorescence activated cell sorting (FACS) technology is a powerful tool to separate single cells based on their unique cell surface marker combinations. Here we applied BD FACS technology to deliver cells for next generation sequencing. Single or small numbers (5, 10, 20, 40, 80) of monocytes, T lymphocytes, B lymphocytes, and NK cells were sorted from normal blood into 96-well PCR plates, based on their specific cell surface markers. Then the cells were hypotonically lysed, mRNAs were reverse transcribed and PCR amplified, and genome-wide transcriptome analyses were performed on each cell by using Illumina robust Smart-Seq technology. Good sequencing libraries were successfully generated from all cell types using Smart-Seq, down to single cell levels. Differential gene expression patterns were reproducibly observed. The in-depth transcriptome analysis showed that although gene expression analysis of single cells had increased noise, hundreds of differentially expressed genes were identified, and different cell types gave cell specific gene expression patterns. These results demonstrate that the combination of flow cytometry and next generation sequencing provides a powerful tool for genome-wide gene expression analysis.

1662W

Improved Computational and Experimental Methods for Targeted DNA Sequencing with Molecular Inversion Probes. *E.A. Boyle, B.J. O'Roak, B.K. Martin, A. Kumar, A.P. Lewis, J. Shendure.* Genome Sciences, University of Washington, Seattle, WA.

Robust target enrichment strategies for next-generation sequencing are necessary for cost-effective clinical genomics as well as for validating the contribution of rare or de novo variants in candidate genes to complex diseases. Hybridization-based capture technologies are limited in that they scale poorly to very large numbers of samples and routinely fail to capture about 5% of the exome, typically in regions of high GC content. Recent studies of sporadic cases of autism spectrum disorder demonstrated the utility of Molecular Inversion Probes (MIPs) for identifying genes in which recurrent de novo mutations contribute to disease risk (O'Roak *et al.*, Science, 2012). However, in this project and others, the utility of MIPs for targeted gene sequencing has been limited by design constraints and poorly captured regions. We sought to address these challenges by developing a flexible new approach to MIP design that better explores the search space of possible designs *in silico* to reduce dropout of GC-rich targets. We first synthesized and tested 12,000 MIPs whose designs explored a broad range of parameter space, and then used these results to empirically train a logistic model for predicting MIP performance. On an unrelated test set, MIP scores from the predictive model were highly correlated with the relative capture efficiencies of individual MIPs (Spearman rank correlation of 0.67). Head-to-head comparisons against previous MIP sets showed improved coverage of poorly captured genes, boosting the proportion of sites meeting per-sample coverage thresholds by as much as 20%, with median per-base coverage rising from 98.4% to over 99.9%. Furthermore, comparison to coverage levels reported on the Exome Variant Server showed comparable or, in some cases, superior coverage across targeted regions. Independent of MIP design, a series of optimizations of the capture protocol improved the relative efficiency of capture of high GC targets tenfold or greater. This work demonstrates that improved modeling of the features affecting MIP capture can be used to successfully predict relative MIP capture efficiencies and, coupled with optimizations of the capture protocol, have the potential to push coverage of target sequences close to 100%. Further improvements to MIP methodology are expected to enable the application of MIP-based sequencing to clinical diagnostics.

1663T

A novel In situ isothermal amplification method for next generation sequencing. *Z. Ma, R. lee, S. Goyals, J. Erikson, K. Lao.* Life Technologies Inc., South San Francisco, CA.

We report a novel alternative approach to generate immobilized monoclonal templates for next generation sequencing applications. This in situ isothermal amplification method can generate over a billion of sub-micrometer size colonies in one lane of 5500 flowchip without multiple cycles. We also demonstrate an alternative paired-end sequencing method using an interstrand DNA photo cross-linking reagent to covalently link the complementary strands of the original templates onto the solid surface. We believe that these approaches could potentially be used on other sequencing platforms and simplify the working flow for clinical genomic applications.

1664F

Enhanced solution based target enrichment using oligonucleotide probes and a novel composition of blocking oligonucleotides. *S.D. Rose, K.L. Popp, C.C. Locklear, A.N. Dvorak.* Product Development, Integrated DNA Technologies, Coralville, IA.

Targeted pull-down enrichment using oligonucleotide probes has become an important tool for those researchers interested in isolating a smaller subset of loci defining a particular pathway or disease state. Using hybridization probe-based enrichment has its advantages over PCR based methods. Probes that are long enough, approximately 120 nucleotides in length are impervious to smaller base variations which allows for discovery of undocumented SNPs. Unlike PCR amplification, deletions or rearrangements can be detected without any additional information about chromosomal location. The improved on target capture percentage through the use of blocking oligonucleotides as part of the hybridization step has been previously demonstrated. A new generation of modified blocking oligonucleotides shows 10-25% increase of 'on target captures' using libraries made with TruSeq™ adapters. Furthermore these new blockers have universal barcoded domains alleviating the need for individual barcode specific blockers.

1665W

A Fast Solution for NGS Library Prep with Low Nanogram DNA Input, for Multiple Sequencing Platforms. *F.J. Stewart, P. Liu, G.J.S. Lohman, E. Cantor, B.W. Langhorst, E. Yigit, L.M. Apone, D.B. Munafo, C. Sumner, D. Rodriguez, V. Panchapakesa, J. Bybee, L.M. Mazzola, T.C. Evans, N.M. Nichols, E.T. Dimalanta, T.B. Davis.* New England Biolabs, Inc., Ipswich, MA.

Next Generation Sequencing (NGS) has significantly impacted human genetics, enabling a comprehensive characterization of the human genome as well as a better understanding of many genomic abnormalities. By delivering massive DNA sequences at unprecedented speed and cost, NGS promises to make personalized medicine widespread in the foreseeable future. To date, library construction with clinical samples has been a challenge, primarily due to the limited quantities of sample DNA available, but also due to the low quality of some samples, such as FFPE samples. To overcome this challenge, we have developed NEBNext® Ultra DNA Library Prep Kit for Illumina, a fast library preparation method using novel reagents and adaptors, including a DNA polymerase that has been optimized to minimize GC bias. This method enables library construction starting with nanogram quantities of DNA, and can be used for both intact and fragmented DNA, such as that found in FFPE samples. Moreover, the workflow is compatible with multiple NGS platforms, enabling use of a single library prep kit for creation of a variety of sequencing libraries. Illumina, 454, and SOLiD 5500 libraries were successfully generated and sequenced on the corresponding platforms to produce high quality sequencing data.

1666T

Profiling of the T-cell receptor variable region gene segment usage in tissue biopsies using Nanostring nCounter. *T. Peters, T. Valensise, N. Cheung, A. Seguin, V. Petitjean, E.J. Oakeley, S. Starck-Schwertz, M. Letzkus, K.J. Johnson, F. Staedtler.* Human Genetics & Genomics, Biomarker Development, Translational Medicine, Novartis Institutes for BioMedical Research (NIBR), Basel, Switzerland.

In order to generate a broad and structurally diverse adaptive immune repertoire, T and B-cell receptor genomic loci undergo non-precise somatic V(D)J segment rearrangements, including templated and non-templated nucleotide additions and deletions. Several studies observed that even in healthy subjects certain TCR V segments are commonly utilized while others are quite rare, also thereby reflecting a snapshot of the currently active adaptive immune system. A shift towards mono- or oligoclonality and the associated usage of dominant TCR V segments is often observed in infection, chronic inflammation, autoimmune disease and cancer. Recently, Zhang et al. described a Nanostring nCounter based "direct TCR expression assay" that allows a rapid and sensitive screen for the usage frequencies of all functional TCR Vα and Vβ segments. Here we report on the modification of the initially described codeset for inclusion of TCR Vγ and Vδ segments as well as the exemplary application for TCR V usage screening of colon and skin biopsies obtained from Crohn's disease and psoriasis patients. As TCR transcripts make up only a minority in total RNA from biopsies, our analysis involved PCR amplification of TCR transcripts and re-transcription to sense RNA for subsequent Nanostring nCounter analysis. We compare the obtained patterns of TCR V segment usage to corresponding profiles generated by massive parallel immune repertoire sequencing. In summary, profiling of TCR V segment usage in tissue biopsies using Nanostring nCounter enables the rapid and efficient detection of expanded T cell clones associated to disease.

1667F

Evaluation of whole genome amplified DNA and reduced genomic DNA for high performance of Illumina SNP microarrays. *C.L. Dagnall[1,2], L.M. Morton[1], B.D. Hicks[1,2], W. Zhou[1,2], X. Deng[1,2], M. Yeager[1,2], S.J. Chanock[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD; 2) Cancer Genomics Research Laboratory, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD.

The high genomic DNA input requirements for whole genome single nucleotide polymorphism (SNP) microarrays can limit the scope of molecular epidemiological studies. We evaluated alternatives to reduce input DNA requirements for Illumina® Infinium® SNP microarrays, using reduced genomic DNA (gDNA) amounts and whole genome amplified DNA (wgaDNA) as inputs into the Infinium® assay protocol. These alternate inputs reduce gDNA template requirements by 75% and 96.67%, respectively. 192 DNA samples from 71 individuals were obtained from multiple biospecimen sources and genotyped using standard and alternative inputs. When compared to genotypes obtained from the standard input amount, we observed 99.98% and 99.73% median concordance with reduced gDNA and wgaDNA, respectively. Our results demonstrate that carefully conducted studies with alternative inputs can yield high-quality genotyping results. These findings enable investigators to consider expansion of ongoing studies, including genome-wide association studies, challenged by small amounts of available DNA.

1668W

Utilizing the QuantStudio™ 3D Digital PCR System for BRAF V600E Mutation Detection in Papillary Thyroid Carinoma and Malignant Melanoma. *K. Hayashibara[1], L. Degoricija[1], E. Springer[2].* 1) Genetic Analysis, Life Technologies, South San Francisco, CA; 2) Institut für Pathologie, Universitätsmedizin der Johannes Gutenberg-Universität, Mainz, GERMANY.

BRAF V600E mutations are present in a high percentage of papillary thyroid carcinomas and malignant melanomas. Problems in routine diagnosis include limited sample quantities, ambiguous results, and the presence of PCR inhibitors, particularly in melanomas. A sensitive, rapid, and inexpensive method for detecting these and other mutations that is not affected by these limitations will be of great utility. Several papillary thyroid carcinoma and malignant melanomas from formalin fixed and native samples were analyzed using both a highly specific amplification refractory mutation system (ARMS) and digital PCR. Comparison of results indicates that digital PCR on the QuantStudio™3D System is able to detect BRAF V600E mutations in all samples where the mutations were detected by ARMS PCR, even when preamplification was required for mutation detection in some samples. Digital PCR was able to detect mutations in samples from melanomas that contained PCR inhibitors. The digital PCR results are also quantitative, allowing the determination of the mutant to non-mutant ratio of the samples.

1669T

Desktop Sequencing Using A Single-Use Cartridge-Based Consumable That Includes Target Enrichment, Amplification and Sequencing. *T. Raz[1], A. Gulamali[1], F. Zhuang[1], H. Ghandour[1], J. Sram[2], J. Healy[1], J. Downer[1], M. Griesbach[1], N. Nerkizian[1], P. Mary[1], S. Haserlat[1], V. Chellappa[1], J. Boyce[1].* 1) GnuBIO Inc., Cambridge, MA; 2) City of Hope, Duarte, CA.

Several major challenges have impeded the widespread application of targeted sequencing assays in the clinical setting. Foremost among these challenges are the high cost and complexity of sample preparation for the next generation sequencing platforms. Current technologies for target enrichment require between two to four days of preparation before samples are ready for the sequencing run. Furthermore, in order for current solutions to approach cost effectiveness, many samples need to be batched in a single run, adding to the complexity, increased likelihood of error and increased turnaround time. The GnuBIO unique desktop sequencer uses a single-use cartridge-based consumable that incorporates all of the preparation steps that are typically performed separately including target enrichment, amplification and sequencing. This drastically simplifies clinical workflows and reduces the operator hands-on time to less than 5 minutes, an unprecedented total. Biochemical reactions take place in a cascading manner, with individual reactions contained inside minute emulsion droplets, which flow through microfluidic channels where they are injected with assay reagents, incubated at the required temperatures, and finally detected for sequencing. Each droplet results in a single DNA sequence read, with reads up to 1000nt long. GnuBIO is collaborating with the City of Hope to develop an oncology targeted gene panel for detection of actionable mutations on the GnuBIO platform. The system is designed to achieve the following: /Starting material: genomic DNA from FFPE samples or other types of samples. /Total run time of less than 4 hours /Sample preparation requires only genomic DNA extraction (no DNA library preparation, enrichment or amplification required). /Robust detection of low frequency alterations (~1%-5%). /High accuracy - with average per base accuracy of 99.999% /Ease of integration with existing analytic and sample tracking workflows /Low per sample cost - $200 per sample /Low instrument cost - $50,000 In addition, the cartridge consumable is a closed system, and can be readily adapted to new panels. The cartridge is a dry consumable, containing all reagents needed to run a single DNA sample, reducing the risk of contamination and error.

1670F

The Advantage of Cold Plasma in Genomic Analysis. *C. Lunn, J. Shieh, P. Hensley.* IonField Systems, Moorestown, NJ.

Technological advances have significantly increased the sensitivity of genomic assay systems. Such advances have placed a premium on lab processes that keep samples contaminant free. That requirement for sample transfer is usually satisfied with costly single use plastic tips. As an efficient and cost effective alternative, IonField Systems has developed the TipCharger™ plasma treatment system, a microplate sized module that uses cold plasma to clean plastic pipette tips. Data shows that a 30-second plasma treatment removes DNA to undetectable levels without generation of any liquid or solid wastes. Use of the TipCharger can provide a valuable alternative that will speed analysis and decrease costs associated with modern genomic analysis.

1671W

Enhanced performance of whole exome and other targeted sequencing of small clinical samples. *J. Langmore, E. Kamberov, S. Yerramilli, T. Tesmer, J. Jessman, M. Carey, M. Carroll.* Rubicon Genomics, Inc, Ann Arbor, MI.

Next generation sequencing (NGS) enables analysis of any or all parts of the human genome in a way that opens new opportunities for clinical research and testing that is at more comprehensive, faster and less expensive than previously possible. For diagnostic applications that require deep sequencing such as cancer mutation discovery or detection from FFPE tissue or plasma, target enrichment by amplification or capture is critical to reach the goals for sequence accuracy and economy of sequencing time and reagents. Target capture using Agilent SureSelect or Roche NimbleGen SeqCap EZ reagents use a recommended amount of about 1000 ng of unamplified input DNA for a successful outcome. Thus it is difficult to use those methods on unamplified FFPE or plasma samples, because the amount of total DNA available is only 1 - 300 nanograms of total DNA. Manufacturer-recommended library amplification protocols specify at least 100 ng input. We have optimized library synthesis, amplification and capture protocols to achieve excellent whole exome and whole kinome results using less than 10 ng unfixed tissue or plasma DNA inputs, or larger quantities of FFPE DNA. We used the Rubicon ThruPLEX-FD library reagents, the Agilent and Roche target capture reagents, and modified protocols (when necessary). Sequencing metrics such as percentage reads on target, target coverage, and SNP or variant calling were equivalent to results using unamplified libraries or other library kits using substantially larger amounts of input DNA recommended using manufacturer-recommended protocols. Libraries made from the smallest amounts of input did produce the expected higher percentages of duplicate reads. The optimized reagents and protocols will be available to researchers and will enable high-quality targeted sequencing to be achieved starting with substantially less input of clinical DNA samples.

1672T

SureSelect strand specific RNA library prep kit provides a fast and streamlined workflow for preparing directional libraries from total RNA. *B. Arezi[1], B. Hsue[1], F. Useche[2], A. Tsalenko[2], B. Novak[2], A. Lucas[2], K. Chen[1], H. Tang[1], H. Hogrefe[1].* 1) Agilent Technologies, La Jolla, CA 92037; 2) Agilent Technologies, Santa Clara, CA 95051.

RNA-Seq is a revolutionary technology for whole transcriptome analysis. Deep sequencing of cDNA has been used to quantify transcript levels, confirm gene annotation, and identify novel transcripts, splice variants, and SNPs. RNA-Seq protocols that preserve strand information are critical for identifying antisense transcripts that play a role in gene regulation, determining the exact boundaries of genes transcribed on opposite strands, and accurately measuring expression levels of overlapping genes. Here, we present the SureSelect Strand Specific RNA Library Prep Kit, which includes all reagents required for isolating polyA RNA and preparing directional RNA libraries for Illumina sequencing from as low as 50ng total RNA, using the dUTP marking method (Parkomchuk et al (09) NAR 37: e123). In a comprehensive comparison of directional RNA-Seq methods, the dUTP marking method was found to provide the most compelling overall balance across all sequencing metrics examined, including strand specificity, library complexity, evenness and continuity of coverage, and accuracy of gene expression profiling (Levin et al (10) Nature Methods 7: 709-715). With our streamlined automation-friendly protocol, which includes master mixes and combined enzymatic steps, researchers can prepare directional cDNA libraries with >99% strand specificity from total RNA in about 5 hours, without the need for gel size selection. A comparison of RNA-Seq to gene expression array shows good correlation (R>0.8) between differential gene expression ratios (MAQCA and B) obtained using Agilent's SureSelect Strand Specific RNA Library Prep kit and SurePrint G3 Human Gene Expression Microarray 8x60K. Finally, we show that SureSelect Strand Specific RNA Library Prep is compatible with target enrichment by performing capture using the SureSelect Human Kinome bait library. Enriched libraries show high target and strand specificity (80% on-target reads; >99% strand specificity), and correlation of mean gene RPKM values between target-enriched and un-enriched libraries (R>0.8).

1673F

A Complete Work Flow for Single Cell Transcriptome Analysis: from Flow Sorting to Gene Expression Analysis. *C. Mason[1], L. Dennis[2], M. Krouse[2], J. Beechem[2], D. Mittar[1], E. Park[1], X.Y. Wang[1].* 1) BD Biosciences, San Jose, CA; 2) NanoString Technologies, Seattle, WA.

Once thought homogeneous populations of cells have recently revealed a startling amount of gene expression variation. New understanding has largely been driven by recent advances in the sensitivity of technologies such as next generation sequencing and sequence detection analysis, which are now capable of analysis at the single cell level. Here we describe how the power of flow cytomery can be exploited to isolate, characterize, phenotype, and purify multiple cell populations at the single cell level. We describe a complete workflow starting with heterogeneous cells that can be isolated at rapid speed by single cell depositions into 96-well plates using the newly developed BD FACSJazz™ flow cytometer from BD Biosciences. We demonstrate the use of new analytical tools such as index cell sorting, which can help track those individual cells during and after sorting. Finally we describe how, using single cell gene expression with the NanoString® platform, we can demonstrate high fidelity of the sorted cells.

1674W

Streamlined Methods for miRNA and Strand Specific RNA Library Construction for the Ion Torrent PGM. *D. Munafo, L. McReynolds, B. Langhorst, C. Sumner, P. Liu, E. Yigit, L. Apone, F. Stewart, J. Bybee, L. Mazzola, E. Dimalanta, T. Davis.* New England Biolabs, Inc., Ipswich, MA.

The Ion Torrent PGM has made it possible to generate large amounts of sequence data in relatively short run times. Consequently, it has become increasingly important to be able to generate large numbers of libraries from a variety of samples quickly and inexpensively. Here we present a fast workflow for construction of RNA libraries for whole transcriptome analysis. miRNA libraries can accommodate low total RNA input with no need for enrichment and are free of adapter-dimer contamination. mRNA libraries retain strand specificity for accurate gene expression quantification and discovery of antisense transcripts. The strand-specific libraries are compatible with poly(A) mRNA as well as ribosomal RNA-depleted total RNA. Using this simple and rapid procedure, library construction time is reduced to one day and the yield of the library is significantly improved. To reduce cost and increase sample throughput, libraries can be barcoded during amplification. The multiplexed libraries can then be pooled before size selection, reducing the number of steps in the workflow.

1675T

Novel Method for Multiplex Small-RNA Library Preparation with Improved Performance and Higher Sensitivity. *D. Rodriguez, D. Munafo, L. McReynolds, B. Langhorst, L. Apone, P. Liu, V. Panchapakesa, C. Sumner, E. Yigit, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, Inc. 240 County Road, Ipswich, MA 01938, USA.

Identification and analysis of small RNA by deep sequencing requires preparation of a di-tagged cDNA library. Most library preparation methods for di-tagged cDNA are based on sequential ligation of adaptors. The excess of adaptors required to provide sufficient library yield leads to adaptor-dimer formation that strongly contaminates the library (unless several gel purification steps are performed to remove unligated adaptors). We have developed a novel method to generate di-tagged small RNA libraries free of adapter-dimer contamination without introducing any additional enzymatic steps or gel purifications. Using this simple and rapid procedure, library construction time is reduced to one day and the yield of the library is significantly improved by enabling the addition of a high concentration of adaptors, thereby increasing the percentage of small RNA molecules included in the library. To reduce sequencing cost and increase sample throughput we have developed a barcode strategy to tag samples during library construction. Up to 24 multiplexed libraries can then be pooled together before size selection, reducing the number of steps in the workflow. Additionally, gel-free size selection greatly simplifies the overall workflow. Our method for generating multiplexed small RNA libraries reduces bias by ligation, increases representation of modified small RNAs and simplifies the workflow during library construction for small RNA analysis and discovery.

1676F

Molecular Indexing for Improved RNA-Seq Analysis. *M. Toloue, J. Risinger, P. Nakashe*. Bioo Scientific, Austin, TX.

Most current Next Generation Sequencing (NGS) library prep methods introduce significant sequence bias. The use of enzyme processing and fragmentation steps can introduce errors in the form of incorrect sequence and misrepresented copy number. Conventional RNA sequencing library construction involves the ligation of a population of cDNA molecules with adapters prior to amplification and sequencing. An inherent weakness of conventional RNA-Seq analysis is that cDNA fragments that amplify more efficiently will unavoidably result in a higher number of reads than cDNAs that do not amplify as well during the library construction PCR step. Therefore, when multiple reads mapping to the same transcript are encountered, it is not possible to determine whether sequenced reads originate from the same or different cDNA molecules. With molecular indexed libraries, each molecule is tagged with a molecular index randomly chosen from ~10,000 combinations so that any two identical molecules become distinguishable (with odds of 10,000/1), and can be independently evaluated in later data analysis. Analysis using molecular indexing information provides an absolute, digital measurement of gene expression levels, irrespective of common amplification distortions observed in many RNA-Seq experiments. This type of indexing requires no additional steps in RNA-Seq workflow and increases the precision of downstream analysis. At low sequencing depths, analysis use of molecular indices is identical to conventional analysis and generates equivalent RPKM values in all applications. As sequencing depth increases, individual molecular resolution also increases. In quantitative RNA-Seq experiments, the molecular indices distinguish re-sampling of the same molecule from sampling of a different molecule. At high sequencing depths, each molecule can be distinguished and the entire library can be analyzed to provide absolute numbers of each molecule. Resolving individual clones of molecules is critical for increasing sequencing accuracy, measuring bias, PCR duplication rates and identifying mutations in complex sample types. While it is well known that library prep methods introduce bias, tools for measuring it are needed if we are to start using NGS for accurate and quantitative gene expression measurements. Toward achieving that goal, we propose the use of molecular indices for all RNA-Seq experiments.

1677W

The Fluidigm® biobanking panel sensitively identifies gender contamination, sample degradation and low quality samples. *M.M. Lee, N.Y. Tuason.* Fluidigm Inc., 7000 Shoreline Court, Suite 100, South San Francisco , CA, 94080.

The lack of standardization and high quality samples from biorepositories can impede the progress of disease and basic research exposing a need for quality control and assessment methods to determine sample utility. The Fluidigm biobanking panel facilitates quality control and assurance studies on DNA samples. The panel consists of 96 assays selected by Dr. Andrew Brooks of Rutgers University to provide critical information regarding sample identity, integrity and quality. In this set of experiments, we will determine the limit of detection for gender contamination, assess DNA degradation detection and evaluate the performance of low quality samples assayed by the Fluidigm biobanking panel on the 96.96 genotyping integrated fluidic circuit in the BioMark™ HD system.

1678T

SNP genotyping using Affymetrix Axiom® Genotyping Solution. *M. Shapero, H. Loi, J. Law, A. Yan, D. Nguyen, C.S. Yu, M. Purdy, R. Kurapati, M. Shirazi, L. Bellon.* Genetic Analysis, Affymetrix, Santa Clara, CA.

The use of high-density DNA microarrays for accurate, cost-effective genotyping of single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (indels) plays an important role in the identification of the underlying genetic basis of common, complex human diseases. As the catalog of both common and rare variants in multiple worldwide populations continues to expand, microarrays are well positioned to capitalize on this evolving information content for targeted genotyping. Here we present an overview of Axiom® 384HT Genotyping Solution in which 384 individual arrays, contained in the footprint of a standard microtiter plate, offer the capability to simultaneously genotype 384 samples at 50,000 variants per sample with a throughput greater than 3,000 samples per week. The solution thus offers a novel technological capability to genotype large numbers of markers at a very high sample throughput in a cost-effective manner. Axiom® Genotyping Solution enables complete automation of DNA target preparation, including DNA amplification and enzymatic fragmentation of post-amplification products, on liquid-handling workstations. Following target preparation, arrays are processed using GeneTitan® Multi-Channel (MC) Instrument. The Axiom® myDesign™ Custom Arrays in the 384-layout can be designed with markers from the Axiom® Genomic Database or from *de novo* SNP discovery initiatives. Axiom 384HT Genotyping Solution retains full compatibility with the existing Axiom instrumentation platform and downstream data analysis software. Genotyping performance consistently achieves an average sample call rate ≥99.0%, average sample concordance to independent DNA genotype information (HapMap) ≥99.5%, and intra- and inter-run reproducibility ≥99.8%. In summary, new Axiom advances with the 384-array layout further extend the platform's capabilities for highly multiplexed genotyping of variants in a single assay. This offers high sample throughput coupled with minimal manual intervention enabling multiple applications including biomarker discovery, sample QC and tracking for large cohorts, and post-GWAS fine-mapping and causal variant analysis.

1679F

Successful Illumina Infinium Beadchip high-density genotyping from fragmented and low concentration samples. *J.C. Tackney[1], D.J. Witherspoon[2], L.B. Jorde[2].* 1) Dept. of Anthropology, University of Utah, Salt Lake City, UT; 2) Dept. of Human Genetics, University of Utah, Salt Lake City, UT.

Currently, high-density and high-quality SNP genotype data cannot be consistently assayed from small amounts of degraded DNA samples. DNA quality and concentration remain a bottleneck both in SNP microarray and next-generation sequencing library creation. While sequencing libraries have successfully been processed from degraded ancient DNA samples, costs are still prohibitively expensive for wide use in forensic DNA laboratories. Generating SNP genotype data from such samples remains a goal in identification of human remains and in crime scene investigations.

Towards that end we created panels of human DNA samples varying in concentration and fragmentation profiles. We first processed these samples through Illumina's proprietary Infinium FFPE DNA Restoration Solution and genotyped them on the HumanOmniExpress-FFPE BeadChip. We determined that we could genotype 1 ng of DNA fragmented to 300 bp with an ~84% call rate and, when compared to genotypes from 100 ng of non-fragmented control DNA, a ~2.2% discordant call rate. Obtaining more than 500,000 usable SNP genotypes from DNA of this quality has not been previously reported and suggests that this method can be applied to forensic DNA samples. This would allow accurate inferences of genetic relationships between samples.

We further processed 100 bp fragmented samples at lower input amounts. We investigated alternatives to Illumina's WGA step by creating sequencing libraries using a standard library preparation technique and a high-sensitivity single-stranded approach (Gansauge, M.T. and Meyer M. 2013. Nat. Protoc. 8(4), 737-48). We report on sequencing libraries as an alternative input for Illumina Beadchips.

This research was supported by DOJ award 2012-DN-BX-K037 and University of Utah Research Foundation and Interdepartmental grants to LBJ.

1680W

Computational pipeline for whole genome sequencing data analysis -- an application to trio families with 22q11 deletion. *J. Cai[1], K. Coleman[2], Z. Zhang[1], B. Morrow[1].* 1) Genetics Dept, Albert Einstein College of Medicine, New York, NY; 2) Human Genetics Dept, Children's Healthcare of Atlanta and Emory University School of Nursing, Atlanta, GA.

Whole genome sequencing (WGS) has become a viable strategy to discover variants associated with human disease. To facilitate WGS data interpretation, we established an automated pipeline for genetic variant (SNVs, INDELs) identification and annotation, using BWA, PICARD, GATK and ANNOVAR. For nonsynonymous SNPs, we predict the effect of amino acid changes by combining the output of BLOSUM62, SIFT and Polyphen2. We derive the MAF of known variants from the 1,000 Genomes Project or the NHLBI Exome Sequencing Project. We also utilize multiple tools including CNVnator, ERDS, and BreakDancer, which consider both depth and split-read information to generate CNV calls. To obtain a reliable set of CNVs, we follow up CNVs that are reported by at least two tools. We annotate known CNVs according to the Database of Genomic Variants. We applied this approach to WGS data from two unrelated families in which the child has velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome and both parents are normal. One child has a heart defect (tetralogy of Fallot; TOF) and the other has a normal heart and aortic arch. The known 22q11.2 deletions were verified by WGS. We identified 3,372,285 SNVs and 762,680 INDELs per sample. A total of 98% and 80.9% of them, respectively, are previously known SNPs found in dbSNP. We phased SNVs and INDELs using both trio and alignment information. This approach allows us to aggregate single base-pair level variants into gene level and dissect compound mutations within one gene. After phasing, we focused on highly possible deleterious variants, including nonsynonymous SNVs that at least one program predicted to be deleterious, stop gain/loss, splicing SNVs and frame-shift INDELs. We also filtered them with MAF <=1% from 1,000 Genomes Project. For each child, we counted the number of deleterious haplotypes for each gene and identified 20 candidate genes with two deleterious haplotypes in the child with TOF but not in the control. As we previously collected 303 heart related genes from literature, we examined whether these genes in the child with TOF were enriched with deleterious haplotypes compared to control, but resulted in no significant difference. We checked the other alleles of 22q11.2, we did not find any highly possible deleterious variants. For CNVs, we detected 1,256 novel deletions and 215 novel duplications per sample. Overall, this approach can be applied to large datasets to identify genetic modifiers of 22q11DS.

1681T

Blueprint: Resources provided by the large-scale Epigenomics project. *L. Clarke, D. Richardson, S. Wilder, P. Flicek, The Blueprint Consortium.* Vertebrate Genomics, European Molecular Biology Laboratory, European Bioinformatics Institute, The Wellcome Trust Genome Campus, Cambridge, United Kingdom.

Blueprint is a large-scale epigenomics project funded through the EC Framework Program 7 as the European Union's entry in the International Human Epigenome Consortium (IHEC). Blueprint aims to provide epigenomic data on several blood cell types both normal and diseased to help the scientific community improve its understanding of haemopoesis. This poster described the pipeline used to analyse the sequence data generated by our experiments and how the community can access the data. Blueprint is generating RNA-seq, ChIP-seq, DNase1-Seq and WGBS-seq data on several different blood cell types. This sequence data is aligned to the genome and and then appropriate signal calling analyses for the data type is undertaken. For the RNA-seq data expression levels are quantified, peaks are called for the ChIP-seq and DNase1-Seq data and the WGBS-seq has hypo- and hyper-methylated regions identified. We also perform some integrated analyses that include differental expression analysis and segmentation of the histone marks using ChromHMM. Blueprint releases new data regularly. On data release the raw sequence data and analysis alignments associated with the release are made available via the European Genome-phenome Archive (EGA) to users given permission by the Blueprint Data Access Committee (DAC). There are three primary mechanisms for release of the non-unique data types, such as expression levels, chromatin states and methylation levels. Our ftp site: (ftp://ftp.ebi.ac.uk/pub/databases/blueprint) holds the raw files and a Track Hub that can be associated with both Ensembl and the UCSC Genome Browser to allow easy visualisation of the data. This poster will primarily focus on these resources. We also provide a BioMart instance (http://blueprint.bsc.es/) to allow users to query the data and a Genomatix Browser (http://blueprint.genomatix.de/) for data visualisation. This Poster gives an overview of the Blueprint project as well as giving details of our analysis pipelines and information about data availability and accessibility for the project.

1682F

Exome sequencing identifies de novo mutations in patients with intellectual disability and epilepsy. *J. Halvardson, A. Zaghlool, AC. Thuresson, L. Feuk.* Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

Exome sequencing has proven to be an effective method to identify disease genes and have lead to the identification of causative genes in a large number of syndromes in recent years. Here we present the results of exome sequencing on 25 trios on a subset of patients with intellectual disability (ID), with additional symptoms of either epilepsy (19 trios) or congenital heart defects (6 trios). All trios have a negative family history and the parents are non-consanguineous. All patients have previously been screened for copy number aberrations using arrays, without clinically relevant findings.

We can report validated *de novo* single nucleotide variants (SNVs) or indels in coding regions in 11 of the 25 trios (44%). Of the 11 genes with validated *de novo* mutations, only two have previously been associated with ID or epilepsy (SCN2A and KCNA1). Thus, a high number of our findings are potentially novel disease genes. A subset of the genes have mouse knockout data that mirrors the patient phenotype, supporting a causative role for the *de novo* mutations. In one patient with severe ID we found a mutation in a highly conserved chromatin remodeling factor, which we are now following up with functional studies. This adds to a growing list of genes involved in chromatin remodeling that have been linked to intellectual disability and neurodevelopmental syndromes. The majority of the validated SNVs were missense mutations of conserved nucleotides in genes expressed in the brain. Several genes are involved in basic cell functions and we are currently investigating the possibilities of performing RNA sequencing on peripheral blood cells from trios to find patterns of expression in patients compared to healthy controls.

During this work we have created an in-house database of all sequenced exomes. This database contains several SNVs seemingly unique as compared to dbSNP, but common in the Swedish population. For example, when only counting SNVs present in 10 or more of the individuals sequenced in our trio project, about 14,000 of these SNVs are not occurring in the dbSNP database. Using this database as a filtering step leads to a significant reduction in the number of false positive SNV calls. Our data highlights the importance of a population specific filtering database when doing trio sequencing.

1683W

An ensemble genotyping approach for whole genome sequencing to reduce erroneous variant calls. *I.H. Lee[1], J.H. Park[1], Y. Choe[1], M.B. Neu[1], K. Lee[2], T. Hambuch[3], I.S. Kohane[1,2], R.C. Green[4], S.W. Kong[1], The MedSeq Project.* 1) Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Illumina, Inc., San Diego, CA; 4) Brigham and Women's Hospital, Boston, MA.

Whole-genome sequencing (WGS) has been effective in identifying causal genomic variants of rare and common complex diseases as well as developing personalized treatment strategies. Accurate genotyping is essential to utilize genome sequence information in clinical settings as aimed by the MedSeq Project, a randomized clinical trial for integrating WGS in medicine. The current consensus is to sequence using multiple platforms to achieve the highest level of accuracy, then to validate all disease-associated variants for a clinical report using Sanger sequencing. In the course of exploring the plausibility of reducing erroneous variant calls in clinical settings, we deployed an ensemble of 8 variant calling algorithms to 17 publicly available paired WGS samples prepared using 2 different platforms. Our ensemble approach significantly reduced possible false positives by 48.8% for single nucleotide variants (SNVs), 50.5% for loss-of-function SNVs, 77.2% for small indels, and 96.4% for de novo mutation (DNM) candidates when we assumed higher false positive rates among platform-specific variants. Among 10 independently validated DNMs in the same trio dataset, 8 somatic or germline mutations were accurately detected by our proposed method compared to none by a pedigree-aware variant calling algorithm. Moreover, we could increase the sensitivity for detecting known disease-causing mutations by combining the results from each algorithm. Our results suggest that the use of multiple variant calling algorithms is a cost-effective way to reduce both false positive and false negative variant calls. The source code of our ensemble genotype calling approach is freely available to the academic research community.

1684T

Investigating the significance of genetic proximity for HLA matched donors and recipients in unrelated allogeneic stem cell transplantation. *A. Madbouly, V. Paunic, M. Maiers.* Bioinformatics Research, National Marrow Donor Program, Minneapolis, MN.

Survival after hematopoietic cell transplantation is dependent on HLA matching between donors and recipients, ethnicity and other factors, most of which are unknown. Donors and recipient ethnic origin is usually self-reported, and it is not clear if matching based on genetic ancestry would result in better survival. While an individual's genetic composition is relatively fixed, self-identified race and ethnicity (SIRE) is a result of self-perception and can change over time. We aimed to investigate the effect of multiple measures of pairwise donor/recipient (D/R) genetic distance, using ancestry informative markers (AIMs), on transplant outcome. The AIMs panel selected for this pilot study consists of 500 autosomal SNPs. DNA samples from 300 fully HLA matched D/R pairs were genotyped for AIMs and were initially clustered using Structure and the 1000 Genomes dataset for reference parental populations. A genetic vector distance was estimated based on the distance between the D/R Structure vectors (describing individual ancestral proportions) in Euclidean space. Additionally, principal component analysis (PCA) was performed on the D/R genotypes for mapping genetic variation and Euclidean distance between D/R pairs was calculated on the PCA map to reflect genetic proximity. While a plateau was not explicitly reached, the optimal number of Structure clusters was estimated at K=5 based on the ad hoc delta(K) measure. The average Euclidean distance between donor and corresponding recipient Structure vectors was 0.347 (range [0.001, 1.376], p = 1). For the PCA, the first, second and third PCs accounted for 3.7%, 1.5% and 0.9% of the total variance respectively. The average PCA distance based on the first, second and third PCs was estimated at 1.4 (range [0.017, 9.18], p = 0.02), 2.05 (range [0.06, 10.3], p = 0.29) and 2.66 (range [0.19, 10.52], p = 0.786). Additionally, the average PCA distance was higher for SIRE mismatched D/R pairs (for the first PC, mean = 2.6, range [0.06, 9.2], p = 0.015) than SIRE matched D/R (mean = 1.14, range [0.01, 5.9], p = 0.99). Further analysis is being conducted on Caucasian sub-populations within the studied D/R cohort and a multivariate transplant outcomes analysis is underway. A power analysis will be performed to estimate the cohort size needed for the second (larger) phase of this study. Genetic-ancestry based analyses are expected to decrease the effect of confounding factors associated with self-reported ethnicity.

1685F

Heterogeneity of global gene expression microarray designs in detecting differentially expressed genes. *G. Malerba[1], D. Noel[2], A. Ferrarini[2], L. Xumerle[1], V. Mijatovic[1], P.F. Pignatti[1], M. Delledonne[2].* 1) Department of Life and Reproduction Sciences, University of Verona, Verona, Italy; 2) Biotechnology Department, University of Verona, , Verona, Italy.

Microarray technology is widely used for gene expression studies by many laboratories worldwide. Microarrays vary for the type of substrate used and for the type and number of oligonucleotide probes implemented. Moreover a disparate list of procedures to subtract background noise (BS) and normalize data (DN) among samples is available. All these factors help to make these reliable tools quite heterogeneous and heterogeneity may play an important role in identifying or not the differentially expressed (DE) genes when conducting global gene expression studies. To address this issue we essayed 4 different microarray platforms to analyze two Vitis vinifera berry developmental stages. Microarray data were processed using 20 different BS-DN combinations. The same RNA samples were also analyzed by RNASeq whose results were used as reference values. Microarray performances in detecting DE were assessed using several measures including correlation between fold-change, classification functions and the area under curve (AUC) of receiver operating characteristic (ROC) curves. The number of DE genes changed from one microarray design to another. Although arrays are commonly deemed reliable tools for gene expression studies their performances are not all the same when conducting studies of differential gene expression.

1686W

Representation of Medical Variation at NCBI: ClinVar, Gene, and MedGen. *D. Maglott, M. Landrum, J. Lee, W. Rubinstein, K. Katz, W. Jang, D. Hoffman, S. Chitipiralla, M. Ovetsky, J. Garner, R. Tully, L. Phan, D. Shao, R. Maiti, R. Villamarin, S. Gorelenkov, S. Sherry, D.M. Church.* Natl Ctr Biotech Info, NIH/NLM, Bethesda, MD.

Genetics professionals need access to organized information about reported relationships among sequence variation, genes, and phenotype to translate advances in genomic technology into medical practice. This presentation summarizes how ClinVar, Gene, MedGen and related tools interoperate to centralize the data, standardize its representation, and facilitate access.

ClinVar aggregates and archives information about sequence variation and assertions made about its clinical significance. ClinVar reports the data in multiple ways to support both interactive queries and integration into workflows. For example, clinicians and researchers may search a DNA or protein location to learn what is known about the clinical significance of variation, both now and previously. Users can access the evidence for or against an asserted phenotype based on its level of expert curation, and how an interpretation may have changed over time. Expert panels and professional societies can also submit assertions about pathogenicity, e.g. in a practice guideline. Aggregation and centralization of variant data from diverse sources such as testing laboratories, genome-wide research and curatorial groups will reduce the burden on professionals who no longer need collate this information independently and enable them to perform an informed and comprehensive assessment.

NCBI's **Gene** database organizes information that defines or is related to a gene: sequence, phenotypes, citations, pathways, interactions, and variations. For example, the reference sequences section, especially the RefSeqGene, reports the sequence standards against which variation is reported. The phenotype section summarizes the names of disorders related a gene, and information about dosage sensitivity. The pathways and interactions sections report other genes with products that, based on their interactions, may contribute to the same phenotype.

MedGen organizes information about conditions such as clinical features, related genes, practice guidelines, ontologies, GeneReviews, OMIM, published literature, consumer resources, and tests registered in the NIH Genetic Testing Registry (GTR). Synonyms and term hierarchies support querying and navigation to related information. The advanced search option allows retrieval based on combinations of clinical features and other concepts, including existence of related sequence variation.

1687T

Meta-analysis of FMRP mRNA Target Datasets Reveals Highly Associated mRNAs Mediated by G-quadruplex Structures Formed by Clustered WGGA Sequences. *J. Suhl[1], P. Chopra[1], S. Warren[1,2].* 1) Dept. of Human Genetics, Emory University, 615 Michael St, Atlanta, GA; 2) Dept. of Biochemistry and Pediatrics, Emory University, Atlanta, GA.

Fragile X syndrome (FXS), a common cause of inherited intellectual disability and a well known monogenic cause of autism spectrum disorder (ASD), is the result of loss or dysfunction of the selective RNA-binding protein FMRP. Since FMRP appears to modulate activity-dependent translation of its target mRNAs at the synapse, a major research effort has been to identify these mRNA targets. Three large-scale studies have attempted to characterize the mRNAs bound by FMRP, each using different methods and each generating lists of putative target genes, leading to distinct hypotheses by which FMRP recognizes its targets either by RNA structure or RNA sequence. However, very little in depth analyses have been performed to identify the level of consensus among the studies. Here, we describe a meta-analysis of these three large-scale FMRP target studies. Consensus lists comprised of mRNAs found to consistently interact with FMRP among the independent methods were generated, conferring a higher likelihood of bona fide association. Additionally, we examined all datasets for sequence elements within the target RNAs to validate the recently reported FMRP binding sequences ACUK and WGGA. We discovered that the sequence WGGA was significantly enriched in multiple FMRP target datasets, supporting this motif as a strong candidate recognition sequence. Unlike WGGA, the ACUK pattern was not enriched in the datasets suggesting that this motif may not be directly recognized by FMRP. The strong enrichment of WGGA in FMRP targets, coupled with the previous suggestion that G-rich secondary structure serves as a recognition element, indicates that the motif may be able to form a G-quadruplex secondary structure. We observed that the WGGA motifs in the FMRP targets are highly clustered and are consistent with the requirements for the G-quadruplex configuration, thus implicating the involvement of both structure and sequence as recognition elements for FMRP. Our analysis identifies the mRNA targets most likely affected by the loss of FMRP and, consequently, the genes critical for neuronal development and synaptic plasticity.

1688F

Real-world performance of five long-range PCR enzymes to amplify ~10kb amplicons from human genomic DNA. *H. Jia, K. Wang.* Zilkha Neurogenetic Institute and Department of Psychiatry, Keck School of Medicine, University of Southern California, Los Angeles.

Long-range PCR remains a flexible,fast,efficient and cost-effective choice for sequencing candidate genomic regions or genes,especially when combined with personal genome sequencers such as MiSeq and Ion Torrent.-Several long-range DNA polymerases are available commercially, and all of them are advertised as being able to amplify up to 15kb or longer genomic DNA.Although it is likely that they may all work well for specific genomic regions under highly optimized conditions,it is unclear what are their real-world performance on randomly chosen amplicons in human genomic DNA. Knowing the different characteristics of these enzymes help select enzymes for use in next-generation sequencing experiments. In the current study,we evaluated five long-range DNA polymerases to amplify three amplicons, with sizes of 5.8kb, 9.7kb, 12.9kb and Tm values of 54.5 ºC, 63.3ºC and 54 ºC respectively.These five long-range polymerases include Invitrogen SequalPrep Long PCR and AccuPrime Taq DNA Polymerase, TaKaRa PrimeSTAR GXL DNA Polymerase, KAPA long Range HotStart DNA polymerase and QIAGEN LongRange PCR Polymerase.The advertised lengths of amplication for these enzymes are 15kb, 20kb, 30kb, 20kb and 40kb, respectively. All experiments were designed according to the reaction mixture and cycling conditions on the manual. We found that PrimeSTAR can amplify all the targets with good quality using two-step unified PCR conditions.SequalPrep and AccuPrime can amplify 5.8kb and 12.9kb but not 9.7kb amplicon,unless PCR conditions are optimized.KAPA can amplify 5.8kb but not the two larger amplicons using their recommended long targets (5kb-18kb) PCR reaction and cycling conditions.QIAGEN cannot amplify any of the three amplicons according to the manual's conditions for very long-range PCR.In practice, many amplicons may need to be amplified at the same time, but primer pairs for them may have very different Tm values.Therefore,we also varied the annealing conditions for selected enzymes, and evaluated how sensitive the enzymes are to alterations of the annealing temperature and extension time.The results indicate that PrimeSTAR polymerase can amplify amplicons with different sizes and Tm values using the same PCR conditions, but others enzymes require alteration of PCR conditions to obtain optimal performance.Our results provided a useful guide to other researchers working on next-generation sequencing of candidate genes or regions by long-range PCR reactions.

1689W

Next-Generation Sequencing and Novel Variant Determination in a Cohort of 86 Familial Exudative Vitreoretinopathy Patients. *J. Salvo[1,2], H. Wang[1,3], K. Wang[3], D. Nguyen[5], K. Zhang[5], R. Chen[1,2,3,4].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Structural and Computational Biology & Molecular Biophysics Graduate Program, Baylor College of Medicine, Houston, TX; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Program in Developmental Biology, Baylor College of Medicine, Houston, TX; 5) Department of Ophthalmology, School of Medicine, University of California, San Francisco, CA.

Background Familial exudative vitreoretinopathy (FEVR) is a developmental disease that can cause visual impairment and retinal detachment at a young age. Four genes involved in the Wnt signaling pathway were previously linked to this disease: NDP, FDZ4, LRP5, and TSPAN12. Identification of novel disease causing variants of this disease allows for a deeper understanding of the disease, better molecular diagnosis, and improved treatment.

Methods 86 patients of families with hereditary FEVR were examined in this study. We used a proprietary capture panel to enrich for 294 known or suspected retinal disease causing genes in humans. Samples were processed using next generation sequencing (NGS) techniques followed by data analysis to identify and classify single nucleotide variants and indels. Sanger validation and segregation testing were used to verify suspected variants. This is the largest study of a FEVR cohort utilizing NGS that we are aware of.

Results Of the cohort of 86, 47 patients were probably solved (55%). Of these samples, 34 were due to rare variants in known FEVR causing genes, 25 of which were novel. The remaining 13 patients were found to have possible disease causing variants in other retinal disease causing genes, including four variants with a high probability of causation (i.e. stop-gain, frameshift, or splicing site mutations).

Conclusions We were able to determine probable disease causing variants in a large number of FEVR patients, the majority of which were novel. Knowledge of these variants will help to further characterize and diagnose FEVR.

1690T

Evaluation of two commercial exome capture kits on the Ion Proton. *F. Gedge, C. Chan.* Biobank, Sanford Research, Sioux Falls, SD.

With an increasing interest in exome analysis for research and clinical diagnostics the number of exome capture kits are available on the market is increasing. The two most published and commonly used are the Agilent SureSelect and NimbleGen SeqCap EZ kits. More recently Life Technologies released their TargetSeq kit for use on the Ion Proton instrument. NimbleGen has adapted their kit for use on the Ion Proton, while Agilent has yet to release an Ion-compatible SureSelect kit. Therefore, we evaluated the Life Technologies TargetSeq and Nimblegen SeqCap EZ v3 kits for exome sequencing on the Ion Proton. Both kits are liquid-phase capture kits, contain about 2 million capture probes and cover greater than 20,000 of the known coding genes plus varying amounts of RNA genes, miRNA binding sites and UTRs. Due to differences in probe design, TargetSeq has a target region of 46.2Mbases whereas the SeqCapkit targets 64Mbases. We compared the performance of these kits on the Ion Proton in terms of capture ability and specificity, coverage of the intended targets and coverage density sufficient for mutation analysis. Exome captures were performed according to the manufacturer's instructions followed by sequencing on the Ion Proton with the P1 chip. Both kits were able to capture their intended targets fairly sufficiently with about 85% of sequence reads being on target and both kits covered more than 92% of their targets with 1x coverage or greater. The remaining target sequence is either not well captured or there is a PCR bias against them. In order for exome sequencing to reliably identify mutations the amount of target covered with at least 20x or higher is important. In this respect, the two kits differ. With the TargetSeq we can sequence greater than 86% with 20x coverage. The SeqCap runs consistently have significantly lower average coverage as well as lower 20x coverage. This difference is likely due to the fact that the SeqCap kit covers almost 40% more bases than the TargetSeq kit, therefore lowering the average coverage per base. Given the need for sufficient coverage, and the fact that current Ion Proton output is limited with the P1 chip, the TargetSeq kit is preferable. However, since the other performance characteristics are similar, the SeqCap kit can be a valuable kit when sequence output on the Proton instrument is improved in the future.

1691F

A single-tube high-plex PCR approach for targeted massively parallel sequencing applied to FFPE-tumour derived material. *T. Nguyen-Dumont[1], B.J. Pope[2,3], F. Hammet[1], M.C. Southey[1], D.J. Park[1].* 1) Genetic Epidemiology Laboratory, Department of Pathology, Medical Building, The University of Melbourne, Victoria 3010, Australia; 2) Victorian Life Sciences Computation Initiative, The University of Melbourne, 187 Grattan Street, Carlton, Melbourne, Victoria 3010, Australia; 3) Department of Computing and Information Systems, The University of Melbourne, Victoria 3010, Australia.

High-plex PCR, as a sequence enrichment method for massively parallel sequencing (MPS), has been made available recently through a range of commercial solutions. However, protocol complexity, expense, and limited design flexibility, reduce their suitability in settings involving modest target size, or requiring low cost and high-throughput. In addition, methods to sequence DNA extracted from formalin-fixed, paraffin-embedded (FFPE) material have achieved variable success so far.

To address these limitations, we have developed Hi-Plex, a PCR-MPS strategy intended for high-throughput screening of multiple genomic target regions. Our library-building approach integrates a simple, automated primer design software developed in-house, enabling control of product size. Featuring permissive, inclusive thermocycling conditions and 'clamp' amplification bias reduction, our protocol is simple due to its single-tube nature. It is also cost and time-effective, using readily available reagents, is independent from expensive instrumentation, and requires minimal optimisation.

In a 60-plex test setting, targeting the breast cancer predisposition genes *PALB2* and *XRCC2*, Hi-Plex applied to 100 ng LCL-derived DNA resulted in 93.33% (56/60), 98.33% (59/60), and 100% of targeted amplicons represented within 5-fold, 10-fold and 12.5-fold of the mean on-target coverage, respectively, despite a broad range of amplicon sequence contexts. 86.94% of the library that mapped to the whole human genome was on-target. When applied to 100 ng FFPE-derived DNA, Hi-Plex resulted in 78.33% (47/60), 91.67% (55/60) and 100% of targeted amplicons represented within 5-fold, 10-fold and 25-fold of the mean, with 97.33% of reads on-target. With 25 ng FFPE-derived DNA, Hi-Plex resulted in 90% (54/60), 98.33% (59/60) and 100% of targeted amplicons represented within 5-fold, 10-fold and 12.5 fold of the mean, with 95.14% of reads on-target.

Hi-Plex represents a powerful new approach for screening of panels of genomic target regions. The benefits of Hi-Plex have been demonstrated in the context of degraded FFPE specimens and low level input DNA, where sequencing data quality was not compromised. Hi-Plex is efficient, cost-effective, and can be used for high-throughput mutation screening both for research and for molecular diagnostic and clinical purposes.

1692W

PhenoTips: Patient Phenotyping Software for Clinical and Research Use. *M. Girdea[1,2], S. Dumitriu[1], M. Fiume[1], S. Bowdin[3,4,5], K. Boycott[6], S. Chénier[7], D. Chitayat[4,8], H. Faghfoury[9,10], M.S. Meyn[3,4,5,11,12], P.N. Ray[3,11,12,13], J. So[9,14], D.J. Stavropoulos[13,15], M. Brudno[1,2,11]*. 1) Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; 2) Centre for Computational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Paediatrics, University of Toronto, Ontario, Canada; 6) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 7) Centre Hospitalier Universitaire de Sherbrooke, Quebec, Canada; 8) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, Toronto, Ontario, Canada; 9) Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 10) University Health Network, University of Toronto, Toronto, Ontario, Canada; 11) Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, Ontario, Canada; 12) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 13) Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 14) Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 15) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, Ontario, Canada.

We have developed PhenoTips, a deep phenotyping tool and database, specifically designed for phenotyping patients with genetic disorders. Our tool closely mirrors clinician workflows so as to facilitate the recording of observations made during the patient encounter. Phenotypic information is represented using the Human Phenotype Ontology; however, the complexity of the ontology is hidden behind a user interface, which combines simple selection of common phenotypes with error-tolerant, predictive search of the entire ontology. The software provides a series of features that help reduce the clinician's workload during the clinical examination. Together with standardized phenotypic data, PhenoTips supports entering demographic information, medical history (including prior laboratory results), family history, various measurements, relevant images depicting manifestations of the patient's disorder, genetic tests and their results, as well as additional notes for each of these categories. A pedigree drawing tool which enables the collection of advanced family histories is currently under development. The software automatically plots growth curves for a variety of measurements, selects phenotypes reflecting abnormal measurements, instantly finds Online Mendelian Inheritance in Man (OMIM) diseases that most closely match the phenotypic description, and can suggest additional clinical investigations that can improve the diagnosis.

PhenoTips is already used both in research studies and in the clinic, including the phenotyping of patients for the FORGE (Finding Of Rare disease GEnes) Canada project (http://care4rare.ca/), and the Undiagnosed Disease Program at the NIH. Our source code and a demo version of PhenoTips are available at http://phenotips.org.

1693T

Sequencing single human and bacterial cells at low coverage for aneuploidy, CNV, and genotyping applications. *E. Kamberov, T. Tesmer, S. Yerramilli, J. Jessman, M. Carey, M. Carroll, J. Langmore.* Rubicon Genomics, Inc., Ann Arbor, MI.

Single-cell analysis by using PCR and array assays is well established for determining aneuploidy, CNV, and genotyping single cells. Next-gen sequencing of single cells presents many barriers to complete and reproducible analysis, including incomplete genome coverage and irreproducible results. In some applications the information from multiple single cell experiments can be combined to increase coverage and reproducibility. However a surprising number of applications can be successfully executed using partial coverage of the genome as long as the coverage is reproducible. These technical applications include a) aneuploidy and copy number variation determination, b) SNP or other single nucleotide variations in a fraction of the genome, and c) identification of complex populations of cells. These technical applications are important for commercial applications such as pre-implantation genetic screening and diagnosis (PGS and PGD), prenatal diagnostics from single or small numbers of fetal cells in maternal circulation, cancer diagnostics from circulating tumor cells, and identification of infectious disease. In these applications coverage can be compromised as long as the partial coverage in reproducible. We have sequenced human and bacterial cells using a version of the Rubicon PicoPLEX technology that is being developed as a single-cell NGS library kit. This PicoPLEX-scD single-cell NGS prep is as simple as the PicoPLEX WGA kit, which is currently used for microarray and PCR studies and diagnostics from single cells. The PicoPLEX-scD prototype kits were used for the above technical applications. Sequencing quality, genome coverage, and reproducibility were measured in flow sorted and microdissected human cells. To verify that the MiSeq NGS could be used for PGS/PGD applications, as many as 48 single human cells were multiplexed on a single lane. Megabase losses or gains of copy number were reproducibly measured with as few as 200,000 clusters per sample. Partial genotyping and variant identification using single cells were also measured. Finally, single cells were studied in mixtures of other genomes. The results show that reproducible incomplete coverage was achieved by NGS of PicoPLEX libraries, and that many of the requirements for diagnostics of single cells could be achieved with minimal numbers of sequencing reads.

1694W

Cis and trans protein quantitative trait loci (pQTLs) identified using a high-throughput protein assay in 297 individuals from the AddNeuroMed cohort: the European collaboration for the discovery of novel biomarkers for Alzheimer's disease.
J.E. Mollon[1,2], M. Sattlecker[1,2], S. Kiddle[1,2], C. Johnstone[1,2], K. Lunnon[1], P. Proitsi[1,2], J. Powell[1], A. Hodges[1,2], S.K. Nelson[3], A. Stewart[3], S. Williams[3], H. Soininen[4], I. Kloszewska[5], P. Mecocci[6], M. Tsolaki[7], B. Vellas[8], S. Lovestone[1,2], S. Newhouse+[1,2], R. Dobson+[1,2]. 1) Institute of Psychiatry, King's College London, London, United Kingdom; 2) NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation Trust; 3) SomaLogic, Boulder, Colorado, United States of America; 4) Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 5) Medical University of Lodz, Lodz, Poland; 6) Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy; 7) 3rd Department of Neurology, Aristotle University, Thessaloniki, Greece; 8) INSERM U 558, University of Toulouse, Toulouse, France.

There has been much success in identifying gene expression quantitative trait loci using high-throughput gene expression measures such as array-based methods or sequencing (RNA-Seq), but high-throughput methods for quantifying proteins are not as widely available. In this study we use an aptamer-based protein assay from SomaLogic which quantifies over a thousand proteins from a single sample. Samples from 106 individuals with Alzheimer's Diseases (AD), 90 with mild cognitive impairment (MCI) and 101 healthy elderly controls were selected from the AddNeuroMed study. Genotyping and imputation resulted in 6,345,198 SNPs for analysis, and a total of 1016 aptamers targeting 1001 proteins were assayed in plasma from the same individuals. We applied regression models using QUICKTEST, measuring single-SNP effects after adjusting for age, gender and 5 genetic principle component axes. A Bonferroni threshold was applied ($p<4.93\times10^{-11}$), adjusting for multiple SNPs and proteins. Significant associations were further modelled in R, adjusting for original covariates as well as disease status and a single PC axis derived from the protein data. New associations with 87 proteins were discovered, and we replicated results from a previous study with overlapping samples. Twelve of the new associations were 'cis', defined as SNPs within the coding region of a gene associated with that gene's protein product. A further 14 proteins were associated with SNPs outside the protein-coding gene but on the same chromosome. Sixty-one proteins had associations with SNPs on different chromosomes ('trans'). The top cis association was between MSP and rs3197999 (3.83×10^{-74}), a SNP in the third intron of MST1, the gene encoding MSP. This SNP is predicted damaging in an alternate transcript of MST1, and has been associated with Crohn's Disease, ulcerative colitis and primary sclerosing cholangitis. The top trans association was between MMP8 and rs12614 (p=2.88x10-70), a non-synonymous (predicted damaging) SNP in FBI12 (an alternate splice variant of CFB). We will describe properties of variants we have found to associate with protein plasma levels and highlight those previously identified as being disease related through GWAS studies. We conclude that the study of genetic effects on intermediary phenotypes such as proteins yields promising results, and may allow us to begin to unravel the mechanisms through which genetic effects manifest in disease.

1695T

Non-additive effects of genes in human metabolomics.
Y.A. Tsepilov[1,2], J.S. Ried[3], C. Gieger[3], K. Strauch[3,4], S. Shin[5,6], N. Soranzo[5], T. Spector[7], Y.S. Aulchenko[1]. 1) Institute of Cytology and Genetics SD RAS, Novosibirsk, Russia; 2) Novosibirsk State University, Novosibirsk, Russia; 3) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 4) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 6) MRC University of Bristol Integrative Epidemiology Unit (IEU), Bristol, United Kingdom; 7) Department of Twin Research and Genetic Epidemiology, King's College London, St Thomas' Hospital, London, United Kingdom.

Genome-wide association studies (GWAS) are widely applied to analyze the genetic effects on phenotypes. In most GWAS the SNP effects are assumed to be additive. The applied models have a lower power to detect non-additive effects (such as recessive, dominant, over-dominant). Therefore, the systematic analysis of non-additive effects can lead to the identification of additional SNPs. Moreover, the knowledge of the underlying SNP effect can be important for a more complete understanding of the genetic determination. Large-scale GWAS on metabolite concentrations and ratios assuming additive SNP effects have revealed several loci. In this study we analyzed a large panel of metabolites with a more general model that covers additive as well as non-additive SNP effects. Serum metabolites and genome-wide SNP data of the population-based cohort KORA F4 (N=1,744) were used for discovery. We calculated GWAS on 151 metabolite concentration, as well as all possible ratios of concentrations (total 22,801 traits) using a two degrees of freedom model (genotypic model). Seventeen loci were found to be genome wide significantly associated (Bonferroni corrected p-value $\leq 2.19\times10^{-12}$) with at least one metabolite (ratio). For each of these loci the four models assuming different SNP effects (additive, dominant, recessive, over-dominant) were tested and the results were compared. For five SNPs a non-additive effect model had the best results (for three the genotypic model was the best, for one the dominant and for one the over-dominant). The additive model was not significant for three of them. In the next step, we replicated these findings using independent TwinsUK study (N=846). The study demonstrated that additional SNPs can be identified with the usage of non-additive SNP effects. The presented strategy of using the genotypic model can increase analysis power and the potential to uncover additional loci involved in the control of human traits.

1696F

Integrating population genetic and functional genomic resources for accurate prediction of deleterious protein-altering variants.
P.J. Walter, M. Sampson, H.M. Kang. Biostatistics, University of Michigan, Ann Arbor, MI.

As a wealth of genomics resources becomes available across multiple species, populations, and tissues, an important area of research will consist of answering the question of how best to leverage this complementary information to unravel the etiology of disease traits. Currently there are many different methods to predict conserved elements through multiple sequence alignment and methods to predict functional impact of a protein-altering mutation. In addition, rapidly increasing amount of population-scale sequencing data allows to estimate the level of selection across human population, providing additional information on top of existing functional resources. These resources can then be integrated to prioritize variants that are likely to be causal in gene-level rare variant tests and post-hoc functional analysis.

Here we integrated existing conservation metrics such as PhyloP and GERP++ scores, predicted functional scores of protein-altering variants including SIFT and Polyphen2, using a support vector machine (SVM) approach trained by known pathogenic and non-pathogenic variants deposited in dbSNP and HGMD to collapse predict the likelihood that a mutation is deleterious. Our integrated variant score predicts the functional impact not only of existing variants, but also of every potential allele (>80 million) of protein-altering single nucleotide variant (SNV) that can be annotated by the latest GenCODE database. Using cross-validation, we estimate that our integrated variant score prioritizes 44.4% of known pathogenic variants over variants observed in 1000 Genomes within top 5% quantile, which is 4.01 and 1.52-fold higher than what could be predicted by GERP++ and Polyphen2 scores, respectively. We demonstrate that novel the variants shown to cause Mendelian forms of rare diseases have strong enrichment of high integrated variant score. In addition, we show that using this integrated variant score as a filter in gene-level burden tests can substantially improve power to identify causal genes, compared to the widely used approach relying on minor allele frequency.

1697W

Mixed model approaches for transcriptome profiling of reciprocal dosage imbalances. *A. Ragavendran[1], I. Blumenthal[1], S. Erdin[1], L. Klei[4], K. Roeder[5], B. Devlin[4,5], M.E. Talkowski[1,2,3].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Departments of Neurology and Genetics, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 5) Department of Statistics, Carnegie Mellon University, Pittsburg, PA.

Methods for analysis of RNA-Seq data are still in a nascent stage of development. Current packages for analyzing differential expression (DE) are developed primarily to accommodate low replication and simple designs. Few methods have been established to model family-based designs. We present some insights from ongoing statistical analysis of RNA-Seq count data associated with reciprocal dosage imbalances, which collectively represent a major source of genetic risk for neurodevelopmental disorders. We analyzed data from two parallel study designs: a family based study from multiplex autism families harboring a 16p11.2 deletion or duplication, and mouse models deleted or duplicated for the syntenic region. Counts generated with RNA-Seq are over-dispersed, thus modeled with negative binomial GLMs (nbGLMs). Analysis of ERRC transcripts (synthetic spike-ins) showed that nbGLM was the best-fit model relative to Poisson or quasi-Poisson fits. This implies the process from sequencing to quantitation introduces additional sources of variance independent of biological variation, which bears further investigation. Of 37 ERCC transcripts in mouse, we detected significant DE (p-value < 0.0001) in 35 transcripts among 3 groups of fold-change (FC) (min conc. 7.5 amols/μl), with the least power in the 0.68 FC group (min conc. 14.6 amols/μl). For statistical inferences in family-based designs, correlations due to shared genetic background are accommodated with mixed effects models, which current packages don't handle. Generalized Linear Mixed Models (GLMMs) are a natural extension to GLMs and we use the *pedigreemm* package in R to fit GLMMs on a gene-by-gene basis. We adopt a Poisson-lognormal GLMM and include random effects for family, while correlations between family members are modeled from pedigree-based estimation of relatedness. We show that two complementary models represent a reasonable approach to study effects of reciprocal dosage imbalance in the genome. The first is to evaluate the effects of deletion/duplication using a regression on the copy number for linear trends in DE. The second is to fit contrasts within a cell-means model to delineate effects specific to loss or gain of function. In both systems, the models capture dysregulation of most genes within the 16p11.2 region from dosage compensation due to copy number state. Inference for GLMMs is an ongoing area of research and we explore randomization procedures to generate empirical p-values.

1698T

Statistical inference of eQTL sharing among many tissues. *T. Flutre[1], X. Wen[2], J. Pritchard[1,3], M. Stephens[1,4].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Biostatistics, University of Michigan Ann Harbor, Ann Harbor, MI; 3) Howard Hughes Medical Institute, Chevy Chase, MD; 4) Statistics, University of Chicago, Chicago, IL.

Using gene expression as intermediate phenotype is now a common approach to interpret associations between genetic variants and organismal phenotypes. Until recently, most studies were performed on a single tissue or cell-type sampled in a hundred individuals. Statistical methods were improved so as to now robustly detect genetic variants associated with changes in gene expression (eQTLs). As typical effect sizes are weak, several recent studies analyzed up to one thousand individuals, showing that most genes have at least one eQTL. However, interpretation of such associations is hampered by the fact that easy-to-sample tissues may often be irrelevant to the etiology of organismal phenotypes of interest. This prompted the NIH to fund the 'genotype - tissue expression' pilot project (GTEx) aiming at building the largest, tissue-wide eQTL data resource to date.

In prevision of such a data set, we recently developed a statistical framework to detect eQTLs with high power by jointly analyzing multiple tissues, and to reliably infer the proportion of tissue-consistent and tissue-specific eQTLs (Flutre *et al*, 2013). As part of the GTEx consortium, we applied our model on the current data set of 9 tissues from 100-200 individuals. This analysis, done in close collaboration with other groups in the consortium, is planned to be presented in the GTEx session.

However, most current methods are unable to efficiently cope with the larger number of tissues that will be available in the future. The reason stems from the use of 'configurations', binary vectors representing activity patterns of eQTLs among tissues. Indeed, a data set of 20 tissues generates 2^{20} configurations (> 10^6). Instead of considering each of them, our improved model learns only the subset of configurations present in the data. Moreover, we don't restrict ourselves anymore to genes expressed in all tissues, by directly analyzing gene expression in terms of read counts. The results obtained with this new model will be presented on the 9 tissues comprising the current GTEx data set. To alleviate the need for permutations in such large-scale studies, we also developed an efficient, yet conservative procedure to control the Bayesian FDR, and evaluate it rigorously by doing permutations.

1699F

A genome-wide assay for regulatory functional potential of sequence variants. *J.M. Peralta[1,3], J. Blangero[1], L.J. Abraham[2], E. Moses[3].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA; 2) School of Chemistry and Biochemistry, University of Western Australia, WA, AU; 3) Centre for Genetic Origins of Health and Disease, University of Western Australia, WA, AU.

DNase I hypersensitivity sites (DHSs) are the result of chromatin remodeling processes that create openings in the chromatin which expose the DNA to digestion by DNase I. A considerable proportion, 34-88%, of the disease-associated variants detected by GWAS studies appear to cluster in DHSs. Here we propose that non-coding variants can have a cis-functional effect that changes the local sensitivity to DNase I, which can be detected in heterozygotes as allele-specific DHSs. This would suggest a potential regulatory role for a variant, since DHSs often correlate with cis-located elements that regulate the expression of a gene, like promoters and transcription binding sites. We call this effect the functional potential (FP) of a variant and we define it for a SNP as a significant deviation from the expectation of equal allele depth coverage at DHSs. To test our approach we selected a 2842123 SNP reference panel (RSNP) from SNPs with heterozygous genotypes in Illumina's Platinum Genomes WGS from sample NA12878. We then measured allele-specific genome-wide chromatin accessibility using mpileup from samtools to call SNPs using the short sequencing reads (reads) obtained from DNase-seq mapped DHS reads from five GM12878 lymphoblastoid cell line replicates (R1-R5) for NA12878 released by the ENCODE Project. The read depth of the two alleles (from the DP4 annotation) at each SNP was used to estimated the FP for a total of 54533 variants (46652 distinct SNPs) that were called from DHSs-mapped reads from all replicates that intersected with RSNP. There was large variation in the number of SNPs called per replicate (mean=10907, sd=6065) and little overlap between replicates (3.6% SNPs, n=1675, seen in > 2 replicates). This was expected due to the short read lengths (20bp) and low read coverage at DHSs (mean= 6.32, min=1, max=1195). Nevertheless, we found that the FP estimates for a SNP were well correlated between the replicates where it was seen (mean r=0.88, sd=0.09). We focused on the set of SNPs overlapping more than 2 replicates and observed significant ($p<9.16\times10^{-7}$) FP estimates for 6.9% (n=116) after correcting for multiple testing. These SNPs showed a very strong allele-specific bias in coverage (95.1% on average), within high-coverage DHSs (mean=170, min=18, max=1195), that was consistent across five replicates. These results suggest that our FP approach is feasible and that it can potentially use other measures of regulatory activity (MNase and FAIRE-seq).

1700W

DeepSAGE Reveals Genetic Variants Associated with Alternative Polyadenylation and Expression of Coding and Non-Coding Transcripts. *D.V. Zhernakova[1], E. de Klerk[2], H.-J. Westra[1], A. Mastrokolias[2], S. Amini[2], Y. Ariyurek[2,3], R. Jansen[4], B.W. Penninx[4], J.J. Hottenga[5], G. Willemsen[5], E.J. de Geus[5], D.I. Boomsma[5], J.H. Veldink[6], L.H. van den Berg[6], C. Wijmenga[1], J.T. den Dunnen[2,3], G.-J.B. van Ommen[2], P.A.C. `t Hoen[2], L. Franke[1].* 1) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands; 2) Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 3) Leiden Genome Technology Center, Leiden, The Netherlands; 4) Department of Psychiatry, The Netherlands Study of Depression and Anxiety, VU University Medical Center, Amsterdam, The Netherlands; 5) Department of Biological Psychology, Netherlands Twin Registry, VU University, Amsterdam, The Netherlands; 6) Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, The Netherlands.

Many disease-associated variants affect gene expression levels (expression quantitative trait loci, eQTLs) and expression profiling using next generation sequencing (NGS) technology is a powerful way to detect these eQTLs. We analyzed 94 total blood samples from healthy volunteers with DeepSAGE to gain specific insight into how genetic variants affect the expression of genes and lengths of 3'-untranslated regions (3'-UTRs). We detected previously unknown *cis*-eQTL effects for GWAS hits in disease- and physiology-associated traits. Apart from *cis*-eQTLs that are typically easily identifiable using microarrays or RNA-sequencing, DeepSAGE also revealed many *cis*-eQTLs for antisense and other non-coding transcripts, often in genomic regions containing retrotransposon-derived elements. We also identified and confirmed SNPs that affect the usage of alternative polyadenylation sites, thereby potentially influencing the stability of messenger RNAs (mRNA). We then combined the power of RNA-sequencing with DeepSAGE by performing a meta-analysis of three datasets, leading to the identification of many more *cis*-eQTLs. Our results indicate that DeepSAGE data is useful for eQTL mapping of known and unknown transcripts, and for identifying SNPs that affect alternative polyadenylation. Because of the inherent differences between DeepSAGE and RNA-sequencing, our complementary, integrative approach leads to greater insight into the molecular consequences of many disease-associated variants.

1701T

The association of urinary Interferon inducible protein-10 levels with kidney allograft rejection. *S. Firasat[1], A. Raza[1], A. Abid[1], T. Aziz[2], M. Mubarak[3], S.A.A. Naqvi[2], S.A.H. Rizvi[2], S. Khaliq[1,4], S.Q. Mehdi[1].* 1) Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 2) Department of Urology, (SIUT), Karachi, Pakistan; 3) Department of Histopathology, (SIUT), Karachi, Pakistan; 4) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan.

Abstract: Interferon-gamma inducible protein-10 (IP-10) is a chemokine secreted by immune and non-immune cells. IP-10 and its receptor CXCR3 have been shown to participate in the immune response against kidney allograft. We aimed to investigate the association of urinary IP-10 levels with rejection episodes in renal transplant patients during the first year post transplant. A total of 206 samples were included in this study. Of these, 118 samples were taken from 93 renal transplants recipients with various histopathological findings of rejection (acute cellular ACR, borderline BLR and acute vascular rejection AVR) and of no-rejection episodes (NAD). Additionally, 88 urine samples were also collected from 88 controls (renal donors). Urinary IP-10 levels were quantified by using the Human CXCL10/IP-10 Quantikine ELISA kit. The data were analyzed using the Statistical Package for Social Sciences (SPSS) and MedCalc software. Statistically significant differences in the urinary IP-10 levels were found between rejection and control groups (p=0.000) and rejection vs. NAD group (p=0.004) groups. Among rejection groups, the mean IP-10 levels for ACR, BLR and AVR were 256.8±52pg/mL, 196±110pg/mL and 113±54pg/mL respectively. The statistical significant difference for IP-10 was found between ACR vs. NAD group (p=0.001) and ACR vs. BLR (p=0.019). ROC curve analysis showed a significant difference in rejection vs. NAD (p=0.004) and rejection vs. control (p=0.0001). At a cutoff value of 8.5pg/mL, the area under roc curve (AUC) was 0.87 with a sensitivity and specificity of 82.3% and 78.4% respectively. However, rejection vs. NAD showed p=0.0013 with 71.8% sensitivity and a 64% specificity at a cutoff value of 27pg/mL and the AUC was 0.70. Renal graft-survival among different urinary levels of IP-10 showed a significant difference in the mean time of graft survival between <100pg/mL and >200pg/mL (9.26±1.1 vs. 3.7±1.0 weeks; log rank p=0.000) and 100-200pg/mL (9.26±1.1 vs. 3.3±0.8 weeks; log rank p=0.017). The results indicate an increased level of IP-10 in the urine of allograft rejection samples that had different histopathological findings.

1702F

A General Statistic Framework for Identifying Genetic Variants of Clinical Significance. *L. Ma, M. Xiong.* University of Texas School of Public Health, Houston, TX.

Fast and cheaper next generation sequencing (NGS) technologies will generate unprecedentedly massive (thousands or even ten thousands of individuals) and highly-dimensional (up to hundreds of millions) genomic and epigenomic variation data. In the near future, a routine part of medical record will include the sequenced genomes. A fundamental question is how to efficiently extract genomic and epigenomic variants of clinical utility which will provide information for optimal wellness and interference strategies. Traditional paradigm for identifying variants of clinical validity is to test association of the variants. However, significantly associated genetic variants may or may not be usefulness for diagnosis and prognosis of diseases. Alternative to association studies for finding genetic variants of predictive utility is to systematically search variants that contain sufficient information for phenotype prediction. To achieve this, we introduce concepts of sufficient dimension reduction (SDR) and coordinate hypothesis which project the original high dimensional data to very low dimensional space while preserving all information on response phenotypes. We then formulate clinically significant genetic variant discovery problem into sparse SDR problem and develop algorithms that can select significant genetic variants from up to or even ten millions of predictors with the aid of dividing SDR for whole genome into a number of sub-SDR problems defined for genomic regions. The sparse SDR is in turn formulated as sparse optimal scoring problem, but with penalty which can remove row vectors from the basis matrix. To speed up computation, we apply the alternating direction method for multipliers to solving the sparse optimal scoring problem which can easily be implemented in parallel. To illustrate its application, the proposed method is applied to simulation data and the NHLBI's Exome Sequencing Project (ESP) dataset as well as the TCGA dataset.

1703W

The n=1 problem in human genetics: identification of rare disease mutations from single genomes. *A.B. Wilfert[1], J.N. Constantino[2], D.F. Conrad[1].* 1) Genetics, Washington Univeristy in St. Louis, St. Louis, MO; 2) Psychiatry, Washington University in St. Louis, St. Louis, MO.

Autism and azoospermia are examples of reproductively lethal diseases with a large population prevalence (about 1%) and extensive genetic heterogeneity (at least 500-1000 genes each). Standard epidemiological methods have begun to falter in the genetic analysis of these traits as cohorts of realistic size will contain many individuals with private, causal mutations that cannot be pinpointed by GWAS or linkage. Likewise, it has been estimated that over 40% of Mendelian disorders remain undiagnosed, which is in part due to sample size limitations. The purest presentation of these challenges is called the 'n=1' problem: given a single genome with a suspected monogenic disease phenotype, can one identify the causal mutation? We have created the first statistical framework that explicitly addresses the n=1 problem. This is an integrative framework that relies heavily on published data sets to consider gene physiology, functional protein-level models and patterns of population genetic variation to model disease and healthy variation. We use variants from the Human Gene Mutation Database (HGMD) and the 1000 Genomes Project to train and test these models. Spike-in analyses show that over 80% of the HGMD variants tested have been prioritized into the top 10 variants within a typical healthy exome and ROC curves show that this integrated prioritization method outperforms each of its component predictors, with an AUC of 0.990. For over 100 Mendelian diseases, each with an average of 5 distinct mutations in HGMD, our method correctly identifies the causal mutation as the most pathogenic event within an otherwise healthy exome, indicating that, in principal, our method solves the n=1 problem for these diseases. We describe applications of this method in two unique cases of autism and azoospermia, identifying candidate mutations in each of these cases, where the genetic etiology was previously unknown. We apply our method to describe the distribution of the most pathogenic mutation detected in 1000 healthy individuals sequenced by the 1000 Genomes Project. We discuss the complications of applying n=1 methodology in the presence of somatic mutation, which may be mistaken as germline variants when sequencing primary cells. While the number of developmentally acquired somatic mutations detectable in a given tissues may be small (e.g. 30-300) we show that these are much more likely to be appear pathogenic than germline variants by pathogenicity assessment tools.

1704T

Meta-analysis of twelve genome-wide association studies (GWAS) identifies novel genetic loci associated with mammographic density phenotypes. *S. Lindström[1], D. Thompson[2], A.D. Paterson[3], J. Li[4], G.L. Gierach[5], J. Stone[6], J.A. Douglas[7], I. dos Santos Silva[8], J. Benitez[9], C. Scott[10], P.A. Fasching[11], L. Baglietto[12], M. Southey[6], G. Giles[12], M. Pollan[9], J. Figueroa[5], F.J. Couch[13], J.L. Hopper[6], P. Hall[4], D.F. Easton[2], N.F. Boyd[14], C.M. Vachon[10], R.M. Tamimi[1,15], Markers of Density (MODE) consortium.* 1) Program in Molecular and Genetic Epidemiology, Department of Epidemiology , Harvard School Of Public Health, Boston, MA, USA; 2) Centre for Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Hormonal and Reproductive Epidemiology Branch, National Cancer Institute, Rockville, MD; 6) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, The University of Melbourne, Melbourne, Australia; 7) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI; 8) Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; 9) Centro Nacional de Investigaciones Oncologicas, Madrid, Spain; 10) Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 11) Erlangen University Perinatal Center, Department of Gynaecology and Obstetrics, Erlangen University Hospital, Friedrich Alexander University of Erlangen-Nuremberg, Erlangen, Germany; 12) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia; 13) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA; 14) Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, Toronto, Ontario, Canada; 15) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA.

Mammographic density reflects the proportion of stromal and epithelial tissues in relation to adipose tissue in the breast and is one of the strongest risk factors for breast cancer. Mammographic density phenotypes are defined by the area of a mammogram that consists of stromal and epithelial tissue (dense area), adipose tissue (non-dense area) and the proportion of stromal and epithelial tissue in relation to total breast area (percent density). Mammographic density has been shown to have high but poorly understood heritability. We have previously conducted a genome-wide association study (GWAS) of percent density within the MODE consortium and identified common genetic variants in the breast cancer susceptibility gene ZNF365 to be associated with percent mammographic density. Here we report results from an extended GWAS of percent density as well as the first GWAS of dense area and non-dense area. In the discovery phase, we conducted a meta-analysis of 12 GWAS of percent density, dense area, and non-dense area in more than 7,300 women of European descent. Ten of the studies used linear regression treating the mammographic density phenotype as a quantitative trait adjusting for age, body mass index (BMI) and menopausal status. Two studies selected women based on the extreme categories of mammographic density (one using percent density and the other dense area) adjusted for age, BMI and menopausal status. A total of 205 SNPs that were associated with at least one density phenotype at a significance level of P10-4 were selected for replication in 5,000-8,300 women independent from our discovery phase. We identified novel genome-wide significant ($P<5\times10^{-8}$) loci for percent density at 5q23.2, 8p11.23 and 22q13.2, loci for dense area at 4q13.3, 6q25.1, 10q21.2, 12q23.2 and 22q13.1, and a locus for non-dense area at 8p11.12. Three of these regions have been identified as breast cancer susceptibility loci. In summary, we report novel associations between common genetic variation and three mammographic density phenotypes. Many of the genetic variants found to be associated with mammographic density are also associated with breast cancer supporting the hypothesis that there is a shared genetic basis between mammographic density and breast cancer risk.

1705F

Significant Role of Combined Alleles of Hormonal Receptors and Hepatocanalicular Transporter Gene in Susceptibility to Cholesterol Gallstone Disease. *A. Srivastava, A. Mishra, B. Mittal.* Genetics, SGPGIMS, Lucknow, India.

Rationale: Cholesterol gallstone disease (CGD), one of the commonest digestive ailments in westernized and developing countries, is caused by a range of cholesterol flush and supersaturation insults. Although this multifactorial disease involves complex interactions among environmental and genetic factors but apart from female gender and increasing age, very little is known about the genetic interactions underlying the phenotypic expression of CGD. Therefore, one or more variations in genes could play critical roles in the diverse pathways which can further progress to cholesterol crystallization in gallbladder. In the present study, we performed multigene interactive analysis as genotyping score, Multifactor dimensionality reduction (MDR) and Classification and Regression Tree analysis (CART) to identify combinations of alleles among the hormonal, hepatocanalicular transporter and adipogenesis differentiation pathway genes in modifying the risk for CGD. Design: The present study recruited a total of 450 subjects including 230 CGD patients and 220 healthy controls who completed an interview and provided blood. The study was approved by the local ethical committee and we followed the norms of Declaration of Helsinki. We analyzed common ESR1, ESR2, PGR, ADRB3, ADRA2A, ABCG8, SLCO1B1, PPARγ2, and SREBP2 gene polymorphisms to find out combinations of genetic variants contributing to CGD risk. Genotyping was carried out by PCR-RFLP, ARMS-PCR and Taqman Assay. Statistical analysis was performed by using SPSS ver16 Results: Single locus analysis by logistic regression showed positive association of ESR1 IVS1-C397T (rs2234693), IVS1- A351G (rs9340799), PGR I/D (rs1042838), ADRB3- T190C (rs4994), ABCG8 G145C (rs11887534), SLCO1B1 CExon4A (rs11045819) and SREBP2 G1784C (rs2228314) with CGD risk. However, the MDR and CART analysis revealed ESR1 IVS1- C397T (rs2234693), ADRB3- T190C (rs4994) and ABCG8 G145C (rs11887534) polymorphisms as the best polymorphic signature for discriminating between cases and controls. The overall odds ratio for the applied multi-analytical approaches ranged from 4.33 to 10.05 showing an incremental risk for cholesterol crystal formation. Thus, the CART and MDR analyses also showed the importance of ESR1 IVS1- C397T and ABCG8 G145C polymorphism in susceptibility to CGD risk Conclusion: Our multigene interactive analysis suggests that ESR1, ADRB3 and ABCG8 genetic variants confer significant risk for cholesterol gallstone disease.

1706W

Bivariate heritability analyses of cardiac conduction and repolarisation measures show a paucity of shared genetic influences. *J. Alghamdi[1], C. Hastie[1], C. Schulz[1], C. Brown[1], L. Hocking[2], M. Luciano[3], D. Porteous[4], A. Morris[5], B. Smith[5], A. Dominiczak[1], S. Padmanabhan[1].* 1) College of Medical, Veterinary and Life Sciences, University of Glasgow, Institute of Cardiovascular and Medical Sciences, Glasgow, UK; 2) Aberdeen Pain Research Collaboration (Musculoskeletal Research), University of Aberdeen, Aberdeen, UK; 3) Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK; 4) Medical Genetics Section, Molecular Medicine Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; 5) Medical Research Institute, University of Dundee, Dundee, UK.

Genome wide association studies have shown predominantly concordant effects for SNPs associated with ECG conduction traits with shared physiologic processes (PR interval(PR) and QRS duration (QRSd)) and mostly discordant effects between traits that reflect different processes (QRSd and QT interval(QTc)). However, these SNPs explain a small fraction of the population variance. To estimate the proportion of shared genetic influences among ECG cardiac conduction and repolarisation measures we performed bivariate heritability in a large cohort of Scottish families. **Methods:** The Generation Scotland Scottish Family Health Study is a nationwide cohort of families with participants recruited between 2006 and 2011. Digital ECG data were coded at the University of Glasgow ECG Core Laboratory. Exclusion criteria were- use of QT prolonging medications at least two months prior to recruitment (n=619; 3.4%), QRSd>120ms, atrial fibrillation or pacemaker. 13,722 individuals (5,082 families) were included in the analyses. Univariate and bivariate heritabilities were calculated using ASREML adjusted for age and sex. **Results:** The study population included 13,722 individuals (5,082 families). The mean age was 47 (SD ±15) years, QTc 411±23 ms, QRSd 230±90.1 ms P-duration(Pd): 106.1±14.2 ms and PR 162±25.1 ms. The age- and sex-adjusted heritabilities were: QRSd - 0.38(0.02), QTc- 0.39(0.02), Pd- 0.15(0.02); PR- 0.37(0.02) (all P<0.01). Bivariate heritabilities were QRSd/QTc: 4.3%, QTc/PR: 0.1%, QRSd/PR: 5.7%, QRSd/Pd: 20%, QTc/Pd: 1.7%, PR-Pd- 19.1%. Shared genetic factors account for only 1%, 1.1% and 0.5% respectively of the phenotypic correlations - QRSd/Pd (corr=0.049), QRSd/QTc (0.254) and QRSd/PR(-0.097). **Conclusions:** Majority of the genes influencing QTc, QRSd and PR heritabilities are specific for each trait with very few genes common to them. This suggests the presence of other undiscovered variants underlying these traits.

1707T

A signal of polygenic inheritance from low frequency variants in case-control genome wide association studies elucidates genetic architecture of common diseases. *Y. Chan[1,2,3], E.T. Lim[1,2,4], N. Sandholm[5,6,7], A.J. McKnight[8], S. Ripke[2,4], M.J. Daly[1,2,4], B.M. Neale[2,4], R.M. Salem[1,2,3], J.N. Hirschhorn[1,2,3].* 1) Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America; 3) Department of Endocrinology, Children's Hospital Boston, Boston, Massachusetts, United States of America; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Massachusetts, United States of America; 5) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Helsinki, Finland; 6) Division of Nephrology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland; 7) Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland; 8) Nephrology Research, Centre for Public Health, Queen's University of Belfast, Belfast, United Kingdom.

In most complex traits and diseases, much of the heritability remains unaccounted for by common variants discovered from genome-wide association studies (GWAS). Thus it has been postulated that rare and low-frequency variants are likely to play a role in accounting for the remaining heritability. To test this hypothesis, we developed a novel method to test for polygenic inheritance specifically from low-frequency and rare variants from GWAS summary statistics. Our method utilizes the fact that there is more power to detect risk than protective variants especially for variants at lower frequencies. This would result in a higher than expected ratio of risk to protective (R/P ratio) variants at a given P-value threshold. We tested our method on published GWAS results in various complex diseases and observed signals consistent with polygenic inheritance from low-frequency variants in type 2 diabetes, schizophrenia and major depressive disorder, but not in bipolar disorder. We also used our method to explore the shared genetic component in overlapping phenotypes related to inflammatory bowel disease (Crohn's disease (CD) and ulcerative colitis (UC)) and diabetic nephropathy (macroalbuminuria and end stage renal disease (ESRD)). The signal from polygenic low-frequency variants was present in both CD and UC, and an attenuated signal persisted when both phenotypes were jointly analyzed, suggesting that although both phenotypes have distinct genetic components, they also share low frequency variants that contribute to a common genetic component in both diseases. Interestingly, while we observed a polygenic signal from low-frequency variants in both macroalbuminuria and ESRD, the signal was lost when the phenotypes were jointly analyzed, suggesting that the genetic variants underlying these two apparently related phenotypes may not overlap substantially, and that both subphenotypes should be studied separately. Thus, our method may help guide the design of future genetic studies of dichotomous traits and diseases.

1708F

Identification of 6 novel type 2 diabetes susceptibility loci using genome-wide association studies imputed from a 1000 Genomes (June 2012) reference panel. *I. Prokopenko[1], C. Ma[2], A.P. Morris[3], H. Chen[4], S.M. Willems[5], Q. Qi[6], K.J. Gaulton[3], M. Li[7], C. Ladenvall[8], N. Van Zuydam[3,9,10], J.S. Ried[11], D. Thuillier[12,13,14], Y. Lu[15], G. Thorleifsson[16], C. Fuchsberger[2], R. Mägi[17], L.J. Scott[2] on behalf of DIAGRAM+ consortium.* 1) Department of Genomics of Common Disease, Imperial College London, London, United Kingdom; 2) Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, USA; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, MA 02118, USA; 5) Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands; 6) Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA; 7) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21209, USA; 8) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University and Lund University Diabetes Centre, CRC at Skåne University Hospital, Malmö, Sweden; 9) Diabetes Research Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee DD1 9SY, UK; 10) Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee DD1 9SY, UK; 11) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 12) European Genomic Institute for Diabetes (EGID) FR 3508, 59000 Lille, France; 13) CNRS-UMR-8199, Lille Pasteur Institute, 59019 Lille, France; 14) Lille 2 University, Lille, France; 15) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 16) deCODE genetics, Reykjavik, Iceland; 17) Estonian Genome Center, University of Tartu, Tartu, Estonia.

To date more than 70 common variant associations have been described for type 2 diabetes (T2D); however these account for <10% of inherited susceptibility. Most T2D-associated variants have minor allele frequency (MAF) >10%. Our goal was to uncover novel sites for T2D by considering lower frequency variation that is not well covered by traditional genome-wide association arrays. To that end, we imputed SNP variants using the 1000 Genomes Phase I reference panel (1,092 individuals, March 2012 release, autosomal chromosomes) within European descent studies with genome-wide association (GWA) data. Following T2D association analysis with/without adjustment for body mass index (BMI), we performed a fixed-effects inverse variance meta-analysis for a total of 17 studies (15 with BMI for adjustment available) representing 19,427 (15,585) T2D cases and 49,483 (39,537) controls. Imputation using the 1000 Genomes reference increased the number of SNPs tested from the previous DIAGRAM+ analysis by >4 fold. 12.1M SNPs were present in at least 50% of the studies. We observed genome-wide significant (p-value<5×10^{-8}) associations at 23 established T2D loci, among which associations at *FTO* and *MC4R* were acting through obesity and were not significant after adjustment for BMI. Within T2D analysis adjusted for BMI, we observed novel associations at 2 common variant loci near *CENPW* (chr6: 126,792,095) and *DGAT1/SCRT1* (chr8: 145,551,200) genes as well as an independent secondary signal within *HMGA2* (chr12: 66,389,968) locus. Analysis of T2D without adjustment for BMI showed 5 novel associations at common variants near *CENPW, ETV1* (chr7:13,887,349), *PLEKHA1/ARMS2* (chr10:124,186,714), *APOE* (chr19:45,411,941) and *HORMAD2* (chr22:30,599,562) not well-tagged in previous imputations. Additionally, we observed an independent secondary association at chr19:46,307,406 within established *GIPR/QPCTL* locus. None of the lower frequency variants outside of known T2D loci was significant at 5×10^{-8}. Imputation with the large multi-ethnic 1000 Genomes reference panel has not revealed novel association signals driven by low MAF variants. However, use of high-density reference SNP set for imputation allowed identification of 6 novel common variant (all MAF>0.08) T2D signals, one of them identified within T2D analysis corrected for the effect of adiposity only.

1709W

Predictive genetic modeling in identifying metabolic syndrome development. *A.T. Kraja, M.A. Province, I.B. Borecki.* Div Statistical Genomics, Washington Univ Sch Med, St Louis, MO.

Metabolic syndrome (MetS) is a combination of risk factors for increased adiposity/ obesity, glucose intolerance/ insulin resistance, dyslipidemia and high blood pressure. To account for non-alcoholic fatty liver disease as a hepatic expression of MetS, we performed a multivariate factor analysis to explore DNA variants associations with 20 interconnected traits in the Family Heart Study, including measures of liver attenuation (LA), alanine transaminase, risk factors for MetS, several measures of fat by CT imaging and a selected number of inflammatory markers. After selecting 133 top SNPs from mixed effects additive association tests with factor scores, the list narrowed further to 45 SNPs when they were tested with each single trait and using a threshold of p-values < 10-6. Of them rs738409 of PNPLA3 associated significantly with LA (p=4.28E-25). Other candidates were also found. But there is no information available in predictive modeling with MetS candidates. We extend this work to predict based on genetic profiles who will develop MetS. We first performed online searches for identifying the pool of variants that influence MetS and its risk factors. Sources of information were: the NCBI Gene Entrez for MetS (332 candidate genes) and MetS risk factors BMI, WAIST, HDLC, TG, INS, GLUC, SBP and DBP, including biomarkers FIB, CRP, PAI-1, IL-6, ICAM-1, ADIP and WBCC. In addition, candidates for MetS (30 genes) and its risk traits from large GWAS were identified. Three main MetS domains were considered: obesity, insulin resistance and dyslipidemia. A variant/ candidate gene was considered to have a contribution to MetS, if it affected at least 2 major domains (obesity (BMI & WAIST) + dyslipidemia (HDLC & TG); glucose intolerance / insulin resistance (INS & GLUC) + HDLC & TG; BMI & WAIST + INS & GLUC). If a variant affects any 2 or more domains, then the MetS score of an individual is represented as the sum of each trait beta coefficient* recoded individual's genotype (0,1,2) * a variable weight. The weight for traits from main domains is 1.0, while for blood pressure and any additional inflammatory marker is 0.25. This creates a MetS index, which can classify individuals with high and low risk for MetS. The scoring system designed is compared with data mining classifying methods (logistic regression, partition trees, k-nearest neighbors, and radial basis machine) on candidate genotypes in the Family Heart Study time 1 and time 2 (~8 years apart).

1710T

Should we account for the random effect of relatedness when using Principal Component Analysis in GWAS? *M. de Andrade[1], J.P. Soler[2].* 1) Div Biomed Statistics & Informatics, Mayo Clinic, Rochester, MN; 2) Depertment of Statistics, University of São Paulo, SP, Brazil.

Studies of human complex diseases and traits associated with candidate genes are potentially vulnerable to bias (confounding) due to population stratification and inbreeding, especially in admixture population. In genome-wide association studies (GWAS) the Principal Components (PCs) method provides a global ancestry value per subject, allowing corrections for population stratification. However, these coefficients are typically estimated assuming unrelated individuals and if family structure is present and is ignored, such sub-structure may induce artifactual PCs. Extensions of the PCs method have been proposed by Konishi and Rao (1992) taking into account only sibship relatedness and by Oualkacha et al. (2012) which can be applied to general pedigrees and high dimensional data. In this work we apply such analysis for estimation of global individual ancestry but admitting PCs extracted from different variance components matrix estimators. For the application we use the GENOA sibship data consisting of European and African American subjects and the Baependi Heart Study consisting of 80 extended families collected from the highly admixture Brazilian population, both with SNPs data from Affymetrix 6.0 chip. All the implementation are done using R package.

1711F

Genome-wide association study of ventricular fibrillation in the setting of acute myocardial infarction. *R. Pazoki[1,2], J.S.S.G. de Jong[2], M.E. Adriaens[2], N. Bruinsma[2], L.R.C. Dekker[3], L. Beekman[2], A.A.M. Wilde[2], M.W. Tanck[1], C.R. Bezzina[2].* 1) Department of Clinical Epidemiology, Biostatistics & Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands; 2) Department of Clinical and Experimental Cardiology, Academic Medical Center, Amsterdam, The Netherlands; 3) Department of Cardiology, Catharina Hospital, Eindhoven, The Netherlands.

Sudden cardiac death from ventricular fibrillation (VF) during acute myocardial infarction (MI) is a leading cause of total and cardiovascular mortality. Although a heritable component plays a role in the determination of risk for VF in the setting of acute MI, the underlying genetic factors remain largely unknown. We previously started exploring the role of common genetic variants in modulation of VF risk by conducting a genome-wide association study (GWAS) in the Arrhythmia Genetics in the NEtherlandS (AGNES) case-control set which consists of individuals with a first acute MI, where cases suffered VF and controls not. The aim of the current study was to carry out GWAS in an extended set of AGNES patients (171 cases, 310 controls) to identify additional susceptibility loci. In total, 1458 patients with a first acute MI, 686 with VF and 767 without VF were included in the current study. The AGNES cases and controls were genotyped using the Human610-Quad and HumanOmni2.5 Illumina arrays. Quality control and principle component analysis were performed using GenABEL package in R. Non-typed single nucleotide polymorphisms (SNPs) were imputed using data from HapMap. Logistic regression models implemented in the ProbABEL program was used to assess differences in the distribution of genotypes among cases and controls. The locus we previously identified on chr21 (rs2824292) exceeded the genome-wide significance threshold for association ($p<5\times10^{-08}$). SNPs exceeding the arbitrary threshold of 1×10^{-05} are currently being meta-analyzed in an independent case-control set of individuals of European descent.

1712W

Genetic determinants of plasma levels of vitamin D binding protein in U.S. black women. *E.A. Ruiz-Narvaez[1,2], L. Rosenberg[1,2], S.A. Haddad[1,2], J.R. Palmer[1,2].* 1) Slone Epidemiology Center, Boston University, Boston, MA; 2) Department of Epidemiology, Boston University School of Public Health, Boston, MA.

Background: Vitamin D plays a key role in the regulation of a variety of metabolic processes such as calcium and phosphate absorption, xenobiotic detoxification, cell proliferation and differentiation, and immunomodulation among others. Vitamin D binding protein (DBP), the major carrier of circulating vitamin D metabolites, determines the free fraction or biologically active component of vitamin D. Genetic polymorphisms may in part determine circulating levels of DBP. **Methods:** In order to identify genetic variants that may explain variation of DBP plasma concentration, we genotyped 23 single nucleotide polymorphisms (SNPs) in vitamin D-related genes in 486 women of the ongoing cohort study Black Women's Health Study (BWHS) who had plasma measure of DBP. We also genotyped 30 ancestral informative markers (AIMs) previously shown to provide a valid estimate of percent African (vs. European) ancestry. We used general regression models with backward stepwise selection process to find the best set of SNPs that explain variation of DBP plasma levels. **Results:** A model that included the 23 vitamin D-related SNPs plus individual African ancestry explained 83% of the variation of DBP plasma concentration; five SNPs were able to explain 82% of the DBP plasma level variation. Most of the DBP variation was explained by SNP rs7041, which represents an Asp-to-Glu change in the position 432 of DBP. The SNP rs7041 explained 74% of DBP plasma concentration variation, and each minor allele was associated with an increase of 2350 nmol/L of DBP (p<0.0001). **Conclusions:** We identified 5 SNPs that explain most of the variation of DBP plasma levels in African American women. These polymorphisms, through their association with DBP concentration, might be predictors of diseases involving vitamin D deficiency.

1713T

Bayesian Polygenic Risk Prediction using Summary Statistics. *B.J. Vilhjalmsson[1,2], J. Yang[3], S. Lindstrom[1,2], A. Gusev[1,2], S. Ripke[2,4], N. Patsopoulos[2,5,6,7], R. Do[2,7], E. Stahl[8], B. Pasaniuc[9], S. Pollack[1], N. Zaitlen[10], H.-H. Won[2,4], S. Kathiresan[2,4], M.E. Goddard[11], N. Wray[3], P.L. De Jager[2,5,6,7], M. Daly[2,4], P.M. Visscher[3], P. Kraft[1,2], N. Patterson[2], A.L. Price[1,2].* 1) Department of Epidemiology, Harvard School of Public Health, Boston; 2) Broad Institute, Cambridge, MA, USA; 3) The University of Queensland, Brisbane, Queensland, Australia; 4) Massachusetts General Hospital, MA, USA; 5) Department of Neurology, Brigham & Women's Hospital, Boston, MA, USA; 6) Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA; 7) Harvard Medical School, Boston, MA, USA; 8) Icahn School of Medicine at Mount Sinai, New York, NY USA; 9) UCLA, CA Los Angeles USA; 10) UCSF, San Francisco, CA USA; 11) The University of Melbourne, Parkville, Victoria, Australia.

As sample sizes in genome-wide association studies (GWAS) of common diseases continue to grow, polygenic risk predictions of common diseases are expected to improve (Chatterjee et al. 2013 Nat Genet; Dudbridge 2013 PLoS Genet). Polygenic risk scores, defined as the sum of estimated allelic effects for markers that achieve a given significance threshold, are currently commonly applied to LD-pruned markers for prediction purposes. However, as we show here, they can be suboptimal under realistic genetic architectures. We propose a computationally efficient, Bayesian polygenic risk prediction method that uses GWAS summary statistics as training data, avoiding logistical and computational challenges of large genotype datasets. Our method adjusts the estimated marker effects by explicitly modeling the underlying distribution of causal effect as a Gaussian mixture. It also accounts for effects of linkage disequilibrium (LD) on indirect associations at non-causal markers and sampling noise in estimating marker effects. Under the assumption that there is at most one causal marker in a region of LD we derive an efficient approximate Bayesian shrink that has linear running time, GoLD (Gaussian posterior mean with LD). Notably, unlike LD-pruning, GoLD does not require any LD information from a reference panel. We show that as the sample size goes to infinity, the use of LD pruning results in a prediction r2 that fails to capture all of the heritability explained by genotyped SNPs, whereas our approach succeeds. Using simulated datasets we compare our approach to the standard thresholding approach on pruned markers and observe that GoLD can increase the prediction r2 by up to a factor 1.23 for reasonable choices of genetic architecture (random 1% of all SNPs are causal) and sample size (20,000 samples). Finally we apply the method to real datasets, including a breast cancer dataset and a schizophrenia dataset with GWAS summary statistics for 35,000 and 22,000 samples respectively. The Bayesian shrink proposed here may enable clinically relevant prediction accuracies at smaller sample sizes than other methods.

1714F

Genetic variations in telomere-maintaining genes and adult-onset cancers. *S. Hwang[1], J.M. Murabito[2], A.P. Reiner[3], A. Aviv[4], D. Levy[1].* 1) 1.National Heart, Lung and Blood Institute's Framingham Heart Study, Center for Population Studies, Framingham, MA; 2) Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts; 3) Department of Epidemiology, University of Washington School of Public Health, Seattle WA; 4) 4.The Center of Human Development and Aging, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, NJ, USA.

Background: Telomere-maintaining variation, especially rs2736100 at the *TERT* gene locus, has been shown to relate to lung, brain, bladder, and colorectal cancers. GWAS have identified SNPs at the telomerase-relating genes *TERT* and *TERC* that contribute to inter-individual differences in leukocyte telomere length (LTL). The current study tested for association between GWAS identified telomere-maintaining SNPs and cancers in a community-based cohort. Methods: Study population was composed of 5475 Framingham Heart Study participants who participated in the SNP Health Association Resource (SHARe) project and were age 45 years or older at the baseline examination. A total of 1748 subjects had a pathology confirmed malignancy including 217 breast, 257 prostate, 142 lung, 144 colorectal, 90 melanoma, and 84 bladder cancer cases. We used an additive genetic model to test whether genetic variation at the *TERT*, *TERT-CLPTM1L*, *TERC*, *NAF1*, and *OBFC1*, loci region (SNPs rs402710g, rs401681g, rs2736100, rs1006969, rs7675998, rs10936599, and rs9420907) is associated with cancer. Generalized estimation equation adjusted for familial correlation was applied to test for significance of genotype by cancer diagnosis adjusted for age, sex, body-mass-index, cigarette smoking, and multiple cancer diagnosis. Results: SNPs at the *TERT CLPTM1L* locus region were marginally associated with any cancer and significantly associated with lung cancer. For those carrying the minor allele of rs402710, a 60% lower lung cancer risk was observed (p=0.0008). Those carrying the minor allele of rs9420907 at the *OBFC1* had a 1.6 -fold lung cancer risk (odds ratio of 1.6, 95%CI 1.15-2.22, p=0.005). Conclusions and Impact: Our community-based long-term follow up study revealed an association between the telomere-mediating genetic variation and cancer risk, lung cancer. Further validation is warranted.

1715W

Genetic analysis of *TOMM40* and *APOE* for the onset of dementia in the Honolulu-Asia Aging Study. *M.W. Lutz[1], D. Goldgaber[2], D.K. Burns[3], A.M. Saunders[1], L.R. White[4], A.D. Roses[1,3].* 1) Duke University Medical Center, Department of Neurology, Durham, NC; 2) Stony Brook University, Stony Brook, NY; 3) Zinfandel Pharmaceuticals Inc., Durham, NC; 4) Pacific Health Research and Education Institute, Honolulu, HI.

Understanding the impact of genetic factors on the onset of dementia and Alzheimer's disease requires longitudinal studies with accurate determination of age of symptom and/or disease onset. Recent work by Roses et al. showed the genetic association of variants in both *APOE* and *TOMM40* with the age of onset of Alzheimer's disease in a prospectively studied, longitudinal Caucasian cohort. Specifically, an informative intronic poly-T variant (rs10524523 or '523) in the *TOMM40* gene was shown to provide more precise stratification of age of onset of Alzheimer's disease than *APOE* alone. This work was the basis for the development of a biomarker risk assignment algorithm that utilizes '523 genotype, *APOE* genotype, and chronological age to predict the risk of developing cognitive impairment within 5 years. *APOE* and '523 allele frequencies differ between ethnicities, and this presentation describes the genetic analysis of *APOE* and '523 for the onset of dementia (time course up to 18 years, median 9 years) in an elderly (>72 years) Asian male cohort (n=649), the Honolulu-Asia Aging Study (HAAS). The HAAS collected data on the development of dementia, risk factors, and neuropathologic abnormalities over a 21-year time frame. In this study, *APOE* and *TOMM40* genotypes were compared to the age of onset of dementia in this prospectively ascertained Asian series. There is a highly significant (p<0.0001) difference in '523 allele frequencies between the Caucasian and HAAS cohorts, with the most significant difference being a two-fold enrichment for the VL allele in the HAAS. VL/VL carriers have been shown to have a late age of onset for dementia: 92 years for 50% affected in both Caucasians and the HAAS cohort. '523 VL/VL carriers have the most common APOE 3/3 genotype, 61% for Caucasians and 69% for Asians. Kaplan-Meier analysis confirmed both *APOE* and '523 as predictors for the onset of dementia, with the '523 genotype providing a resolution of approximately 2-10 years between the median age of onset for dementia where 50% of the carriers of each genotype are affected. This study provides comparative data on *APOE* and '523 genotypes and age of onset of dementia that supports the generalization of the biomarker risk assignment algorithm to the Japanese ethnicity. Further work is needed to generalize the relationship between '523 genotype and age of onset of dementia across multiple ethnicities.

1716T

Molecular screening of CFTR Gene mutations in North Indian Asthmatic Children; a case-control study. *P. Dixit[1], S. Awasthi[1], N. Maurya[1], S. Agarwal[2], M. Srinivasan[2].* 1) Dept. of Paediatrics, King George's Medical University, Lucknow, India; 2) Dept. of Medical Genetics, S.G.P.G.I.M.S., Lucknow.

Background & Objective: Asthma is a complex genetic disorder. Cystic Fibrosis trans-membrane conductance regulator (CFTR) gene is an asthma susceptibility gene and carriers for this gene may develop obstructive pulmonary disease like bronchial asthma. Therefore, the objective of the study was to assess the association of CFTR gene mutations in genetic susceptibility to asthma. Material and methods: It was a hospital based case-control study, 250 bronchial asthma cases and 250 age and sex matched controls, aged 5 months-15 years were recruited for the study. Cases included were those children presenting symptoms of asthma and excluded were clinically suspected cystic fibrosis (CF) or sweat chloride level > 60 mmol/L or suffering from other respiratory diseases. Spirometery and sweat chloride test were performed in all subjects. All cases were further categorized into four different categories as per Global Initiative for Asthma criteria (GINA) guidelines: mild intermittent (83), mild persistent (96), moderate persistent (52), and severe persistent (19). Genotyping was performed for 24 CFTR gene mutations; R553X, N1303K, 621+1G>T,1717-1G→A, R668C, L997F, G576A, R75Q, R560T, V520F, Q493X, Δ1507, 3849+10KbC→T, R1162X, 3659delC, G85E, Y122X, 1078delT, 1898+1G→A, R347P, R347H, A455E, 2183AA→T, 2789+5G→A by ARMS- PCR and PCR- RFLP method. Data was statistically analyzed by Chi-square and student t test. Two-sided P value considered significant. Results: Of 250 cases and 250 controls, heterozygous allele of R553X mutation was found in 4 (1.6%) asthmatic cases and 2 (0.8%) controls. Value of FEV1/FVC ratio was significantly lower in heterozygous individuals compared to wild type homozygous individuals (P value<0.05). No significant difference was observed in the genotype and allele frequency of R553X mutation among cases and controls (P value - 0.685; Odds Ratio -1.339, Confidence Interval: 0.755-2.374). Similarly, no significant difference was observed in sweat chloride levels in heterozygous and wild type individuals among cases and controls (P value- 0.059 and 0.116 respectively). Furthermore, all wild type homozygous alleles were observed in remaining 23 CFTR gene mutations. Conclusion: We conclude that, R553X mutation of CFTR gene was not significantly associated with bronchial asthma in North Indian asthmatic population. This may imply that heterozygosity may be related with a silent obstructive pulmonary profile.

1717F

The Association of HLA-DQB1 Alleles with Nonalcoholic Fatty Liver Disease in Turkish Population. *S. Katrinli[1], L. Doganay[2, 3], Y. Colak[2], E. Senates[4], O. Ozturk[3], C. Ulasoglu[2], N. Karatas[1], I. Tuncer[2].* 1) Molecular Biology and Genetics, Istanbul Technical University, Istanbul, Turkey; 2) Department of Gastroenterology, Medeniyet University, Goztepe Teaching and Research Hospital, Istanbul, Turkey; 3) Department of Gastroenterology, Umraniye Teaching and Research Hospital, Istanbul; 4) Department of Gastroenterology, Dicle University, Medical School, Diyarbakir, Istanbul.

Introduction: Nonalcoholic fatty liver disease (NAFLD) is probably the most common liver disease in western and westernized countries, with an overall prevalence up to 30%. The spectrum of NAFLD grades from simple fatty liver (hepatic steatosis) to nonalcoholic steatohepatitis (NASH) which is its more progressive form. Hepatocellular apoptosis caused by activation of immune cells is an important feature contributing to the pathogenesis of NAFLD. Highly polymorphic human leukocyte antigen (HLA) gene is closely associated with immune response. The role of HLA class II molecules is to present antigens to CD4+ T lymphocytes which are recently shown to be increased in Nash patients (Inzaugarat *et al.*, 2011). Moreover, recent studies emphasize that HLA-DQB1 polymorphisms influence individual immune response and thus involve in the outcome of chronic viral liver diseases and drug induced liver injury. Hence, we thought that HLA-DQB1 polymorphisms may also have an effect on the progression of NAFLD and thus, we aim to investigate the association between HLA-DQB1 alleles and NAFLD in Turkish population. **Methods:** A study group of 93 biopsy proven NAFLD patients (46 male and 47 female, mean age 42±10,2 years) and a gender and age matched comparison group consisting 101 healthy people (48 male and 53 female, mean age 41±9.5) are recruited. The alleles of HLA-DQB1 are detected by sequence-specific primers-polymerase chain reaction (SSP-PCR) (Olerup *et al.*, 1993). **Results and Conclusions:** Among thirteen HLA-DQB1 alleles analyzed in this study, DQB1*0604 is significantly more frequent among the NAFLD patients compared to healthy controls (12.9% vs 2%, ψ^2=8.6, P=0.003, OR: 7.3 95% CI: 1.6-33.7). In addition the frequency of DQB1*0302 is significantly higher in the healthy control group than the NAFLD patients (24.8% vs 7.5%, ψ^2=10.4, P=0.001, OR: 0.2, 95% CI: 0.1-0.6). These results suggested DQB1*0604 might be associated with NAFLD and DQB1*0302 might be a protective allele against NAFLD.

1718W

Genetic susceptibility of FCER2 gene variants with asthma and its severity in north Indian children; a case-control study. *N. Sharma[1], S. Awasthi[1], S. R Phadke[2].* 1) Department of Paediatrics, King George's Medical University, Lucknow-UP, Lucknow, India; 2) Department of Medical Genetics, SGPGIMS, Lucknow -UP, India.

Background & objective: Several strong evidences are available for genomic factors predisposing asthma in children. Previous studies have shown that the activated low-affinity IgE receptor (encoded by FCER2 gene) plays a pivotal role in allergic immune response, regulation of immunoglobulin E (IgE) levels and other associated inflammatory mediators. Any disturbance in the delicately balanced immune response due to genetic variations in the form of single nucleotide polymorphisms (SNPs) with in FCER2 gene may result in insufficient functions of this immune modulator gene, which can enhance the susceptibility to asthma and its severity. The present study involves investigation of association of the SNPs rs2834072 and rs7249320 (located 7 bp 3'of exon 9 and intron respectively in FCER2 gene) with asthma and its severity. Methods: Case-control based genetic association study was performed among 550 children (275 asthmatic children and 275 normal controls) from sub-population of north India (Indo-Europeans linguistic subgroup + Caucasoid morphological subtype) using polymerase chain reaction-restriction fragment length polymorphism and DNA sequencing analysis. The diagnosis and classification of the asthma severity was made according to Global Initiative for asthma guideline. Data were analysed using chi-square test and logistic regression model. Results: Statistical analysis demonstrated significant association between asthma and genotypes frequency of both the SNPs rs2834072 and rs7249320. Calculated statistical values for rs2834072 were heterozygous (AG) Odds Ratio (OR) 1.417, 95% Confidence Interval (CI) 0.988 to 2.032, P= 0.058 homozygous mutant (GG) OR 2.130, CI 1.242-3.653, P= 0.006 and combined mutant (AG+GG) OR 1.521, CI 1.110 to 2.185, P= 0.0129. Similarly, significant association observed in rs7249320 for homozygous variant (CC) OR 2.767, CI 1.532 to 4.997, P= 0.001, variant allele (C) OR 1.494, CI 1.158 to 1.928 and combined mutation (AC+CC) OR 1.400, CI 1.00 to 1.959, P= 0.0494. A positive association between asthma and haplotypes AG, CA, CG were also found (P= 0.014, 0.014, 0.000 & OR 1.460, 1.475, 2.503 respectively). Interpretation & conclusions: Our data indicate that the rs2834072 & rs7249320 variations of FCER2 gene are strongly associated with asthma susceptibility and might be risk factor for asthma among north Indian subpopulation.

1719T

Significant association of Perilipin polymorphisms with LDL-cholesterol level. *I. Naka[1], R. Kimura[2], T. Inaoka[3], Y. Matsumura[4], J. Ohashi[1].* 1) University of Tsukuba, Tsukuba, Ibaraki, Japan; 2) Transdisciplinary Research Organization for Subtropics and Island Studies, University of the Ryukyus, Okinawa, Japan; 3) Department of Environmental Sociology, Faculty of Agriculture, Saga University, Saga, Japan; 4) Faculty of Health and Nutrition, Bunkyo University, Kanagawa, Japan.

Perilipins and hormone-sensitive lipase (HSL) play key roles in the cellular regulation of triglyceride deposition and mobilization. Perilipins, which are encoded by perilipin 1 gene (*PLIN1*) on 15q26, are primarily expressed at lipid droplet surface in adipocytes and steroidogenic cells. Phosphorylation of perilipins allows HSL to break down triglycerides in lipid droplets into glycerol and fatty acid chains. Several single nucleotide polymorphisms (SNPs) of *PLIN1* have been shown to be associated with levels of lipid fractions. In this study the associations of three *PLIN1* SNPs (rs2289487, rs894160, and rs1052700) and an *HSL* SNP (rs34845087) with total-cholesterol, HDL-cholesterol, LDL-cholesterol, and triglyceride levels were evaluated in 127 adult subjects living in Tonga. A multiple regression analysis adjusted for age, sex, and body mass index (BMI) revealed that a copy of rs894160-T allele significantly decreased serum LDL-cholesterol by 11.04 mg/dl (*P*-value = 0.0350) and also a copy of rs1052700-T allele significantly decreased serum LDL-cholesterol by 12.39 mg/dl (*P*-value = 0.0248). No other significant associations were observed. The rs894160 and rs1052700 were in linkage disequilibrium with each other (D '= 0.87 and r^2 = 0.73). The haplotype analysis revealed that a copy of risk haplotype (rs894160-T and rs1052700-T) was significantly decreased serum LDL-cholesterol by 12.33 mg/dl (*P*-value = 0.0193). Our results suggest that *PLIN1* polymorphisms affect the levels of serum LDL-cholesterol in Tongan adults.

1720F

Genetic risk models: model size and confidence intervals of the risk estimates. *Y. Shan[1,2], D.T. Smelser[3], G. Tromp[3], H. Kuivaniemi[3], D.E. Weeks[1,2].* 1) Department of Biostatistics, Graduate School of Public health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 3) The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA, USA.

Disease risk estimation plays a very important role in disease prevention. With more precise risk estimates, people can make more informed decisions about disease prevention and treatment. Many studies, which estimate disease risk using logistic regression models applied to a given set of risk SNPs, have found that the larger the number of risk SNPs in the risk model, the greater the ability to predict the risk. However, the width of the confidence interval (CI) is often not considered in the evaluation of the risk model. In fact, when risk estimates are based on meta analysis results, confidence intervals are often not even estimated. Furthermore, a risk estimate with a larger CI from a larger model may not be practically better than a similar risk estimate with a smaller CI from a smaller model. Here, we explore the relationship of predicted disease risk, the size of the model, and the width of the CI of the risk using both simulated data and real data. We developed a formula and conducted a simulation of how much a person's risk could change when a single SNP was added into the model. We found that the width of the CI depended strongly on the magnitude of the estimated risk. In addition, we found that the CI width appeared to be positively correlated with the size of the model at the population level, but at the individual level, the CI width did not necessarily follow the size of the model. In other words, there was larger proportion (approximately 60%) of individuals having a bigger CI than a smaller CI, when one more SNP was added into the model. This study contributes to our understanding of relationship of disease risk, size of the model and the width of CI. This work was supported by a Collaborative Research Award on Translational Genomics, part of The Commonwealth Universal Research Enhancement (CURE) program of the Pennsylvania Department of Health.

1721W

Genetic Risk Score Modeling in Age-Related Macular Degeneration.
J.N. Cooke Bailey[1], J.D. Hoffman[1], L.M. Olson[1], W. Cade[2], N. Schnetz-Boutaud[1], P. Mayo[1], M. Allen[1], A. Agarwal[3], M.A. Brantley[3], W.K. Scott[2], M.A. Pericak-Vance[2], J.L. Haines[1]. 1) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 2) Hussman Institute of Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 3) Vanderbilt Eye Institute, Vanderbilt University Medical Center, Nashville, TN.

Age-related macular degeneration (AMD) is a progressive neurodegenerative disease that is the leading cause of blindness in elderly individuals in developed countries. AMD risk is mediated by genetic and environmental factors. The majority of identified genetic risk for AMD appears to be influenced by single nucleotide polymorphisms (SNPs) in ARMS2/HTRA1 and CFH genes, which account for approximately 50% of the genetic risk for AMD. These along with several additional loci of lower size effect account for less than 65% of the genetic component of AMD. Most recent publications implicate 19 SNPs as the known mediators of AMD risk (Fritsche et al, Nat Gen 2013, 45: 433). In a Caucasian sample of 1196 AMD cases (individuals with intermediate and/or advanced [non-exudative and/or exudative AMD, AREDS grades 3-5, respectively] AMD in at least one eye) and 726 non-AMD controls, we genotyped these 19 SNPs and evaluated them for association with AMD, using a logistic regression model adjusting for age of exam and gender. SNPs associated at a Bonferroni-corrected P<0.0026 included rs10790924 in ARMS2, rs9542236 in B3GALTL , rs429608 in C2/CFB, rs2230199 in C3, and rs10737680 in CFH; several others trended toward significance. Including smoking status (available for 76% of individuals), as a covariate did not alter inferences of association. To evaluate overall genetic risk for AMD, we generated a weighted genetic risk score (GRS) based on the 19 SNPs and their published effect sizes and compared GRSs between cases and controls. We detected a significant difference in mean risk scores between the groups (P<1×10-5), indicating that these SNPs aid in differentiating between cases and controls. Interestingly, when patients with exudative AMD were stratified based on unilateral (n=378) or bilateral (n=341) disease, we detected a significant difference in cumulative risk scores between the groups (P=3×10-4, age and gender included as covariates; P= 5×10-4 when smoking status was also included as a covariate). Additionally, comparing the unilateral exudative AMD patients (AREDS grade 2-4 in the fellow eye) to those with bilateral exudative AMD, the risk score trended significantly higher as AMD severity rose in the fellow eye. These results highlight not only the utility of GRSs in AMD, but also that there may be a significant difference in genetic risk between individuals with different subtypes of this disease.

1722T

Genetic variation at the *CY2C19* gene associated with Metabolic Syndrome susceptibility in a South Portuguese population.
V. Gaio[1], A. Fernandes[2], F. Mendonça[3], F. Horta Correia[3], A. Beleza[2], A. Gil[1], M. Bourbon[1], A. Vicente[1], C. Dias[1], M. Barreto da Silva[1]. 1) Instituto Nacional de Saude Doutor Ricardo Jorge, Lisboa, Portugal; 2) Laboratório de Saúde Pública Dra. Laura Ayres, Faro, Portugal; 3) Administração Regional de Saúde do Algarve, Faro, Portugal.

Metabolic syndrome (MetS) is a cluster of conditions ~ increased blood pressure, high blood glucose level, excess body fat around the waist and abnormal cholesterol levels ~ that occur together, increasing the risk of heart disease, stroke and diabetes. In Portugal, the MetS prevalence is estimated to be 27,5% with regional variations, being highest in the Alentejo (30,99%) and lowest in the Algarve (24,42%), constituting a public health problem. Although for clinical settings, a binary definition of MetS enabling a yes or no diagnosis is useful, it is clear that dichotomizing a continuous outcome variable reduces the statistical power of the MetS association studies. Therefore, the aim of the present study is to identify genetic risk factors involved in MetS etiology, using a continuous MetS score. To achieve our goal, a principal component analysis was performed to compute a score using the six normalized risk factors for MetS (waist circumference, diastolic and systolic blood pressure, glucose, triglycerides and HDL blood levels), with a higher MetS score indicating a less favorable MetS profile. After calculating this score, an association study was performed using 37 SNPs in candidate genes involved in MetS related diseases. A total of 206 subjects, including 119 women and 87 men (mean age: 56,31± 16,37 years, range: 26-91 years) were included in this analysis. We found 4 SNPs significantly associated with higher MetS scores (rs4244285 (*CYP2C19*), rs279871(*GABRA2*), rs1647(*NPY*) and rs1142345(*TPMT*)). P-values are $4{,}36\times10^{-4}$, $1{,}3\times10^{-2}$, $1{,}7\times10^{-2}$ and $9{,}76\times10^{-3}$ respectively. After correcting for multiple testing only rs4244285 (*CYP2C19*) remains significant (p=0,016). In addition, we have performed a multiple regression analysis considering the *CYP2C19* genotype as the independent variable, adjusted for age. The resulting model explains 17% of the MetS score variance. After adding the remaining SNP genotypes that do not survive the multiple testing correction, the same model is able to explain 23,1% of the score. Our findings support the evidence of an association between *CYP2C19* rs4244285 gene polymorphism and the MetS score, emphasizing the importance of lipid metabolism, thought cytochrome P450 enzymes, in the MetS etiology. However, further studies will be necessary to replicate these findings in different populations as well as functional studies to clarify the role of this variant in the etiology of MetS.

1723F

Hearing function and loss: `A complex multistep strategy' to identify genetics and environmental factors.
G. Girotto[1], D. Vuckovic[1], A. Buniello[2], K. Steel[2], P. Gasparini[1]. 1) Medical Genetics IRCCS 'Burlo Garofolo'-DMS, Univ.Trieste , Trieste, Italy; 2) King's College London, Guy's Campus, London, United Kingdom.

The analysis of complex genetic traits/diseases such as normal hearing function (NHF) has long been an enigma of genetic biology and medical sciences and the majority of genes/loci and environmental/lifestyle factors involved still need to be detected. Here we propose a multistep approach based on: A) Genome-Wide association study (GWAS), B) expression studies, C) replication with general association studies and D) pathway analysis and E) genotype-phenotype relationship to detected genetics factors. As regards environmental factors a linear/logistic regression has been performed. For the genetic studies, we have selected a list of 19 genes recently identified in a large GWAS on different quantitative hearing traits (step A) carried out on 3417 individuals from isolated populations/communities located in Europe. These genes have been evaluated at the expression level, 12 of them showing a clear staining in the mouse cochlea of wild-type mice (at 4 and 5 days postnatal). In particular, 5 of them (Arsg, Slc16a6, Dclk1, Gabrg3, Csdm1) show strikingly specific expression in the cochlea (e.g. at the top of sensory hair cells and in the marginal cells of the stria vascularis) while the other 7 (Ptprd, Grm8, GlyBP, Evi5, Rimbp2, Ank2, Cdh13) are located in multiple cell types in the cochlea (step B). In the next step (C), 9 out of these 12 genes have been successfully replicated in independent cohorts from Caucasus and Central Asia. Moreover, 8 of these replicated genes fit in the same pathway which include a known HL gene (*MARVELD2*) (step D); finally, to look for genotype-phenotype relationship, the audiometric profiles (i.e. mean values at each frequency) of the 3 genotypes of the most associated gene-variant have been analyzed. In particular, 7 genes out of the 9 replicated (*CDH13, GRM8, ANK2, SLC16A6, ARSG, RIMBP2 and DCLK1*) showed a peculiar audiometric pattern with relevant differences for each genotype further supporting their role in NHF (step E). As regards environmental factors, among all the investigated variables (smoking, chocolate, coffee, tea, wine, beer, dairy products, spirits), only coffee consumption was associated at low and high frequencies (p=0.006) while the intake only at high frequency (p=0.003). These data demonstrate the usefulness of these combined methods in providing new insights into the molecular basis of NHF and may suggest new targets for hearing impairment treatment and prevention.

1724W

GWAS on an admixed Chilean sample of Cases and Controls to identify the genetic basis of the phenotypic variability in 22q11 microdeletion syndrome. *S. Eyheramendy[1], F. Manevy[1], M. Ramirez[1], C. Vial[2], K. Espinoza[2], J.C. Rivera[2], G. Repetto[2].* 1) Statistics, Pontificia Univ Catolic, Santiago, Chile; 2) Center for human genetics, Faculty of Medicine, Clinica Alemana, Universidad del desarrollo.

In this study we are interested in finding the genetic basis of the incomplete penetrance of the cardiovascular phenotype in 22q11 microdeletion syndrome (22q11DS) in a Chilean admixed sample of cases and controls. Genome-wide association studies (GWAS) have been a popular strategy to try to localize the genes or mutations that increase/decrease disease susceptibility in homogeneous populations. Unfortunately, the standard methodology utilized in GWAS needs to be adapted in the study of admixed populations. In order to assess for the association of a genetic variation witha disease, it is necessary to know the disease status of the individuals as well as the local ancestry of the mutations on the individuals. Chilean history shows that the general Chilean is mainly a mixture between Europeans (predominantly people from Spain but also from Germany, France, UK, Croatia, etc.) and Native Americans, in contrast with other Latin Apopulations which have an African component as well. In this study we performed a comparison of local ancestry estimation methods in terms of their performance and sensitivity towards different choices of ancestral populations in the association outcome. We consider ancestral populations from Native American, Europe and Africa to perform the comparison and show that indeed the association outcome can be sensitive to the choice of ancestral populations and local ancestry estimation algorithm. We also performed a genetic structural analysis to estimate distances between individuals of different ancestry. Funded by Fondecyt-Chile Grants 1120813, 1100131 y 1130392.

1725T

Multi-step LASSO approach identifies a chromosomal segment associated with working memory. *V. Freytag[1], L. Gschwind[1], A. Milnik[1], D. de Quervain[2,3], A. Papassotiropoulos[1,2,4], C. Vogler[1,2,4].* 1) Molecular Neuroscience, University of Basel, Basel, Switzerland; 2) Psychiatric University Clinic, University of Basel, Basel, Switzerland; 3) Cognitive Neuroscience, University of Basel, Basel, Switzerland; 4) Department Biozentrum, Life Sciences Training Facility, University of Basel, Basel, Switzerland.

Background: Working memory is a heritable trait, likely to involve a large number of variants with small effect sizes as it has been reported for other complex traits, e.g. human height. Several studies have already applied GWAS, starting to elucidate the molecular underpinnings of human working memory. Yet, a large fraction of the genetic contribution remains to be unraveled. Multidimensional approaches (MDA), capable of simultaneously modeling the effects of multiple SNPs, are a proposed solution to deal with the polygenic architecture of complex traits. MDAs can identify associations between phenotypes and a subset of markers, whose individual marginal effects cannot be detected by the standard univariate analytical approach. Due to the high number of genetic markers in comparison to the number of individuals, dimensionality reduction is often a prerequisite for the application of MDAs, which can be achieved by taking advantage of a priori biological knowledge. Here we exploit the fact that in mammalian genomes, colocalized genes have a tendency to be co-expressed and functionally related.

Methods: An initial sample of N=1039 healthy young Swiss individuals that underwent working memory assessment using the n-back paradigm was subsequently subjected to genotyping with the Affymetrix SNP Array 6.0. We followed a multi-step approach, starting with pruning SNPs for strong LD ($r^2>0.8$) on a genome-wide level. SNPs showing an uncorrected nominal ($p<0.05$) association signal were partitioned into chromosomal segments of 20 MB pairs. Within each segment a penalized regression model (LASSO) was trained on the remaining SNPs and combined with a bootstrap-based stability selection procedure to identify subsets of informative SNPs. Statistical significance of the subsets of SNPs was assessed in a second independent replication sample of N=719, by testing the correlation (unilateral Pearson's correlation test) between the model prediction and the actual observed phenotype. Resulting p-values were corrected for the number of multiple tests performed.

Results: Several subsets of SNPs reached nominal significance. On chromosome 5q34-35, a set of 46 markers exceeding a stability fraction of 0.9 survived correction for multiple comparisons (uncorr. $p < 0.00003$, corr. $p < 0.007$), suggesting involvement of this region in working memory. Among those, 19 SNPs mapped to intragenic regions, yielding 10 candidate genes for further investigation.

1726F

Genetic Dynamic Model for Temporal Quantitative Trait with both GWAS and Next-Generation Sequencing Data. *G. Getie, M. Xiong.* University of Texas School of Public Health, Houston, TX.

Traditional quantitative genetics has primarily studied traits that are measured at a specified location or time. The traits are investigated as isolated and static variables. However, in real biologic world, many quantitative traits change over time. These quantitative traits are repeatedly measured as functions of time or complete curves. In other words, these traits are observed either as continuous random functions, or on a dense grid. In the past several decades, researchers have primarily focused on (1) the role of individual genetic variation in determining the diseases and (2) the steady-state molecular and clinical phenotypes. Little attention has been paid to determining how the genetic variations and environmental perturbation act together to dynamically alter regulations and metabolism leading to the emergence of complex phenotypes and diseases. The concept of a dynamic system that originally arose from the Newtonian mechanics is now widely used as a basic framework for scientific research. Dynamic models not only focus on functional values themselves, but also focus on the rates of changes of functions. We term the rate of changes of a function as a derivative of the function. Dynamic models allow us to simultaneously model both the function itself and its derivatives that can be used as either responses or predictors. Therefore, they can capture information on both the curve and its derivatives. Genetics incorporating both functions and their derivatives into the model is referred to as dynamic genetics. There is increasing recognition and appreciation of modeling dynamics of quantitative traits and studying how the quantitative traits are evolved in response to perturbation of the genetic and environmental factors. In this talk, we will present a general dynamic genetic model for functional quantitative with both GWAS and NGS data. Two stage methods will be developed to estimate the parameters in the dynamic genetic model and test statistic will be proposed to test association of genetic variants with a temporal quantitative trait. By intensive simulations we show that the proposed test statistic has the correct type 1 error rates and much higher power to detect association than the current existing methods. The proposed method is applied to Starr County Sleep Studies where oxygen saturation of 767 individuals with 795,736 SNPs were measured for 380 minutes on average.

1727W

Does Haplotype Tests Gain More Power than Collapsing Tests in Pedigree-based Association Studies for Detecting Rare Variants. *W. Guo, Y. Yao.* Division of Intramural Division Program, National Institute of Mental Health, National Institute of Health, USA.

It has been well known that both common and rare variants contribute to complex disease etiology. Recent genome-wide association studies (GWAS) have been successfully applied in investigating the effects of common variants in explaining the genetic factors in human common diseases. However, it is still a challenge to identify the impact of rare variants, which are abundant (1 every 17 bases) in human population as the development in next-generation sequencing technologies. A number of statistical tests have already been developed to analyze collapsed rare variants identified by association tests. Here, we propose a weighted haplotype-based approach under the statistical frame of pedigree disequilibrium test, to test for the effect of rare variants with sequencing data in general pedigrees. We evaluated our method and compared it with the common collapsing strategy in the sequencing setting through extensive simulations. The new method clearly show enhanced statistical power over existing collapsing method when the variants are extremely rare, or the proportion of causal variants is large as well as the proportion of variants that increase risks. Following testing and evaluation, we have submitted weighted haplotype-based pedigree test in a computer program to the Comprehensive R Archive Network (CRAN) for general use.

1728T

Localization of causal variants at loci with multiple signals of association.
F. Hormozdiari[1], E. Kostem[1], E. Kang[1], B. Pasaniuc[3,4], E. Eskin[1,2,4]. 1) Computer Science, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA, Los Angeles, CA; 3) Department of Pathology and Laboratory Medicine, Geffen School of Medicine, UCLA, Los Angeles, CA; 4) Inter-Departmental Program in Bioinformatics, UCLA, Los Angeles, CA.

Although genome-wide association studies have successfully identified thousands of risk loci for complex traits, only a handful of causal variants have been successfully identified. Several fine-mapping studies are currently underway in an effort to pinpoint causal variants for common disease. Although several loci have been shown to contain multiple causal variants, current statistical methods for fine-mapping rely on simplistic assumptions of a single causal variant per locus to estimate posterior probability of variants being causal (WTCCC et al Nature Genetics 2012). In this work we propose a new framework for statistical fine-mapping for causal variants. As opposed to current approaches that make assumptions about the number of causal variants at a given locus, our framework considers all models with arbitrary number of causal variants to properly estimate the posterior probability of a SNP being causal. A direct benefit of our approach is that we provide well-calibrated confidence sets of SNPs that are guaranteed to contain the true casual SNPs with high confidence (i.e. 95% confidence). In other words not all the SNPs picked in our set are casual but we claim we capture the true casual SNPs with high probability (probability higher than 95%). We assessed the performance of our method using both simulated and real data. We indicate in this work the idea of peaking the top significant SNPs or using the conditional method, which are the common methods to find the casual SNP, have high false negative rate. Using the simulation data we can illustrate our methods obtain a set of SNPs which contain the true causal SNPs with 95% probability and the average number of SNPs picked by our method is only twice the number of true casual SNPs implanted in the region.

1729F

A Phenome-Wide Association Study (PheWAS) Exploration of Multiple Traits at Baseline in AIDS Clinical Trial Group (ACTG) Protocols.
C.B. Moore[1,2], A. Verma[2], D.H. Johnson[1], E.S. Daar[3], R.M. Gulick[4], R. Haubrich[5], G.K. Robbins[6], S.A. Pendergrass[2], D. Haas[1], M. Ritchie[2]. 1) Vanderbilt University, Nashville, TN; 2) Center for System Genomics, The Pennsylvania State University, University Park, PA; 3) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 4) Weill Cornell Medical College, New York, NY; 5) University California San Diego, CA; 6) Harvard University, Cambridge, MA.

Phenome-wide association studies (PheWAS) evaluate the associations between genetic variation and wide range of complex traits. We applied this approach to the baseline data from four AIDS Clinical Trial Group (ACTG) protocols using > 30 pre-treatment laboratory measures from 2547 individuals of mixed ancestry. We imputed data using Impute2 and a cosmopolitan reference panel, resulting in 5,954,294 SNPs. Due to the unique nature of this cohort, we divided the data to seek replication of our PheWAS results. We used linear regression to calculate associations for each laboratory measure, adjusting for top principle components, age, and sex. To assess replication we compared results from study A5095 to the results from combined studies A384, A5142, and A5202. A total of 1319 SNPs had replicating significant associations (test p-value ≤1e-03, replicated p-value ≤1e-01). Top results were associated with total bilirubin and mapped to UDP- glucuronosyltransferase (*UGT*), responsible for glucuronidation (top result: test p-value 5.44e-35, replicated p-value 7.69e-30). Members of the *UGT* family have been previously associated with Gilbert syndrome and hyperbiliruninemia. The second cluster of SNPs were associated with high-density lipoprotein (HDL), located on chromosome 16 near or in *CETP* (top result: test p-value 5.60e-07, replicated p-value 3.65e-07). These SNPs have been previously associated with serum cholesterol and metabolic measures. A total of 49 SNPs were associated with more than one measurement. For example, 35 SNPs were found to be significant for low-density lipoprotein measurements (LDL) and total cholesterol. These SNPs mapped to *COPG2*, *TSGA13*, and *FAR2*. The coatomer gamma subunit (*COPG2*) is a component of the protein complex required for budding from Golgi membranes. It influences the Golgi structural integrity and processing, activity, and endocytic recycling of LDL receptors. *TSGA13* is in close proximity to *COPG2*. *FAR2* encodes a dehydrogenase/reductase enzyme involved in converting fatty acids to fatty alcohols. In summary, we were able to identify robust SNP associations for multiple baseline laboratory measurements. Our results serve as a validation of whole-genome PheWAS, and show the utility of this approach for clinical studies. Future work includes PheWAS with additional ACTG data, including on-treatment data.

1730W

Fast And Robust Association Testing For High-Throughput Testing.
Y.H. Zhou, F.A. Wright. Biostatistics, University of North Carolina, Chapel Hill, NC.

In the analysis of genotype-phenotype associations, the investigator often relies on parametric testing for high-frequency variants, but must resort to exact testing or permutation analysis for low-frequency variants. Such a two-stage procedure is cumbersome, and this difficulty is compounded for datasets with numerous rare variants. We describe the moment-corrected correlation (MCC) method for association testing, which provides a close approximation to exact test p-values, but with greatly reduced computation. The approach can be used for any phenotype distribution, and covariates can be handled by residualization or using covariate stratification. MCC is very fast, and can perform the association analysis for a genome scan in minutes. MCC is also, very general and it can be applied on others type of high dimensional ‘omics data, including gene expression data from microarrays or RNA-Seq. We illustrate the wide applicability of MCC as a screening and testing tool using a variety of ‘omics datasets.

1731T

A Novel General Framework for Imaging Genetics Analysis with Next-generation Sequencing Data.
N. Lin, M. Chen, M. Xiong. University of Texas School of Public Health, Houston, TX.

Imaging Genetics is to investigate the relationship between the individual genetic variation and variation in brain wiring (connections), structure, and intellectual function. The traditional statistical methods for imaging genetic data analysis often simply summarizes three dimensional image data into an overall image measurement and tests association between a single genetic variant with the summarized simple statistic. These methods ignore image variation at positional level. In addition, the traditional statistical methods for imaging genetic data analysis are originally designed for common variants and are difficult to be applied to next-generation sequencing (NGS) data for their low power, large number of tests and high proportion of missing data. To overcome these limitations, we develop a novel general framework for imaging genetics analysis with NGS data in which we offer creative mathematical formulation of the imaging-genetic analysis as a new functional linear model with three dimensional functional responses and a functional predictor. We will take image variation at the positional level into account and use high dimension data reduction techniques to develop a statistic that collectively test association of all genetic variants within a gene or a genomic region with medical images. By intensive simulations, we demonstrate that the three dimensional functional linear model for imaging genetics analysis has the correct type 1 error rates and much higher power to detect association than the current methods. The proposed method was applied to glioblastoma (GBM) dataset where we extract MRIs of 78 GBM patients present in the Cancer Imaging Archive (TCIA) corresponding to patients in the Cancer Genome Atlas (TCGA) and RNA-seq and Exom sequencing data of 78 same GBM patients from TCGA dataset. Our preliminary results show that the proposed statistic has much smaller P-value than that of traditional method.

1732F

A Bayesian Approach to Expression Quantitative Trait Loci (eQTL) Mapping Based on Biological Pathway Knowledge. *I.S. Chang[1,2], T.Y. Chen[3], C.H. Chen[1], C.A. Hsiung[2].* 1) National Institute of Cancer Research, National Health Research Institutes, Miaoli, Taiwan; 2) Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan; 3) Institutes of Statistics, National Tsing-Hua University, Hsinchu, Taiwan.

While genome-wide association studies (GWAS) have successfully discovered and replicated thousands of SNPs associated with various traits/diseases, it is a challenge to gain biological insights regarding this association, because over 80% of them are not in coding region. One of the main approaches in the follow up of a GWAS is to examine whether these SNPs from GWAS are associated with the expression levels of certain genes. This leads to the so-called expression quantitative trait loci (eQTL) studies. In the current approaches, it is common to consider the expression data of 25K probes and as many SNPs as possible simultaneously so as to explore new phenomena or generate new hypotheses. While intuitive, these approaches often ignore the correlation between the gene expressions and do not make use of biological pathway information. We propose a Bayesian approach to eQTL that alleviate these concerns. In particular, the prior distribution makes use of the classical heritability concept in genetics and the pathway information from biological databases like GO, KEGG, BIOCARTA, etc. The former helps to avoid the often too conservative practices in genomic studies and the latter helps to provide biological interpretation and avoid reproducibility issue. In addition, the true-path rule is considered in the model construction, which says that if a gene is associated to a pathway, it is also associated to all the pathways along the path up to the root in the database. A carefully designed MCMC algorithm is proposed to sample the posterior distribution for inference. Our algorithm can handle 25K expression probes and 1000 SNPs. Simulation studies and real data analyses are conducted to evaluate the performance of this method.

1733W

A Novel Exact Test for Association for Small Sample Case-Control Studies. *L. Ehwerhemuepha, S. Alexandria, C. Rakovski.* Chapman University, Orange, CA.

We propose a novel exact test for association between a multiallelic marker and a phenotype with small sample case-control data. In these settings, the case-control genotype data follow multinomial distributions but classical large-sample chi-square contingency table methods are not applicable. The approach enumerates all samples under the non-completely specified null hypothesis of equality of the underlying multinomial distributions and calculates p-values as the sum of the probabilities of the samples as likely or less likely to occur than the observed data. We performed an extensive simulation study to assess the type I error rates under various null hypothesis and at several alpha levels. The number of all possible samples is a product of large binomial coefficients that make the full calculation computationally intensive. Thus, we developed a fast version of the algorithm that reduces the computational time by a factor of 1000 based on the idea of a selective removal of samples with very low probabilities while controlling the precision of the estimated exact p-values. Our results show that the new method possesses a conservative type I error in all scenarios due to the absence of adjustment in this nonparametric technique for the estimation of common multinomial probabilities under the null. It is a viable association approach that attains moderate power to detect deleterious mutations with very large effect sizes even with small sample data. The proposed method is readily extendable to haplotype data and even multimarker genotype data with haplotypes phase uncertainty. In the latter case, the EM algorithm can be implemented to determine weighted haplotype assignments of all subjects based on the haplotype pairs compatible with the unphased genotypes.

1734T

SNPx: fast testing of SNPxSNP interactions using only summary statistic data. *S. Bacanu, T. Bigdeli, D. Lee.* Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.

Analyses of genome wide association studies (GWASs) and meta-analyses have mainly focused on detecting univariate association between trait and single nucleotide polymorphisms (SNPs). However, SNPxSNP interactions are likely to play important role in the etiology of complex phenotypes/diseases and, thus, detecting such interactions might provide new insights into the biological basis of these phenotypes. Nonetheless an exhaustive testing for these interactions has been prohibitively difficult due to i) the lack of access to individual level genotype data used in GWASs/meta-analyses and ii) extremely heavy computational burden required for a) the imputation of unmeasured SNPs and b) testing of all pairwise SNP interactions. To alleviate the burden of testing SNPxSNP interactions, we propose a novel method/software for detecting interaction between SNPs in a genome scan (GWAS or whole genome sequence scan), by using only summary statistics. The proposed method employs the conditional expectation formula for multivariate normal variates to impute i) summary statistics of unmeasured SNPs and ii) interaction statistics between all nominally significant SNPs. Based on these univariate summary statistics and their correlations, as estimated from a relevant reference panel (e.g. 1000 Genomes and UK10K), we construct the omnibus SNPxSNP test. We applied the method to summary statistic data from the Psychiatric GWAS Consortium (PGC) stage 1 and detected a number of interesting and unexpected interactions among SNPs in various pathways.

1735F

Efficient and powerful set tests using phenotype prediction. *C. Lippert, C. Kadie, J. Listgarten, D. Heckerman.* Microsoft Research, Los Angeles, CA.

Set tests are rapidly gaining in importance, proving useful for testing associations between phenotypes and sets of (possibly rare) variants defined by genes, pathways, or regions of the genome. We introduce a new methodology for such tests that yields improved power over widely-used kernel-based methods (such as SKAT), while controlling for type-I error. The test we consider uses a test statistic that is the difference in out-of-sample (cross-validated) phenotype prediction accuracy between using all and no variants in the set. This statistic has been used previously, but obtaining *P* values has been computationally inefficient for genome-wide testing, because it has required a huge number of permutations for reliably estimating the extreme tail of the null distribution, the region of most interest. Here, we introduce a substantial speed up of the permutation approach, wherein test statistics for a small number of permutations are computed (typically, 10 permutations per test), and then the tail of the distribution of pooled test statistics is fit to a simple parametric null distribution. One example of a parametric distribution that we have found to work well in practice is a mixture distribution between a constant zero component and a gamma distribution. We illustrate the approach on several synthetic and real datasets, including data from the Wellcome Trust Case Consortium. The approach is extremely general as it can be used with various linear and nonlinear classification/regression models and various measures of prediction accuracy and does not rely on asymptotic assumptions.

1736W

Analyzing genome-wide associations with high dimensional phenotypes in the GALA II study. *C.R. Gignoux[1], J.M. Galanter[1], K.A. Drake[1], H. Aschard[2], D.G. Torgerson[1], L.A. Roth[1], S.S. Oh[1], P. Kraft[2], C.D. Bustamante[3], N.A. Zaitlen[1], E.G. Burchard[1], The GALA II Investigators.* 1) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 3) Department of Genetics, Stanford University, Stanford, CA.

Studies of complex traits often measure a large number of phenotypic variables beyond the specific measurement of interest. Ideally these can help power studies of disease severity and identify potential disease heterogeneity. However, these measurements often contain missing data and can be highly correlated, noisy and difficult to interpret on their own. To identify genetic associations with high dimensional phenotypes, we leveraged an extensive number of measurements in the GALA II study of childhood asthma in Latinos.

We used a modified principal component analysis (PCA) capable of handling missing data to decompose the 75 phenotypes into uncorrelated components and performed a genome-wide association study for each PC using standard linear regression. We then performed a meta-analysis across all PC-GWASes to identify variants associated with any phenotype variables.

We found that for most PCs, scores were drawn from continuous distributions. Importantly the structure was not defined by a small number of PCs: PC1 explained ~8% of the total variance, while the top 10 PCs together still explained less than 40%. Yet the orthogonality ensured that PCA separated different components of the disease process: the top three PCs related largely to severity of disease (including frequency of symptoms and exacerbations), lung function, and anthropometrics (age, height, weight, and obesity). We also identified lower PCs emphasizing other aspects of atopic disease (e.g. eczema/skin rash).

We identified several genome-wide significant hits in our combined GWAS across all PCs. These variants tended to be less common (frequencies of 1-5%), and clustered in several pathways relevant to asthma - including a variant in *SMAD3*, a gene in the *TGF*-β signaling pathway previously associated with asthma in Europeans and European Americans. Additional associations involved the trafficking of toll-like receptors and a gene targeted in COPD treatment. Importantly, identifying these biologically relevant genes requires joint examination of the full spectrum of measurents in GALA II. By taking advantage of the data-rich sampling performed in GALA II we were able to efficiently identify novel significant variants important in the disease process, even across a large number of phenotypes.

1737T

A flexible association model for the analysis of multiple diseases, and its application to inflammatory bowel disease. *L. Jostins, G.A. McVean, International IBD Genetics Consortium.* Wellcome Trust Centre for Human Genetics, University of Oxford, UK.

There is growing interest in understanding the shared and distinct genetic risk pathways that underlie different human diseases. However, these investigations can be confounded by differential power and control sharing, making loci falsely appear distinct and shared respectively. To address these issues we need a cross-phenotype model that is capable of combining evidence across multiple diseases for both locus discovery and fine-mapping, and testing specific hypotheses about locus sharing (e.g. showing a signal to be phenotype-specific). It must also be robust to differences in power, sample size and control sharing between cohorts, and be computationally efficient.

We propose a Bayesian multinomial logistic regression approach that fulfills the above criteria. Diseases are modeled as independent categories, with per-disease log odds ratios drawn from a multivariate normal prior representing correlations in risk across diseases. Global model parameters are fitted across all established loci, and are in turn used to calculate per-locus Bayes factors (BFs) summarizing the combined evidence for association, and for locus sharing or phenotype specificity. For tens of thousands of samples these calculations run genome-wide in a few hours on a laptop. The method allows regression techniques such as covariates, conditional analysis and pairwise epistasis (e.g to test for phenotype-specific interaction).

We apply this method to 30,837 inflammatory bowel disease (IBD) patients and 22,442 controls genotyped using the Immunochip by the International IBD Genetics Consortium. IBD is largely made up of two diseases (Crohn's disease, CD, and ulcerative colitis, UC): 101 loci were shown with high confidence (BF > 10) to be shared by both forms, compared to 12 CD specific, 14 UC specific and 36 inconclusive loci. For example, a SNP upstream of *GIPC2*, previously believed to be CD specific, was shown to be shared. While the UC association alone was not significant (OR=1.05, $p = 0.07$, compared to OR = 1.13, $p < 10^{-5}$ in CD), the joint analysis showed strong evidence of sharing (BF > 100). We also use this method to test for differential enrichment of functional terms (GO and KEGG) in CD and UC. Finally, we fine-map 78 densely genotyped loci by simultaneously considering data from both CD and UC, and compare the results to standard fine-mapping techniques.

1738F

Association study of genetic polymorphisms of SLC5A11 and ISYNA1 gene with the risk of Neural Tube Defects in a high-risk Chinese population. *J. Guo, JH. Wang, XW. Wang, C. Ji, Z. Guan, Q. Xie, ZQ. Zhu, B. Niu, T. Zhang.* Capital Institute of Pediatrics, Beijing, China.

Neural tube defects (NTDs) are among the commonest and most severe disorders of the fetus and newborn. Maternal nutritional status is associated with the occurrence of NTDs. Myo-inositol (MI) deficiency during embryonic development can cause cranial NTDs in both rats and mice, and some human NTD pregnancies also have lower maternal MI concentrations than unaffected pregnancies. In the inositol metabolic pathway, Solute carrier family 5 (Sodium/Glucose Cotransporter), Member 11(SLC5A11) is exclusively responsible for apical MI transport and absorption in intestine and the encoded protein of Myo-Inositol 1-phosphate synthase A1 (ISYNA1)plays a critical role in the myo-inositol biosynthesis pathway by catalyzing the rate-limiting conversion of glucose 6-phosphate to myoinositol 1-phosphate. Our hypothesis is that SNP of SLC5A11 and ISYNA1 gene may affect the maternal MI concentration and then disturbances the process of neural tube close. In the present study, we carried out a case-control study in a Chinese population of Shanxi Province, a high-risk area for NTDs. The case group was consisted of 150 pregnant women with NTDs and there were 279 pregnant women with normal newborns in the control group. Totally 8 tagSnps of SLC5A11 and ISYNA1 gene were genotyped by the MassARRAY platform and analyzed. The result showed the all the 8 SNPs in the present study were in Hardy-Weinberg equilibrium. In SLC5A11 gene, CC+CT genotype of rs274077 decreased risk for NTDs compared with those harboring the TT genotype (OR=0.64, 95%CI: 0.42-0.96, P = 0.029). After stratifying NTDs group into anencephalus, spina bifida and cenencephalocele, AG+GG genotype of rs8057788 increased 3.24 times risk of NTDs compared with AA genotype (OR=3.24, 95% CI, 1.11-9.47, P = 0.023). In ISYNA1 gene, none of the SNP was associated with the NTDs risk. Then we predicted the binding capacity to the potential transcription factor of the different genotype of the above 2 SNP using the bioinformatics method. The rs274077 is located in the conserved sequence of signal transducers and activators of transcription (STATx), and rs8057788 is located in the conserved sequence of transcription factor c-Rel and NF-kappaB. Our result suggested the polymorphism in SLC5A11 may affect the transcription factor binding ability, therefore, might regulate the expression and transcription of SLC5A11 gene to increase the risk of NTDs.

1739W

Established susceptibility loci do not infer cognitive impairment as measured by TICS-M in multiple sclerosis patients. *M.F. George[1], E. Elboudwarej[1], F.B.S. Briggs[1], H. Quach[1], R. Whitmer[2], L. Shen[2], A. Bernstein[3], C. Schaefer[2], L.F. Barcellos[1,2].* 1) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, CA, USA; 2) Kaiser Permanente Division of Research, Oakland, CA, USA; 3) Palm Drive Hospital, Petaluma, CA, USA.

The prevalence of cognitive impairment may be as high as 70% in individuals with multiple sclerosis (MS), and has been shown to affect social and emotional function, maintenance of employment, and overall quality of life. The identification of genetic variants that predict severe clinical MS outcomes is critical to understand disease mechanisms and guide development of effective therapeutics. We investigated the association between established risk loci and disease progression measured by cognitive impairment. In total, 932 white MS patients and 576 controls identified from the Kaiser Permanente Medical Care Plan, Northern California Region were studied. A weighted genetic risk score (wGRS) combined the weighted odds ratio (OR) from each of 52 established MS risk loci. The weight for each variant was the natural log of the OR for each allele determined from a previous GWAS (N≈30,000, Nature 2011). An unweighted genetic risk score (GRS) was calculated to capture variant allele counts per person. HLA-DRB1*15:01 carrier status was also examined. Cognitive status was determined using a validated telephone interview cognitive status (TICS-M) assessment tool, and accounted for level of education. Each participant was questioned regarding orientation; registration and free recall; attention and calculation; comprehension; language and repetition; and delayed recall. These seven areas of cognition were also examined as sub-scores. Linear regression models were utilized; adjusted models controlled for age, gender and known environmental risk factors. As expected, overall cognitive score was lower among cases compared to controls ($p=4\times10^{-4}$), as well as several sub-scores (orientation $p=3\times10^{-5}$; registration p=0.013, comprehension p= 0.048, and delayed recall $p=2\times10^{-3}$). When testing in cases only, no association between individual risk variants, wGRS, GRS, or HLA-DRB1*15:01 and cognitive score was observed. This study is the first to investigate the relationship between genetic risk scores (weighted and unweighted) and established MS risk loci on cognitive status in MS cases. Results suggest genetic factors contributing to cognitive impairment in MS are different from those that predispose to disease onset.

1740T

The role of rare genetic variants in host genetic control of anti-mycobacterial immunity. *J. Manry[1,2], A. Cobat[1,2], E. Schurr[1,2].* 1) McGill International TB Centre, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 2) Departments of Medicine and Human Genetics, McGill University, Montreal, Quebec, Canada.

The tuberculin skin test (TST) measures the intensity of anti-mycobacterial acquired immunity and is used to diagnose latent infection (i.e. without clinical symptoms) with Mycobacterium tuberculosis. In a previous genome-wide linkage study, we identified two loci that have an impact on TST reactivity. The TST1 locus impacts on TST zero versus nonzero while TST2 impacts on the intensity of the developing immune response. Present experimental data do not support a strong effect of variants with a minor allele frequency (MAF) > 2% on both phenotypes leading to the conclusion that TST responses are under control of rare variants. Exome sequencing was performed on 96 individuals from the founder generation of the families that led to the detection of the TST1 and TST2 linkage peaks. A bioinformatic pipeline for variant discovery was setup, and 2461 variants were identified in genes located within the two linkage intervals (1305 in the TST1 locus and 1156 in the TST2 locus). To reduce the number of candidate variants, several filters were applied to identify those variants that are most likely to impact on the two phenotypes. In a first step, the exclusion of variants that corresponded to synonymous amino acid changes or that were not in Hardy-Weinberg equilibrium allowed us to reduce the list to 430 variants. Next, variants that had previously been genotyped in a high resolution scan of common variants of the linkage intervals were eliminated reducing the list to 351 variants. Interestingly, 239 of these 351 variants were found in families contributing substantially to linkage (maximum LOD score for the family > 0.1). Indeed, 51 variants in 26 genes in the TST1 locus and 25 variants in 11 genes in the TST2 locus were only present in those families.. This genotyping effort will allow us to identify genes with enrichment of variants that segregate with TST reactivities.

1741F

The Empirical Assessment of Statistical Power of Rare Variant Association Methods. *K. Hao[1,2,3], H. Chen[4], H. Zhou[4], C. Molony[4], H. Dai[3,4].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Asian Cancer Research Group, Inc., Wilmington, DE; 4) Merck Research Laboratories, Boston, MA.

Backgrounds The role of rare genetic variation in the etiology of complex disease remains as an active research field. The next-generation sequencing technologies were powerful in identifying genetic variants responsible to Mendelian disorders, but had limited success in dissecting complex diseases. Several novel statistical methodologies have been recently proposed to assess the contribution of rare variation to complex disease etiology. So far, their statistical power was assessed mainly on simulated phenotypes. Methods We quantified the statistical power of popular rare variants association methods empirically using the eQTL framework. That is the relative statistical power is proportional to the number of eQTLs captured at fixed FDR. Results On N=100 human non-tumor liver tissues, we measured gene expression trait with Affymetrix HU133 Plus2 chip, and conducted whole genome DNA sequencing (WGS) at 30X coverage. Also, we measured the genotype using Illumina 650Y array. Results of conventional GWAS methods with 1000G imputation was compared to those of rare variants association methods, including cohort allelic sum test (CAST), weighted sum test (WST) and SKAT. When only focusing on SNV of MAF<5%, none of the rare variants association methods have meaningful statistical power. When including all SNVs, rare variants association methods offer comparable power to single marker test (ie, GWAS methods) Conclusion WGS has no advantage in term of statistical power over SNP array in identifying eQTLs. Gene-level tests offer higher power due to less multiple testing penalties. SNVs on exons, splicing site or UTRs were more likely to control gene expression levels.

1742W

Enhanced ability to replicate findings as a rationale for conducting marker-set tests. *J. Arbet[1], K. Grinde[2], C. Fu[3], A. Benitez[4], M. O'Connell[5], N. Tintle[6].* 1) Winona State University, Winona, MN; 2) St. Olaf College, Northfield, MN; 3) Massachusetts Institute of Technology, Cambridge, MA; 4) Brown University, Providence, RI; 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

Over the last decade numerous approaches have been proposed to simultaneously analyze multiple individual variants for association with a phenotype. These approaches span methods which combine rare variants (minor allele frequency less than 5%), common variants, rare and common variants and pathway based approaches. All of these methods hope to provide improved power versus single variant approaches by (a) reducing multiple testing penalties and (b) aggregating evidence from multiple causal variants by creating biologically meaningful sets. While simulation studies have shown that for particular genetic architecture and based on the proportion of non-causal variants, multi-marker approaches can provide improved power in a discovery (stage 1) analysis, little work has explored the impact of single vs. multi-marker tests on the power to replicate findings in an independent (stage 2) study. We present general results for power in independent samples when using multiple marker tests vs. single marker tests, and demonstrate that the advantages of multiple marker approaches are a function of the number of markers combined, the proportion of non-causal variants, the strength and direction of the causal variant association with the phenotype and the multiple testing penalty being used. Practical guidance on use of multiple marker tests with a goal of maximal power for replication is provided.

1743T

Maximizing the power in Principal Components Analysis of Correlated Phenotypes. *H. Aschard[1], B. Vilhjalmsson[1], C. Wu[1], N. Greliche[2], PE. Morange[3], B. Wolpin[4], DA. Tregouet[2], P. Kraft[1].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) INSERM UMR_S 937, ICAN Institute for Cardiometabolism And Nutrition, Pierre et Marie Curie University, Paris 6, France; 3) INSER UMR_S 1062, Aix-Marseille Université, Marseille F-13385, France; 4) Dana-Farber Cancer Institute, department of Medical Oncology, Boston, USA.

Background: Principal Component analysis (PCA) is a useful tool that has been widely used for the multivariate analysis of correlated variables. It is usually applied as a dimension reduction method: the few top principal components (PCs) explaining most of total variance are tested for association with a predictor of interest, and the remaining PCs are ignored. This strategy has been widely applied in genetic epidemiology, however some aspects of this analytical technique are not well appreciated in the context of single nucleotide polymorphisms (SNPs) testing.

Method: We reviewed some of the theoretical basis and behavior of PCA when testing for association between a SNP and two correlated traits under various scenarios. We then evaluated through simulations the power of a few different PCA-based strategies when analyzing up to 100 traits. We applied these methods for the genome-wide association test of five coagulation traits in 685 subjects from the MARTHA study, and the association test of 700 candidate SNPs with 79 metabolites levels in 1190 individuals from four U.S. prospective cohorts.

Result: We show that contrary to widespread practice, testing the top PCs only can be dramatically underpowered since PCs explaining a low amount of the total phenotypic variance can harbor a substantial part of the total genetic association. We also demonstrate that PC-based strategies can only achieve a moderate gain in power in the presence of positive pleiotropy, but have great potential to detect negative pleiotropy (e.g. positive correlation and opposite genetic effects) or genetic variants that are associated with a single trait highly correlated to others. Real data applications confirm these results: the combined analysis of the five PCs from the coagulation traits identified two new potential candidates SNPs, which had strongest associations with the 5th PC, while the combined analysis of half of the PCs from the metabolite traits explaining the less of the total variance (12% of the total when combined) identified two new variants that likely affect a single trait.

Conclusion: We identified major improvements to standard PCA-based strategies for the analysis of correlated traits. Their implementations led to the identification of new genetic variants that would have been missed by standard approaches.

1744F

What now? Post-hoc approaches for gene-based, rare variant tests of association. *A. Benitez[1], J. Arbet[2], K. Grinde[3], C. Fu[4], M. O'Connell[5], N. Tintle[6].* 1) Brown University, Providence, RI; 2) Winona State University, Winona, MN; 3) St. Olaf College, Northfield, MN; 4) Massachusetts Institute of Technology, Cambridge, MA; 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

To date, rare variant testing approaches have focused on maximizing statistical power to identify genes showing significant association with disease. The statistics can accommodate a combination of risk, protective, and non-causal variants and can weight each variant. Although increasingly complex test statistics and weighting strategies may improve power, they also may make interpretation of a significant association result more difficult. Determining which variant(s) in the gene are causal and estimating their effect is crucial toward planning replication studies and characterizing the genetic architecture of the locus. This problem is analogous to the situation in a one-way ANOVA analysis, where evidence in the omnibus test says that 'at least one group mean is different than the others,' but it takes post-hoc analyses to identify specific group means as being different than others. Recent work by our group has classified general characteristics of two large classes of gene-based rare variant tests. Using this framework, we have explored the ramifications of choice of gene-based test statistic on post-hoc analyses attempting to identify causal variants. Furthermore, we have evaluated the overall quality and consistency of different single marker association statistics (e.g. single marker test p-values, difference in allele counts between cases and controls, relative risk estimates) in identifying the most likely causal variants within a gene. Finally, we have explored approaches to identify the most likely subset of causal variants within a gene, when incorporating a priori biological information (e.g., position, function). In this presentation we will present the results of our findings, ultimately providing a set of best practices for applied researchers.

1745W

FARVAT: a fast and efficient rare variant association Tool for dichotomous trait with extended families. *S. Choi[1], S. Lee[1], M.M. Nöthen[2], C. Lange[3], T. Park[1,4], S. Won[5].* 1) Bioinformatics Program, Seoul National University, Seoul, South Korea; 2) Institute of Human Genetics, University of Bonn, Bonn, Germany; 3) Dept of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 4) Dept of Statistics, Seoul National University, Seoul, South Korea; 5) Dept of Applied Statistics, Chung-Ang UNiversity, Seoul, South Korea.

Family-based samples are genetically more homogeneous than the population-based samples and the analysis of rare variant with family-based samples have been often recommended. However in spite of the importance of family-based samples for rare variant association analysis, few statistical methods have been suggested. In this report we propose a new statistical method for the analysis of rare variant with extended families. The proposed method is robust against the population substructure, and can be applied to both binary and continuous traits. Depending on the choice of working matrix, our method can be a burden test or variance component test, and this approach is extended to the SKAT-O statistic. We empirically showed that the proposed methods perform better than the existing methods such as PedCMC and FPCA (MC Wu et al. AJHG, 2011). Furthermore the time complexity for the proposed method is O(M3 + N2M) and we found that the analysis of the whole genome sequence data for 1000 individuals in extended family design can be conducted within a few hours. The proposed method is implemented in C, and the proposed method was applied to gene-based analysis of Schizophrenia. All samples were collected from Germany, and the whole genomes for 36 trios which consist of affected offspring and unaffected parents were sequenced. There were 9,216,373 bi-allelic variants available, and 31,046 variants MAF of which minor allele frequencies were less than 0.05. Using 13,053 impact genes, we identified several candidate genes at the 5% significance level after Bonferroni adjustment, which are related to schizophrenia.

1746T

8q24 Risk Alleles and Prostate Cancer in African-Barbadian Men. *C.D. Cropp[1], C.M. Robbins[2], A.J.M. Hennis[3], J.D. Carpten[2], L. Waterman[3], R. Worrell[3], J.M. Trent[2], M.C. Leske[4], S.Y. Wu[4], J.E. Bailey-Wilson[1], B. Nemesure[4].* 1) Statistical Gen Branch, IDRB/NHGRI/NIH, Baltimore, MD; 2) Integrated Cancer Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ; 3) Department of Biological & Chemical Sciences, University of the West Indies, Bridgetown, Barbados; 4) Department of Preventive Medicine, Stony Brook University Medical Center, Stony Brook, NY.

African American men (AA) exhibit a disproportionate share of prostate cancer (PC) incidence, morbidity and mortality compared to other groups. Several genetic association studies have implicated select loci in the 8q24 region as increasing PC risk in AA. We evaluated the association between previously reported 8q24 risk alleles and PC in African-Barbadian (AB) men, also known to have high rates of PC. Ten previously reported tag SNPs were genotyped in 447 AB men with PC and 385 AB controls from the Prostate Cancer in a Black Population (PCBP) study. Only rs2124036 was nominally significant in AB men, (OR = 2.0, 95% CI (1.0-4.3), P=0.06) for the homozygous C/C genotype after correction for multiple testing. We also conducted a meta-analysis including our AB population along with two additional African-Caribbean populations from Tobago and Jamaica for SNPs rs16901979 and rs1447295. A significant association resulted for the rs16901979 A allele (Z score 2.75; p=0.006; summary OR= 1.21 (95% CI: 1.01-1.46)). Our findings may indicate: i) the presence of a founder effect; ii) the selected SNPs not being tagged to an ancestral haplotype bearing the 8q24 risk allele(s) in this population; or iii) inadequate power to detect a true association. Additional GWAS and sequencing studies are underway to further interrogate any potential contribution of the 8q24 region to PC in this West African-derived population.

1747F

Powerful methods for including genotype uncertainty in tests of Hardy-Weinberg Equilibrium. *C. Fu[1], M. O'Connell[2], A. Benitez[3], J. Arbet[4], K. Grinde[5], K. Liu[7], A. Luedtke[8], A. Beck[9], N. Tintle[6].* 1) Massachusetts Institute of Technology, Cambridge, MA; 2) Miami University of Ohio, Oxford, OH; 3) Brown University, Providence, RI; 4) Winona State University, Winona, MN; 5) St. Olaf College, Northfield, MN; 6) Dordt College, Sioux Center, IA; 7) Harvard University, Cambridge, MA; 8) University of California - Berkeley, Berkeley, CA; 9) Loyola University Chicago, Chicago, IL.

Most genotyping technologies now generate posterior probabilities reflecting genotype uncertainty. Increasingly, methodological work seeks to utilize these posterior probabilities in downstream statistical analysis. In many cases, use of posterior probabilities (e.g., the dosage) can yield improved statistical power. Recent work by our group and others has shown that for standard single marker tests of association use of the dosage provides a nearly optimal use of posterior probabilities. However, prior to association testing, a standard quality control step in the analysis of SNP genotypes is to test for departures from Hardy-Weinberg Equilibrium (HWE). In this presentation we propose two novel strategies for testing HWE using posterior probabilities, and compare to a third recently proposed option. We demonstrate that these novel strategies provide improved power and proper control of the type I error rate as compared to existing methods.

1748W

A clustering approach for mapping rare variants based on mutual association. *S. Ghosh, S. Deb.* Human Genetics Unit, Indian Statistical Institute, Kolkata, India.

Although genome-wide association studies have successfully identified a large number of common variants underlying various complex disorders, a substantial proportion of the total genetic variation in a trait still remains unexplained. It is becoming increasingly evident that the 'Common Disease Common Variant' paradigm needs to be modified and the 'missing heritability' may possibly be explained by rare variants that could not be identified using genome-wide association studies. However, the major impediment in identifying rare variants is that one would require huge sample sizes to detect differences in allele frequencies between cases and controls. Thus, most existing methods are based collapsing multiple variant sites using different statistical algorithms. Motivated by the combined multivariate and collapsing (CMC) algorithm, we develop a clustering mechanism of rare variant sites, but based on their mutual extent of association rather than similarity in allele frequencies as proposed in CMC, thereby reducing the possibility of combining functional and non-functional variants. We use the Fisher's exact test to identify blocks of variant sites such that the initial site in each block is associated with all other sites in the block. The test for case-control association is then performed within each block by comparing the proportions of affected and unaffected individuals carrying at least one copy of a rare variant and is based on a variance stabilizing sine transformation. We carry out extensive simulations under different rare variant models and compare the false positive rate and the power of our proposed method with some of the popular competing methods: CMC, adaptive SUM, WSS, TestRare, RareCover and Kernel-based Adaptive Clustering. We find that the proposed test procedure yields more power than the existing approaches, especially with increasing sample size, while maintaining the correct size.

1749T

Evaluating the impact of genotype errors and uncertainty on gene-based rare variant tests of association. *K. Grinde[1], C. Fu[2], J. Arbet[3], A. Benitez[4], M. O'Connell[5], N. Tintle[6].* 1) St. Olaf College, Northfield, MN; 2) Massachusetts Institute of Technology, Cambridge, MA; 3) Winona State University, Winona, MN; 4) Brown University, Providence, RI; 5) Miami University of Ohio, Oxford, OH; 6) Dordt College, Sioux Center, IA.

The new class of gene-based rare variant tests of association have usually been evaluated assuming error-free genotype information. In reality, rare variant genotypes, whether from next generation sequencing or imputation will reflect genotype uncertainty (e.g., genotype likelihoods, dosages or errors), and, ideally, subsequent rare variant tests should be robust to this uncertainty. Errors and uncertainty in SNP genotyping are already known to impact statistical power for single marker tests on common variants and, in some cases, inflate the type I error rate. Recent results show that uncertainty in genotype calls derived from sequencing reads are dependent on several factors, including read depth, calling algorithm, number of alleles present in the sample, and the frequency at which an allele segregates in the population. Imputation accuracy of rare variants is dependent upon the frequency of the allele to be imputed, as well as the size of the reference panel and its genetic relatedness to the study sample. We have recently proposed a general framework for the evaluation and investigation of gene-based rare variant tests of association. We have now extended this framework to incorporate rare variant genotype uncertainty and error models to precisely relate factors affecting genotype uncertainty to the power and type I error rate of rare variant tests. This work provides a realistic framework for assessing power and type I error and suggests genetic disease model, study design (e.g., imputation? sequencing depth?) and test statistic combinations that are particularly impacted by or resistant to genotype uncertainty.

1750F

Meta-Analysis of Gene-Level Associations for Rare Variants Based on Single-Variant Statistics. *Y.J. Hu[1], S.I. Berndt[2], S. Gustafsson[3], A. Ganna[3,4], J. Hirschhorn[5,6,7], K.E. North[8], E. Ingelsson[3,9], D.Y. Lin[10], Genetic Investigation of ANthropometric Traits (GIANT) Consortium,.* 1) Biostatistics and Bioinformatics, Emory University, Atlanta, GA.Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 30322, USA; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA; 3) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University Hospital, 751 85 Uppsala, Sweden; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden; 5) Divisions of Genetics and Endocrinology and Center for Basic and Translational Obesity Research, Children's Hospital, Boston, Massachusetts 02115, USA; 6) Metabolism Initiative and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA; 7) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; 8) Department of Epidemiology and Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC 27599-8050, USA; 9) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK; 10) Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7420, USA.

Meta-analysis of genome-wide association studies (GWAS) has led to the discoveries of many common variants associated with complex human diseases. There is a growing recognition that identifying̀ causal'' rare variants also requires large-scale meta-analysis. The fact that association tests with rare variants are performed at the gene level rather than at the variant level poses unprecedented challenges in the meta-analysis. First, different studies may adopt different gene-level tests, so the results are not compatible. Second, gene-level tests require multivariate statistics (i.e., components of the test statistic and their covariance matrix), which are difficult to obtain. To overcome these challenges, we propose to perform gene-level tests for rare variants by combining the results of single-variant analysis (i.e., p-values of association tests and effect estimates) from participating studies. This simple strategy is possible because of an insight that multivariate statistics can be recovered from single-variant statistics, together with the correlation matrix of the single-variant test statistics, which can be estimated from one of the participating studies or from a publicly available database. We show both theoretically and numerically that the proposed meta-analysis approach provides accurate control of the type I error and is as powerful as joint analysis of individual participant data. This approach accommodates any disease phenotype and any study design and produces all commonly used gene-level tests. An application to the GWAS summary results of the Genetic Investigation of ANthropometric Traits (GIANT) consortium reveals rare and low-frequency variants associated with human height. The relevant software is freely available.

1751W

Recursive organizer (ROR): an analytic framework for sequence-based association analysis. *X. Huang[1], L.P. Zhao[2].* 1) AbbVie Inc., North Chicago, IL. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA School of Public Health, University of Washington, Seattle, WA.

The advent of next-generation sequencing technologies affords the ability to sequence thousands of subjects cost-effectively, and is revolutionizing the landscape of genetic research. With the evolving genotyping/sequencing technologies, it is not unrealistic to expect that we will soon obtain a pair of diploidic fully-phased genome sequences from each subject in the near future. Here, in light of this potential, we propose an analytic framework called, recursive organizer (ROR), which recursively groups sequence variants based upon sequence similarities and their empirical disease associations, into fewer and potentially more interpretable super sequence variants (SSV). As an illustration, we applied ROR to assess an association between HLA-DRB1 and type 1 diabetes (T1D), discovering SSVs of HLA-DRB1 with sequence data from the Wellcome Trust Case Control Consortium (WTCCC). Specifically, ROR reduces 36 observed unique HLA-DRB1 sequences into 8 SSVs that empirically associate with T1D, a four-fold reduction of sequence complexity. Using HLA-DRB1 data from Type 1 Diabetes Genetics Consortium (T1DGC) as cases and data from Fred Hutchinson Cancer Research Center as controls, we are able to validate associations of these SSVs with T1D. Further, SSVs consist of nine nucleotides, and each associates with its corresponding amino acids. Detailed examination of these selected amino acids reveals their potential functional roles in protein structures and possible implication to the mechanism of T1D.

1752T

Valid permutation tests for genetic case-control studies with missing genotypes. *D.D. Kinnamon, E.R. Martin.* John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Permutation tests are indispensable tools in the analysis of genetic case-control studies. However, many permutation tests proposed in the genetics literature are justified only by heuristic arguments under the assumption of complete genotype data. We therefore developed a rigorous theoretical framework for constructing valid permutation tests for genetic case-control studies with unrelated subjects and missing genotypes arising from a variety of missing data processes. We began by specifying a nonparametric probability model for the observed data in such a study. Using group-theoretic arguments, we then established two conditions that together guarantee an exact level-α Monte Carlo permutation test for data generated under this nonparametric probability model. We showed that one of these conditions is not satisfied for the most frequently used Monte Carlo permutation test and that this test is guaranteed to be level α only for missing data processes with certain characteristics. We therefore proposed an alternative Monte Carlo permutation test that we showed is exact level α as long as all covariates influencing the missing data process are identified and recorded. We supplemented our theoretical development with Monte Carlo simulations for a variety of test statistics and missing data processes. We concluded that, while permutation tests can be extremely useful in genetic case-control studies, they must be constructed with careful consideration of the process generating missing genotypes to avoid inferential errors.

1753F

Heritability of Gene Expression Variation. *Z. Liu[1], J. Huang[1], W. Cookson[2], M. Moffatt[2], L. Liang[1].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) National Heart and Lung Institute, Imperial College, London SW3 6LY, UK.

We proposed a statistical approach to estimate the heritability of global gene expression variability, named variance heritability (abbreviated in this paper as vH2), in Epstein-Barr virus-transformed lymphoblastoid cell lines in two large datasets from nuclear families of British descent recruited through a proband with childhood eczema (MRCE, n=550, based on Illumina platform) and asthma (MRCA, n=405, based on Affymetrix platform). This vH2 estimates additional genetic effects on expression to those captured by narrow sense (additive) heritability, including dominant effect, GxG and GxE interactions. The median heritability of global gene expression variability across all probes in MRCE and MRCA were 0.04 and 0.04, the third quartile (Q3) were 0.12 and 0.14, respectively. Thirty percent of probes in MRCE and 34% of probes in MRCA have estimated vH2 larger than 0.1. Compared with estimates from randomly permuted data (median=0.004, Q3=0.08, 20% probes with H2>0.1 in MRCE; median=0.002, Q3=0.09, 22% probes with H2>0.1 in MRCA), our estimated results implied the existence of genetic variants regulating the variance of gene expression. For probes targeting the same transcript, the heritability estimates were consistent across the two independent datasets (MRCE and MRCA) and were robust to expression array platforms. We further mapped the heritability of gene expression variance to individual genetic loci using Illumina 300K SNPs and 39 millions imputed 1000Genomes SNPs and INDELs. Our results suggest new genetic regulatory variants that might be missed by standard eQTL mapping studies that focused on the mean level of gene expression rather than its variability.

1754W

On the simultaneous testing for large genomic regions: A clustering approach for rare variants. *s. Lutz[1], C. Lange[2].* 1) Biostatistics, University of Colorado, Aurora, CO; 2) Biostatistics, Harvard School of Public Health, Boston, MA.

In case-control studies, we propose a general analysis frame-work in which hundreds of genetic loci with allele frequency less than 1% can be tested simultaneously for association with case-control status. The approach is built on spatial- clustering methodology, assuming that rare variants that are associated with the target phenotype cluster in certain genomic regions. In contrast to standard methodology for rare-variant analysis, which has focused on collapsing methods, the proposed approach does not depend on each rare variant being associated with the case-control status in the same direction and with similar effect sizes. Using simulation studies, the properties of the approach are evaluated. In an application to a genome-wide association study (GWAS) for chronic obstructive pulmonary disease (COPD) in the COPDGene study, we illustrate the practical relevance of the proposed method.

1755T

A kernel based multilocus genetic association test for longitudinal quantitative phenotype data. *I. Mukhopadhyay, P.K. Mandal.* Human Gen Unit, Indian Statistical Inst, Kolkata, India.

Technological advances permit us to collect data on thousands of single nucleotide polymorphisms (SNPs) for each individual in a genetic association study. However, current paradigm of analyzing such dataset focuses on looking at association with the phenotype only at a single SNP level. Due to limited success of genome wide association study using this method, some multilocus association methods have been proposed in recent literature. Even then an important issue is often ignored while collecting such data. The phenotype data, whether qualitative or quantitative, is usually collected only once. However, due to medication and other factors like age, food habits etc the quantitative phenotype value may vary over time. Thus a longitudinal phenotype data i.e. repeated measurements on the quantitative phenotype of an individual at different time points might be more informative to get a more vivid picture of the genetic architecture. In this work we have developed a multilocus genetic association testing method using longitudinal quantitative phenotype data to see any genetic association between any genomic region of interest or genes (rather than a single SNP) with the phenotype. Our method explores the association between kernel based genotype similarity and phenotype similarity over different time points and combine information of such association over multiple markers. This method is flexible enough to adjust the effect of other covariates, if any and also can takes care of the missing value for the covariates at some time points. We have also derived the asymptotic distribution of the test statistic that will help to calculate the p-value faster than a permutation technique.

1756F

Adaptive approaches for combining multiple rare variant association tests provide improved power across a wider range of genetic architecture. *M. O'Connell[1], A. Benitez[2], J. Arbet[3], K. Grinde[4], C. Fu[5], B. Greco[8], A. Hainline[6], N. Tintle[7].* 1) Miami University of Ohio, Oxford, OH; 2) Brown University, Providence, RI; 3) Winona State University, Winona, MN; 4) St. Olaf College, Northfield, MN; 5) Massachusetts Institute of Technology, Cambridge, MA; 6) Dordt College, Sioux Center, IA; 7) Vanderbilt University, Nashville, TN; 8) University of Michigan, Ann Arbor, MI.

Over the past five years, numerous gene-based rare variant tests of association have been proposed, each of which attempt to combine variants within a gene or region of interest into a single association statistic, with a goal of providing more power than a strategy which analyzes each variant separately. Simulation results have shown that many of these individual tests provide good power for particular genetic architectures, but not others. We have developed a general strategy for combining any two or more gene-based rare variant tests using an adaptive approach, which yields a single p-value representing the cumulative evidence for association across the set of gene-based tests. For example this strategy can take any threshold based test and turn it into a variable-threshold test, combine similar tests (similar statistic with alternative weighting strategies), or combine substantially different tests (e.g., burden tests and variance components tests). Using simulation we provide guidance on the tradeoff between power gains and test robustness versus the number of tests being combined, a result which is based on the correlation structure of the tests are under the null hypothesis of no association. Finally, we demonstrate how recent results from our group which suggested a substantially different gene-based test which is robust to high proportions of non-causal variants, combined with other popular tests (burden and variance component tests), can provide improved power across a wider range of genetic architecture.

1757W

Genome-wide association study of allergic rhinitis. *C. Schaefer[3], J. Liu[1,2,3], L. Shen[3], T. Hoffmann[1,2], M. Kvale[2], Y. Banda[1,2], D. Ranatunga[3], N. Risch[1,2], J. Witte[1,2], E. Jorgenson[3].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Kaiser Permanente Northern California Division of Research, Oakland, CA.

Allergic diseases are complex and the risk of developing allergies is thought to have a strong genetic basis. To examine the role of genetic factors on allergic rhinitis, we conducted a genome-wide association study in 61,957 non-Hispanic White subjects from the Kaiser Permanente Genetic Epidemiology Research on Adult Heath and Aging (GERA) cohort, which includes genotypes on over 675,000 SNPs. We identified several variants that were associated with allergic rhinitis at a genome-wide significant level. The most significant SNP associations identified are located on chromosome 2 within the gene IL1RL1 (rs2160203, p=3.8*10^-10). IL1RL1 encodes a member of the interleukin 1 receptor family and has previously been reported to be associated with asthma. In addition, three SNPs (rs3806933, rs34624588, and rs1438673) that lay within a region of chromosome 5 containing the genes TSLP and WDR36 reached genome-wide significance (5*10^-8). TSLP promotes T helper type 2 cell responses and is associated with immune response in various inflammatory diseases, including asthma and allergic inflammation. WDR36 is involved in T-cell activation and has been shown to be associated with bronchial asthma and eosinophilic esophagitis. In summary, we detected genome-wide significant SNP associations for allergic rhinitis, in loci that have been previously reported to be associated with asthma and allergic disease.

1758T

Disentangling Pooled Triad Genotypes for Association Studies. *M. Shi, D.M. Umbach, C.R. Weinberg.* Biostatistics Br, NIEHS, Res Triangle Park, NC.

Association studies that genotype affected offspring and their parents (triads) offer robustness to genetic population structure while enabling analyses to probe maternal effects, parent-of-origin effects, fetal-maternal interaction, and gene-by-environment interaction. We study case-parents designs that use pooled DNA specimens to both reduce genotyping costs and make good use of limited available specimens. We assume that the assay for each diallelic autosomal locus counts the number of variant alleles among the 2h copies present in a pool of h individuals. We reduce the genotypes required per pool from 3h to three by randomly partitioning the available triads into pooling sets of h triads each and then creating three pools from every pooling set, one pool each for mothers, fathers, and offspring. Data analysis proceeds via log-linear modeling using the expectation maximization algorithm, where the pseudo-complete data are disaggregated allele counts. The analysis provides relative risk estimates and assessment of maternal and parent-of-origin effects. It also accommodates genotyping errors and missing genotypes. We compare the power of our proposed analysis for testing offspring and maternal genetic effects to that based on a difference approach considered by Lee and that of the gold-standard based on individual genotypes, under a range of allele frequencies, missing father proportions, genotyping error rates and modes of inheritance. Power calculations show that the pooling strategies incur only modest reductions in power if genotyping errors are low, while reducing genotyping costs and conserving limited specimens. We illustrate our procedure with data from a study of oral clefts.

1759F

Meta-Analysis of Genome-Wide Association Studies in Myopia in Nine Populations. *C. Simpson[1], R. Wojciechowski[2], V. Verhoeven[3,4], P. Hysi[5], M. Schache[6], X. Li[7], M. Hosseini[8,9], L. Portas[10], F. Murgia[10], K. Oexle[11,12], A. Paterson[8,9], V. Vitart[13], C. Hammond[5], P.N. Baird[6], M. Pirastu[10], J. Rotter[7], C.C.W. Klaver[3,4], T. Meitinger[11,12], D. Stambolian[14], J.E. Bailey-Wilson[1].* 1) Inherited Disease Research Branch, NHGRI, NIH, Baltimore, MD; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA; 3) Department of Ophthalmology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands; 4) Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands; 5) Department of Twin Research and Genetic Epidemiology, King's College London, St. Thomas' Hospital, London, UK; 6) Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, Australia; 7) Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 8) Institute of Medical Science, University of Toronto, Toronto, Canada; 9) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Canada; 10) Institute of Population Genetics, National Research Council, Sassari, Italy; 11) Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology I, Neuherberg, Germany; 12) Institute of Human Genetics, Technical University Munich, Munich, Germany; 13) Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK; 14) Department of Ophthalmology, University of Pennsylvania, Philadelphia, USA.

Myopia is a common refractive error which affects at least a third of most populations. Both genetic and environmental factors influence myopic development. It has a significant impact on the lives of affected individuals and carries high economic costs associated with treatment, loss of productivity and co-morbidity from vision impairment. Recent genome-wide association studies (GWAS) have identified a number of loci associated with myopia and refractive error. Here we report results of a large meta-analysis of myopia in nine cohorts, for a total of 17,787 individuals of European ancestry and replication in a further 8 cohorts for a total of 7953 individuals. Genotypes in each population were imputed to HapMap2 and analyzed separately by each group. Cases were defined as a spherical equivalent of -1 diopters (D) or worse and controls were defined as > 0D. Individuals between 0 and -1D were coded as unknown. Analyses were performed including age, sex and years of education, plus the first 3 principal components to adjust for population structure. Meta-analysis was performed in METAL using the sample size schema. Due to large differences in numbers of cases and controls for some studies, effective sample sizes were calculated using the formula recommended by the authors of METAL. Genomic control was used to adjust for any residual structural differences between populations. 3 SNPS were identified as genome-wide significant with $P < 5e-8$, rs10113215 on chromosome 8q12.1 and rs1370156 and rs2028099 both in a previously reported locus on chromosome 15q14. For replication, SNPs with $P < 1e-5$ were identified and all SNPs within 500kb each side of that SNP selected, for a total of 20,431 SNPs. The replication threshold was set by calculating the effective degrees of freedom using the Ramos method. SNPs will be considered to replicate where the p value < 0.0026. The replication analyses are ongoing and will be presented.

1760W

Identifying disease susceptibility variants using pleiotropy and independent genome-wide association studies. *A. Skol.* Genetic Medicine, Dept of Medicine, Univ Chicago, Chicago, IL.

Pleiotropy is the phenomenon whereby a single gene influences multiple traits. Given that observing more than one disease cosegregating within a family is not uncommon, and that several studies have mapped the same gene to multiple traits, it would appear that pleiotropy is not an unusual phenomenon in disease genetics. Here, I explore the ability to use pleiotropy to identify variants or genes that are common between diseases using independent GWAS data, meaning that the traits of interest are measured in independent groups of individuals. The rational for looking for pleiotropic loci is two-fold: first, most studies will be insufficiently powered to identify most of the trait or disease influencing variation, however if two traits share a genetic underpinning then we should find contributing variants in the intersection of the studies' marginally significant results, and second, pleiotropic loci will provide novel insights into the common genetic etiology of seemingly disparate disease. I will compare three methods for detection: meta-analysis, intersection among the tails of the GWAS test statistics, and false discovery rate of p-values from one GWAS within the tail of the other. I will present results demonstrating the power of each of these approaches, how to determine significance thresholds that maximize power, what effect sizes are detectable, and how to correctly accommodate linkage disequilibrium.

1761T

An extrapolation approach for estimating genome-wide significance for whole genome sequencing studies and region-based tests. *C. Xu*[1,2] *on behalf of the UK10K Statistics group.* 1) Lady Davis Institute, JGH, Montreal, Canada; 2) Dept. of Biostatistics, McGill University.

Rare genetic variation derived from whole genome sequencing studies is often analyzed by defining a series of windows or genetic regions, and then evaluating the association between phenotype and all the genetic variability in each region. However, the choice of region boundaries, test statistics, and model parameters for such tests may be somewhat arbitrary, and repeated analyses may be performed making different assumptions. Therefore, establishing an appropriate correction for multiple testing is a challenging problem. In recent work, we have used pilot data from the UK10K consortium whole genome sequencing study to propose an empirical approach to establishing significance thresholds. This is combined with extrapolation from small genomic regions to the whole genome to make the empirical evaluations computationally feasible. Here we evaluate the dependence of the resulting estimated significance thresholds on sample size and minor allele frequency. In our investigations, we show that significance thresholds, for a combined analysis including univariate tests for common SNPs and region-based tests for rare genetic variation, are driven by the common SNP variation, and that we saw little dependence of the results on sample size. We recommend a genome-wide significance threshold of 1e-08 for such a combined analytic strategy.

1762F

Methods for Association Analysis and Meta-Analysis of Rare Variants in Families. *S. Feng, D. Liu, M. Zawistowski, X. Zhan, G.R. Abecasis.* Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Recent advances in exome sequencing and the development of exome genotyping arrays are enabling explorations of the contribution of rare variants of clear functional consequence to complex traits. Decreasing cost of sequencing and the low cost of exome chip genotyping allow increasingly large cohorts to be studied. For these rare variants, single variant association tests can lack power and a variety of association tests that group rare variants by gene or functional unit have been proposed as alternatives. Here, we describe family-based association tests for rare variants that allow analysis of a variety of quantitative traits, with or without covariates, and show how these tests can be applied in meta-analysis settings. The key idea beyond our method is that various gene-level test statistics can be reconstructed from single variant score statistics and that, when the linkage disequilibrium relationships between variants are known, the distribution of these gene-level statistics can be derived and used to evaluate significance. This idea can be used to derive a variety of gene-level association tests in families and also to enable meta-analysis. We have implemented family-based burden test, SKAT and variable frequency threshold tests. The methods have been implemented in freely available C++ code. To investigate power, we simulated families with various structures and sizes and genes comprising between 1kb and 25kb of sequence, each explaining 1-5% trait variance. In each gene, 10-30% rare variants were chosen to be causal. Our simulations show that, using alpha=2.5×10-6, when all variants influence the trait in the same direction, the variable threshold method has the largest power. For a sample of three-generation-family of size 10 with 1000 founders or 2500 individuals, VT is predicted to obtain ~80% power when 80% variants are causal for a gene of 1kb explaining 2% of trait variance. When variants influencing traits in both directions co-exist in the same gene, when half of them are trait-decreasing, our implementation of the SKAT test performs the best. For a sample of three-generation-family of size 16 with 2000 founders or 8000 individuals, SKAT is predicted to obtain ~85% power when 80% variants are causal and 50% of them are trait decreasing for a gene of 1kb explaining 4% of trait variance. Type-I error was examined using simulations under the null hypothesis and they were well controlled.

1763W

Testing for differences between multiple groups in high-throughput sequencing data using Bayesian multi-scale Poisson models. *H. Shim*[1], *E. Pantaleo*[1], *M. Stephens*[1,2]. 1) Dept Human Genetics, Univ Chicago, Chicago, IL; 2) Dept Statistics, Univ Chicago, Chicago, IL.

High-throughput sequencing technologies are now routinely applied at a genome-wide scale to collect a variety of phenotypic data. Testing for differences in these data between multiple groups is frequently encountered in genomics applications (e.g., eQTL mapping using RNA-seq and detecting differences in transcription factor binding across tissues using DNase-seq or ChIP-seq) Most approaches perform a test for differences using summary statistics such as total number of reads mapped to a gene or window although the original data consists of the counts of reads mapped to each base along the genome. These typical approaches have limited power to identify diverse patterns of signals and their results are often sensitive to window size. For example, these approaches could miss associations when different bases within a window show effects in opposite directions and partially cancel each other out, leading to a small overall effect on total number of reads over the window. Here we present statistical methods for testing for differences in the original data, which leads to increased power by using full information from the data. Specifically, our methods consider the data as an inhomogeneous Poisson process and test for differences in underlying intensities using Bayesian multi-scale Poisson models. In addition to testing for differences, our approach aims to provide a better interpretation of the analysis such as which parts and features of the data are driving the observed signals. We illustrate the proposed methods on DNase-seq data from 70 HapMap Yoruba LCLs. Moreover, we modify the proposed method to detect differences in gene expression between multiple tissues measured by paired-end RNA-seq data. Our results demonstrate that the proposed approaches can considerably increase power to detect associations compared with conventional window-based approaches.

1764T

Detecting association of rare variants by testing an optimally weighted combination of variants for censored survival outcomes. *X. Wang, X. Zhao.* Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI. 1240 N 10th Street Milwaukee, WI 53205.

There is increasing evidence showing that rare genetic variants are important for identifying patients with aggressive diseases who need prioritized treatments. Identifying these patients is critical in improving survival outcomes. In addition, knowledge of the rare genetic variants can help us gain further understanding the biological processes underling aggressive diseases progression. Next-generation sequencing technology allows sequencing the whole genome of large groups of individuals with survival outcomes, and thus it makes directly testing prognostic rare variants possible. However, recently developed statistical methods for detecting association of rare variants are not applicable to survival data. Survival data are often subject to censoring and use the Cox proportional hazards model. The score test in a Cox proportional hazards model is seriously anticonservative for rare variants. In this paper, we use the signed square root of the log partial likelihood ratio test statistic to replace the score test and propose a Variable Wight Test for testing the effect of an optimally weighted combination of variants on a survival outcome. Our new test is based on optimal weights which are analytically derived under a certain criterion. Extensive simulation studies show that the type I error rates of the new test are under control. In order to evaluate the performance of the new test, we are conducting extensive simulation studies for power comparisons. Furthermore, we will illustrate the proposed test with a real data application.

1765F

Detecting association for low-frequency variants by the standardized linkage disequilibrium in case-control genome-wide screens. *C. Xing, C-Y. Lin, H-C. Ku.* McDermott Ctr, Univ Texas SW Med Ctr, Dallas, TX.

In genetic association studies a conventional test statistic is proportional to the correlation coefficient between the trait and the variant, with the result that it lacks power to detect association for low-frequency variants. Considering the link between the conventional association test statistics and the linkage disequilibrium measurement r-squared , we propose a test statistic analogous to the standardized linkage disequilibrium D-primer to enhance the power of detecting association for low-frequency variants with moderate to large effect sizes. We examined its validity and showed it is more powerful than the conventional methods to detect association for low-frequency variants in a genome-wide screen setting by both simulation and real data analysis.

1766W

An optimal quasi-likelihood-based burden test for rare-variant association. *X. Wu[1], H. Zhu[1], D. Liu[2].* 1) Department of Statistics, Virginia Tech, Blacksburg, VA; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

With the advent of next-generation sequencing (NGS) technologies, rare-variant association testing is receiving increasing amount of attention in genome-wide association studies. As a single rare variant shows little variation due to low allele frequency, standard testing methods for single-variant suffer insufficient power. A general solution is burden test, which assesses the genetic effect of a collapsed set of variants. However, currently available burden tests are usually proposed for population-based genetic studies thus may not be appropriate in the presence of related individuals. Furthermore, the weighting scheme in those burden tests is often empirical and lacks theoretical justification. In this paper, we propose an optimal MQLS burden test (OMBT), a flexible and computationally efficient method for testing multiple-variant (rare or common) association with disease-related traits (quantitative or dichotomous). OMBT is a generalization of the traditional MQLS/MASTOR test in order to accommodate rare variants under linkage disequilibrium (LD). Compared with other burden tests, OMBT has several advantages: (1) it takes into account dependence among individuals and among variants, (2) it can incorporate individuals with missing data, and (3) it provides optimal weights to guarantee maximized power for association testing. Simulation studies demonstrate that OMBT achieves correct type-I error and improved power over other competing methods. We apply the proposed method to a sequencing-based genetic study using data from the SardiNIA Medical Sequencing Discovery Project.

1767T

An empirical validation of random effects and Bayesian meta-analysis models. *R. Ahn, C. Garner.* Department of Epidemiology, Sprague Hall, Room 318, University of California, Irvine, Irvine, CA 92697-3905.

Genome-wide association study (GWAS) meta-analysis is a statistical method that is now routinely used to combine either individual-level data or summary statistics from multiple studies to increase the power to detect the small genetic effect sizes of common alleles and to decrease the number of false-positive associations. As new samples do not have to be genotyped, the cost to perform a GWAS meta-analysis with sample sizes in the tens of thousands or even in the hundreds of thousands is substantially lowered. As a result, hundreds of GWAS meta-analyses have been performed over the last few years and have substantially increased the number of risk loci discovered and replicated for several different phenotypes. Amongst the many approaches to GWAS meta-analysis, the most popular has been fixed effects meta-analysis because it is the most powerful approach. However, fixed effects meta-analysis ignores the potential heterogeneity that may exist between studies by assuming that the effect of a risk allele is the same across all studies. We have reanalyzed a two-stage GWAS meta-analysis using 292K single nucleotide polymorphisms (SNPs) directly genotyped in 4,533 celiac disease cases and 10,750 controls from 4 different countries in the discovery stage. In the replication stage, 131 SNPs were directly genotyped in 9,451 celiac disease cases and 16,434 controls from 11 countries. In the original study, a fixed effects meta-analysis approach was used. The purpose of our analysis was to determine if a novel random effects or Bayesian meta-analysis that accounts for between-study heterogeneity will yield results that differ from the results that were originally published. Here we present evidence that a SNP at a loci in chromosome 1 (RUNX3) that was previously reported to show genome-wide significance ($P < 5 \times 10\text{-}8$) was not genome-wide significant in either the discovery stage or the replication when the between-study heterogeneity was accounted for by the random effects or Bayesian meta-analysis approaches. This study highlights the need to carefully investigate between-study heterogeneity and the implementation of GWAS meta-analysis models that control for heterogeneity.

1768F

Methods and tools for fast efficient mixed-models based whole-genome association analysis for large cohorts and multiple phenotypes. *Y. Aulchenko.* Lab of recombination and segregation analysis, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russian Federation.

Whole-genome association (WGA) analysis is a tool of choice for identification of loci underlying complex human traits. In WGA scans, association of millions of genetic polymorphisms measured with SNP arrays and/or genome re-sequencing is analyzed in relation to phenotypes in thousands of individuals. The mixed models methodology allows powerful WGA analysis in case when the trait under analysis is highly polygenic and/or when genetic (sub)structure is present in the sample. However, the analysis under mixed models is computationally challenging, especially for large cohorts and/or multi-trait "omics" data sets. Here, I will review recent advances in mixed-models based WGA analyses, and describe a number of methods, algorithms, and software tools we have developed to facilitate the analysis of cohorts including tens of thousands of participants and analysis of "omics" data sets potentially including hundreds of thousands of phenotypes.

1769W

PODKAT: a non-burden test for associating complex traits with rare and private variants. *U. Bodenhofer, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University, Linz, Austria.

High-throughput sequencing technologies have facilitated the identification of large numbers of single-nucleotide variations (SNVs), many of which have already been proven to be associated with diseases or other complex traits. Since association tests considering individual SNVs independently are known to be underpowered, different collapsing strategies have been proposed to consider multiple SNVs occurring in a region simultaneously. Such strategies can be classified into burden tests and non-burden tests, an important representative of which is the acclaimed Sequence Kernel Association Test (SKAT) by Wu et al. Several large sequencing studies, such as, the 1000 Genomes Project, the UK10K project, or the NHLBI-Exome Sequencing Project, have consistently reported a large proportion of private SNVs, that is, variants that are unique to a family or even a single individual. Non-burden tests like SKAT are typically utilizing correlations between SNVs to increase statistical power - a strategy that is not applicable to private SNVs, since singular events are generally uncorrelated. Burden tests are potentially able to deal with private SNVs, but only if the number of private SNVs occurring in a region is correlated with the trait under consideration. Moreover, burden tests have a disadvantage if deleterious and protective SNVs occur together in the same region. We propose the Position-Dependent Kernel Association Test (PODKAT). By means of a position-dependent kernel approach, PODKAT can potentially detect associations of rare and private SNVs with the trait under consideration even if the burden scores are not correlated with the trait. PODKAT assumes that, the closer two SNVs are on the genome, the more likely they have similar effects on the trait under consideration. This assumption is fulfilled as long as deleterious, neutral, and protective variants are grouped sufficiently well along the genome. PODKAT can be applied to testing focused regions as well as to whole-exome and whole-genome association testing. We evaluated PODKAT on simulated genomic data with simulated traits (both quantitative and dichotomous) and real sequencing data with simulated and real traits to illustrate its potential for association testing involving rare and private variants.

1770T

Model-based pathway enrichment analysis and prioritization of genetic variants in enriched pathways yields novel putative susceptibility loci for rheumatoid arthritis and type 1 diabetes. *P. Carbonetto[1], M. Stephens[2].* 1) Dept. of Human Genetics University of Chicago Chicago, IL; 2) Depts. of Statistics and Human Genetics University of Chicago Chicago, IL.

Many common diseases appear to be highly polygenic, modulated by a large number of genetic factors each with a small effect on disease risk. As a result, standard single-marker analyses of genome-wide association studies are unable to identify most variants conferring risk to disease; many individual SNP contributions to disease risk are so subtle that they cannot be distinguished from correlations that occur by chance. To address this problem, researchers have developed statistical methods that aggregate information over groups of related genes, such as genes in common pathways, to identify gene sets that are enriched for variants associated with disease. However, these methods often fail to answer a critical question: which genes and variants in the enriched pathways are associated with disease risk? We have developed a model-based approach that not only systematically interrogates pathways for enrichment of disease associations, but also uses this information to prioritize variants assigned to these pathways, in an attempt to promote these variants above background noise. A key feature of this approach, which distinguishes it from other enrichment methods, is that it is able to uncover additional loci contributing to disease risk. We describe our results from applying this approach to genome-wide studies of Crohn's disease (CD), rheumatoid arthritis (RA) and type 1 diabetes (T1D). These analyses yield many additional putative associations compared to standard single-marker analyses. Because of the predominant role of the MHC in RA and T1D, we developed methods to assess enriched pathways beyond enrichment of disease associations in the MHC, and found that this was important for reliably identifying pathways, and non-HLA associations. For CD and RA, most of the additional disease associations (7 of 8 non-HLA associations) are corroborated by other studies and large-scale meta-analyses, validating the usefulness of the methodology. The one novel disease association with RA is a region near TP73, and it appears to be promising candidate. For T1D, the results very strongly support the connection between IL-2 signaling and development of T1D, and prioritization of IL-2 signaling genes yields strong evidence for 7 additional non-HLA candidate disease regions. Of these, 4 have been validated by other genome-wide studies of T1D (IL2, IL2RA, CLEC16A and C1QTNF6), and 3 constitute novel T1D loci (regions containing RAF1, MAPK14 and FYN).

1771F

Rapid linear mixed model methods for large-scale genome-wide association studies. *W-M. Chen[1], A. Manichaikul[1], S.S. Rich[1], M. Cushman[2], M.M. Sale[1].* 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Department of Medicine, University of Vermont, Colchester, VT.

Population stratification is known to confound association results in genome-wide association studies (GWASs). Recently, Linear Mixed Models (LMMs) have received attention as a flexible way to adjust for population structure using variance components. Multiple LMM algorithms were recently published that included exact maximum likelihood methods with computational complexity quadratic to sample size and fast approximations with computational complexity linear to sample size. Although existing LMM methods have proven to be computationally feasible for GWAS scans, several limitations exist. First, current methods have inflated type I errors in the presence of missing genotypes and misspecification of polygenic estimators; second, current efficient algorithms for LMM require analyzing one trait, limiting the number of traits that can be analyzed in a reasonable amount of time. Here, we present two novel fast LMM methods with computational complexity linear to sample size: (1) an approximation fast LMM method, and (2) an exact fast LMM method. As score tests, the proposed LMM test statistics are developed to be robust to missing genotypes, missing phenotypes, phenotypic distributional assumptions, and misspecification of polygenic parameters. The computational improvement of our implementations over existing fast LMM methods includes efficient computation of the relatedness matrix using the KING algorithm, rapid regression-based estimation for polygenic analysis in the approximation LMM method, and rapid genome scans in a matter of seconds for both fast LMM methods. In a test dataset consisting of 2,400 samples each typed at ~800,000 SNPs and 1,000 quantitative traits, it takes 40 minutes to perform 1,000 GWAS scans (*i.e.*, 2 seconds per GWAS scan) using the approximation fast LMM method and 9 hours using the exact fast LMM method on a single CPU. The genome scan results using the proposed fast LMM methods are highly correlated (*e.g.*, with correlation 0.994) with the results using the exact maximum likelihood methods, with the same order of significance of association at the top GWAS hits. Efficient implementations in our software tool KING makes it feasible to conduct large-scale GWAS scans for 10,000s of traits with small amount of computing resources.

1772W

An effective association testing procedure incorporating admixture mapping information. *G. Gao, W. Chen.* Department of Biostatistics, Virginia Commonwealth University, Richmond, VA.

Admixed populations are formed by recent admixture of two or more ancestral populations. For instance, African Americans often have recent genetic ancestry from both West Africans and European Americans. As a result of admixture, the variation in genome-wide ancestry (a type of stratification) in admixed populations can be a confounding factor in association tests, causing false positive and false negative findings. This makes gene mapping more complicated in admixed populations than in non-mixed populations (such as Europeans). For gene mapping in admixed populations, admixture mapping tests use admixture linkage disequilibrium (LD) and can only identify a causal variant in a large chromosomal region (several Mbs). Admixture LD occurs during admixture over the past several hundred years when large chromosomal segments are inherited from a particular ancestral population. To identify a causal variant in a small region, association tests that correct for local ancestry have been developed. These tests can use the background LD (traditional LD) existing in the ancestral populations and therefore can often identify a causal variant in a region with less than a few hundred Kbs). However, these tests can have relatively low power. Recently, to acquire increased power, several joint association tests that combine information from admixture mapping tests and association tests that correct for local ancestry have been proposed. However, these joint tests are more appropriate to identify a causal variant in a large chromosomal region (several Mbs). Our simulation studies showed that these joint tests could not control family-wise error rates (FWERs) in genome wide association studies if the null hypothesis is to test if a SNP is in background LD with the causal variants. In this study, we propose an effective association testing procedure for analysis of genome wide SNP data that incorporates appropriate amount of information from admixture mapping into association tests that correct for ancestry. Our simulation studies indicates that the association testing procedure not only controls family-wise error rates, but also can have improved power compared to the association tests that correct for local ancestry. We applied the association testing procedure to association studies in a data set from the Multi-Ethnic Study of Atherosclerosis project.

1773T

Mixed model association statistic with correction for case-control ascertainment provides large increase in power. *T. Hayeck[1], N. Zaitlen[2], B. Vilhjalmsson[1], S. Pollack[1], J. Yang[3], G. Chen[3], M. Goddard[4], P. Visscher[3], N. Patterson[5], A. Price[1].* 1) Harvard School of Public Health, Boston, MA; 2) University of California, San Francisco; 3) University of Queensland, Brisbane, Australia; 4) University of Melbourne, Melbourne, Australia; 5) Broad Institute, Cambridge, MA.

We introduce a mixed model association method for ascertained case-control studies that increases test statistics at causal markers by up to 23% vs. existing mixed model methods, with well-controlled false-positive rate. It is widely known that appropriate modeling of case-control ascertainment can produce large increases in power for case-control studies with fixed-effect covariates (reviewed in Mefford & Witte 2012 PLoS Genet), but such increases in power have not yet been obtained for models that include random effects. Here, we improve upon existing mixed model methods (e.g. Kang et al. 2010 Nat Genet, Zhou & Stephens 2012 Nat Genet) using a score statistic computed from posterior mean liabilities (PML) under the liability threshold model. The PML of each individual is conditional not only on that individual's case-control status, but also on every other individual's case-control status and on the genetic relationship matrix obtained from the data. For example, disease cases with higher genetic relationships to other disease cases will be assigned a larger PML than disease cases with lower genetic relationships to other disease cases. The PML are estimated using a multivariate Gibbs sampler, with the liability-scale phenotypic covariance matrix based on the genetic relationship matrix and a heritability parameter estimated via Haseman-Elston regression on case-control phenotypes followed by transformation to liability scale (Lee et al. 2011 AJHG). The Gibbs sampler does not iterate over SNPs, and overall running time is comparable to existing mixed model methods. In simulations of unrelated individuals, our statistic outperformed existing mixed model methods in all scenarios tested, with the magnitude of the improvement depending on the severity of case-control ascertainment. For example, we observed 23%, 8% and 2% improvements in test statistics at causal markers for ascertained case-control studies simulated at thresholds of 3, 2 and 1 (corresponding to disease prevalence of 0.13%, 2.3% and 16%), with well-calibrated test statistics at null markers. In conclusion, a large increase in power over existing mixed model association methods is available for ascertained case-control studies of diseases with low prevalence.

1774F

Testing the spatial correlation of association signals from two genome scans. *J. Hecker[1], D. Propopenko[1], P. Costa[1], E. Silverman[2], M. Naylor[3], S. Weiss[2], C. Lange[1,2,3,4], H. Loehlein Fier[1].* 1) Institute of Genomic Mathematics, Bonn, Germany; 2) Channing Laboratory, Brigham and Women's Hospital, Boston, United States; 3) Department of Biostatistics, Harvard School of Public Health, Boston, United States; 4) German Center for Neurodegenerative Diseases(DZNE), Bonn, Germany.

It is widely recognized that complex diseases like i.e. Alzheimer's disease, Cardio-Vascular-Diseases, Diabetes, Asthma, psychiatric diseases, and many others are characterized by a interaction of genetic, environmental and individual factors (e.g. Hunter 2005, Manolio et al. 2006, Risch et al. 2009, Manolio et al. 2008, Goldstein 2009, Thomas 2010).

As a result usually multiple correlated traits exist that describe the syndromes of a complex disease. Pinpointing and quantifying genetic risk factors of a complex disease based on several correlated traits however implies a challenge, since different traits might be differently associated with the genetic profile of the analyzed subjects (e.g. Wang 2012, Yang and Wang 2012). In order to understand the genetic architecture of complex diseases it is therefore crucial to unravel the genetic dependencies between correlated disease traits (pleiotropic effects).

In this communication, we propose a method to test the spatial correlation of two GWAS-association signals. For this, we need to combine the information about the positions and the corresponding p-values. We utilize kernel density estimation based techniques to locate clusters of significant p-values. With this knowledge we can use similarity measures, i.e. the L2-distance, to test the positional overlap of the identified clusters. We conduct a simulation study to assess how well our method is able to detect spatial correlation of GWAS signals and apply our method to real data.

1775W

Variable selection for GWAS with linear mixed models yields improved power and control of type I error. *D. Heckerman[1], O. Weissbrod[2], N. Fusi[1], C. Kadie[3], R. Davidson[3], C. Lippert[1], J. Listgarten[1].* 1) Microsoft Research, Los Angeles, CA; 2) Computer Science Department, Technion, Haifa; 3) Microsoft Research, Redmond, WA.

Linear mixed models (LMMs) are now routinely used to correct for genetic structure in genome-wide association studies (GWAS). At their core, LMMs rely on the estimation of a genetic similarity matrix, which encodes the pairwise similarity between every two individuals in a cohort. These similarities are estimated from single nucleotide polymorphisms (SNPs) or other genetic variants. Traditionally, all available variants are used to estimate the matrix. Here, we provide both theoretical and empirical evidence that such use is non-optimal, and that the careful selection of variants can lead to improved results. Theoretically, a linear mixed model is equivalent to a form of linear regression, where the SNPs that determine the genetic similarity matrix in the LMM view are covariates in the linear-regression view. Taking the latter view strongly suggests that the inclusion of variants relevant to the phenotype and the exclusion of variants irrelevant to the phenotype should lead to improvements in analysis. Empirical results confirm this argument, using a simple variant selection algorithm that searches over various sets of SNPs to identify those that maximize cross-validated prediction accuracy. To keep the search practical, we order SNPs for each fold by their univariate linear-regression P values on the training data for that fold. We then use increasing numbers of SNPs by this ordering, measuring prediction accuracy on the out-of-sample test set. Next, we average the prediction accuracy over each fold. Finally, we identify the number of SNPs that optimized this average. Essential to this procedure is out-of-sample optimization of a key parameter of the LMM representing the ratio of the variance explained by noise to the variance explained by the genetic similarity matrix. In-sample optimization of this parameter leads to lower quality GWAS performance. We apply this variant selection algorithm to several phenotypes from the WTCCC1 and WTCCC2 data, demonstrating significant, simultaneous improvements in control of type I error and power.

1776T

Association Studies of Imputed Genotypes Using Expectation-Maximization Likelihood-Ratio Test. *K. Huang[1], Y. Li[1,2].* 1) Department of Biostatistics, The University of North Carolina, Chapel Hill, NC; 2) Department of Genetics, The University of North Carolina, Chapel Hill, NC.

Genotype imputation has become standard practice in modern genetic studies. As sequencing-based reference panel continues to grow, we have increasingly more well imputed markers but at the same time also more imputation markers with relatively low quality. Here, we propose new methods that attempt to more elegantly incorporate uncertainty when analyzing imputed genotypes. We consider two scenarios: 1) when posterior probabilities of genotypes are estimated or 2) when only imputed dosages are available. When posterior probabilities are estimated, we developed an expectation-maximization (EM) likelihood-ratio test (LRT) for association studies. When only dosages are observed, instead of modeling dosages directly, we first sample the probabilities of all three possible genotypes and then apply the EM-LRT test on the sampled probabilities for EM-LRT. Extensive simulations have shown that type I error rates of the EM-LRT tests under both scenarios are protected. Regarding power, EM-LRT-Prob offers enhanced statistical power across the whole spectrum of imputed quality and EM-LRT-Dose has similar power performance as EM-LRT-Prob and better than standard methods that model dosages directly, especially for markers with relatively low imputation quality ($Rsq \leq 0.3$). Application to real datasets will also be shown.

1777F

Bayesian sparse models of high-dimensional correlated traits in related and unrelated individuals. *V. Iotchkova[1, 2, 3], J. O'Connell[4], A. Dahl[4], J. Marchini[3, 4].* 1) The EMBL-European Bioinformatics Institute, Hinxton, United Kingdom; 2) The Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 3) Department of Statistics, Oxford University, Oxford, United Kingdom; 4) The Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom.

Genetic association studies of complex traits have yielded a wealth of biologic discoveries. However, up to now studies have been mostly carried out under simplistic assumptions, for instance analyzing one trait and one SNP at the time, thus failing to capture the underlying complexity of these datasets. In order to move beyond simple GWAS the joint analysis of complex, highly-dimensional datasets represents an important extension of phenotype-genotype associations with great potential. The move to high-dimensional phenotypes raises many new statistical problems. For example, how to model the large diversity of possible different types of phenotypes, accounting for sporadically missing phenotype data, modelling sparse signals and hidden factors, how to model relationships and networks between phenotypes and how to assign statistical significance. We have developed a Bayesian model designed to detect association when correlations exist between high-dimensional phenotypes and between individuals due to relatedness. We impose a sparsity assumption through the use of a 'spike and slab' prior on the genetic effect sizes that allows us to learn the subset of traits that are associated with each SNP or genetic region. In this way our method is able to carry out model inference (association testing) and model selection at the same time. To allow for related individuals and/or protect against population structure our method generalizes single-trait linear mixed model analysis to high-dimensional traits. We use variational bayesian methods to efficiently fit our model, which facilitates approximate fully Bayesian inference. A key property of our method is that it can handle missing phenotype data, which we impute as part of the model fitting process. To illustrate the power of our method we have applied it to a study of high-dimensional gylcomics phenotypes on 960 individuals from an isolated population where individuals show appreciable levels of relatedness. At the MGAT5 association on 2q21 (rs1257220), which was originally discovered in a much larger meta-analysis, our method clearly detects an association (log10 Bayes Factor = 5.6), whereas approaches based on combining single-trait analyses uncover little signal. On simulated datasets of high-dimensional traits our approach has significantly more power than simpler approaches when the true signal is sparse, can accurately impute missing phenotypes and control for population structure and relatedness.

1778W

Effect size estimation in the stage of planning a replication study. *C. Kuo[1], D. Zaykin[2].* 1) Department of Community Medicine & Health Care, University of Connecticut, Farmington, CT; 2) Biostatistics Branch, National Institute of Environmental Health Sciences, NIH.

Large-scale association and sequencing studies generate millions of results. The top hits usually sorted by P-value are of interest but the smallest P-values can result from true signals or simply occur by chance. A common strategy to separate true effects from null effects is to carry forward a number of top hits to a replication study. In the replication study, the number of top hits can be chosen to control family wise error rate or false discovery rate. The sample size that depends on the effect sizes such as odds ratios or non-centralities must be carefully chosen to achieve a specified power. Due to the winner's curse, the effect sizes that rank on the top tend to be inflated by naive estimates. Previous methods that estimate effect sizes accounting for the winner's curse mainly fall into 'conditional likelihood approaches' or 'Bayesian approaches'. It has been criticized that the conditional likelihood approaches don't use the population information provided by samples. The Bayesian approaches are immune to the winner's curse while they require specifying a prior, which if not correctly specified would bias the results. Motivated by population genetics theory, we assume that the effect sizes follow a Gamma with unknown shape and scale parameters. Following that, it is straightforward to incorporate the distribution in a conditional likelihood and also to derive Bayes estimates. The methods we propose outperform existing methods and are robust to the assumed effect size distribution. Our methods are general in that they are developed for P-value and can be easily adapted for different effect size measures.

1779T

Greater power for kernel-based tests using the likelihood ratio. *J. Listgarten, J. Xiang, C. Kadie, D. Heckerman, C. Lippert.* eScience, Microsoft Research, Los Angeles, CA.

Recently, tests for association between a phenotype and region- or gene-based sets of SNPs have attracted interest, in particular, for rare variants. One of the dominant approaches is to use a variance-component model, with a score test, such as implemented in SKAT. The score test in this setting can easily be computed in closed form, and has known asymptotic results, making computation of P values extremely efficient. In addition, because the score test does not require fitting of alternative model parameters, this test is much faster than a likelihood-ratio test (which does require such fitting). Finally, in light of the fact that asymptotic results for likelihood-ratio tests in this setting require unrealistic assumptions, one might conclude that the score test is the obvious method of choice. Surprisingly, however, we have found that there is good reason to use a likelihood ratio test over a score test in this setting. In particular, we have found that (1) the likelihood ratio test offers a substantial increase in power , and (2) that with just a handful of permutations, and a particular null distribution parametric form, that p-values can be computed reasonably efficiently (while controlling type I error). We systematically examined use of our approach on a number of real and simulated data sets spanning a wide range of important settings including case-control and continuous phenotypes, rare and common variants, and data confounded by relatedness and population structure. Overall, we found power to be greatly increased (about 50%) while still controlling type-1 error. To overcome the computational burden, we also describe a number of speedups for both tests that make genome-wide testing in the presence of population structure and relatedness feasible on data sets involving tens of thousands of samples, even for the likelihood-ratio test.

1780F

Bayesian mixed model association statistics in linear time. *P. Loh[1,2], N. Patterson[2], A.L. Price[1,2,3].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Department of Biostatistics, Harvard School of Public Health, Boston, MA.

Linear mixed models are a powerful statistical framework for identifying associated loci and avoiding confounding. Existing methods rely on the computation of a genetic relationship matrix, which requires time $O(MN^2)$ (where N = #samples and M = #SNPs) and implicitly assumes an infinitesimal genetic architecture in which all variants are associated. Here, we propose a very fast O(MN) coordinate descent algorithm that can achieve the same results as existing methods, but can also be generalized to non-infinitesimal genetic architectures to increase power. In each iteration of the algorithm, the estimated effect size of each SNP is updated with its posterior mean given the estimated components of variance, the prior distribution on SNP effects, and the residual phenotype conditional on current estimates of all other SNP effect sizes. We define a retrospective score statistic computed from the phenotypic residuals upon convergence, which typically occurs in fewer than 10 iterations. When an infinitesimal (normal) prior distribution of effect sizes is used, the posterior mean effect sizes are best linear unbiased predictions (BLUPs) and optimize a ridge regression problem. Our iterative method thus calculates BLUP coefficients and residuals by applying coordinate descent on the vector of SNP effect sizes. In this case our statistic is equivalent to the GRAMMAR-Gamma mixed model association statistic (Svishcheva et al. 2012 Nat Genet); we observed a correlation of 1.000 in simulations. When modeling non-infinitesimal genetic architectures, our method achieves improved power as a consequence of more accurate estimation of sparse SNP effects: for example, in simulations in which only 5% of independent markers are causal (with the number of samples roughly equal to 10 times the number of causal markers), our approach achieves a 10% increase in test statistics at causal markers as compared to existing methods.

1781W

Near equivalent calibration and power of joint and meta-analysis for association analysis of quantitative traits. *C. Ma, T. Blackwell, M. Boehnke, L.J. Scott, the GoT2D investigators.* Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

In genome-wide association studies of quantitative traits (QTs), investigators typically analyze common genetic variants using linear regression within each study, and combine association results across studies using fixed-effects meta-analysis. For common variants (minor allele frequency [MAF] ≥ 5%), linear regression is well-calibrated, and meta-analysis has near comparable power to joint analysis of the combined individual-level data (Lin and Zeng, 2010). In current sequencing and dense chip association studies, investigators wish to identify trait-associated low frequency variants (MAF < 5%), for which the calibration of linear regression analysis and the relative power of joint and meta-analysis has not been investigated. Here, we assess and compare the calibration and power of linear regression in joint and meta-analysis for QT association analysis across all allele frequencies, but in particular for low frequency variants, and examine the impact of sample size, additional covariates, and non-normally distributed QTs.

For common and low frequency variants (with minor allele count ≥ 5), simulation results show that: (1) linear regression-based joint analysis of normally-distributed QTs is well-calibrated; (2) sample-size weighted meta-analysis is well-calibrated and only slightly less powerful than joint analysis; and (3) inverse variance weighted meta-analysis is slightly anti-conservative. These results continue to hold when modeling the effects of additional covariates. For non-normally distributed QTs, joint and meta-analysis can become anti-conservative, but inverse-normal transformation of the QT remedies this problem while resulting in only modest loss of power. We are currently analyzing quantitative trait data from the GoT2D sequencing study to assess the generalizability these simulation-based findings to real data.

1782T

Estimating causal variant allele frequency, and thus efficacy of sequencing, at genetic loci identified by GWAS. *P.F. O'Reilly[1], C.J. Hoggart[2].* 1) Epidemiology & Biostatistics, Imperial College London, London, United Kingdom; 2) Genomics of Common Disease, Imperial College London, London United Kingdom.

Genome-wide association studies (GWAS) have provided a large number of genotype-phenotype associations, but these can only be assumed to highlight genetic regions harbouring causal variants rather than identifying the causal variants themselves. However, sequencing of such susceptibility loci makes identifying the causal variants feasible, which is a crucial step in translating the findings from GWAS into medical application. Most genetic regions identified by GWAS have not been interrogated for the causal variants via sequencing, largely due to the significant cost involved, and when they have, this has tended to be on an 'ad hoc' basis according to the resources of individual studies or the significance of association signals at certain loci. The availability of the 1000G reference panel for imputation means that future GWAS will capture > 95% of common variants, so that subsequent sequencing is only worthwhile if the causal variant(s) is rare and has the power to be detected in the sample. Here, we introduce a novel method that estimates the posterior probability of the allele frequency of putative causal variants at a genetic locus, which can be used to determine whether sequencing is likely to yield causal variants or whether it is more likely that they are already present among the available genotyped or 1000G imputed SNPs. Our method exploits the SNP GWAS association P-values, their minor allele frequency and Linkage Disequilibrium (LD), to calculate the likelihood that the data are consistent with a causal variant(s) of a given effect size and frequency. While the probability computed is dependent on several modelling assumptions, we show how the method can be used to prioritise the sequencing of genetic loci with known effects on human traits or diseases based on GWAS results.

1783F

Testing Association without Calling Genotypes Allows for Systematic Differences in Read Depth and Sequencing Error Rate between Cases and Controls. *G.A. Satten[1], H.R. Johnston[2], A.S. Allen[3,4], Y.J. Hu[2].* 1) Division of Reproductive Health, CDC, Atlanta, GA; 2) Department of Biostatistics and Bioinformatics, Emory University, Atlanta GA; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham NC; 4) Duke Clinical Research Institute, Durham NC.

Background: The quality of genotype calling for next-generation sequence data depends on read depth. Loci with high coverage can typically be called reliably, while those with low coverage may be difficult to call. In an association study, if case participants are sequenced to a greater depth than controls, the difference in genotype quality can introduce a systematic bias. This can easily occur when historical controls (e.g., data from The 1000 Genomes Project) are used as controls.

Methods: We propose directly comparing the proportion of calls for the minor allele between cases and controls, rather than comparing genotypes. We show how this proposal can be used to perform both single-marker test and gene-level test of rare variants (e.g., using modified versions of the burden test, variable threshold test or SKAT). We also show how this proposal can be extended to the situation where the per-call read error rate differs between cases and controls.

Results: Using simulated data and theoretical results, we demonstrate our proposals yield valid tests even in the presence of systematic differences in coverage rate between cases and controls, and show that in these situations, tests based on genotype have inflated size. We also show that power gains are possible using designs where we increase the number of controls while decreasing the read depth (while keeping total reads constant).

1784W

Meta-Analysis of Sequencing Studies Under Random-Effects Models. *Z. Tang, D. Lin.* Department of Biostatistics, University of North Carolina at Chapel Hill.

Meta-analysis of genome-wide association studies (GWAS) has led to the discoveries of common genetic variants for virtually every complex human disease. Recent advances in sequencing technologies have made it possible to extend association studies to rare variants. Because larger sample sizes are required to detect rare variants than common variants (with similar effect sizes), meta-analysis is as important to sequencing studies as to GWAS. Several research groups have recently developed meta-analysis methods for gene-level associations with rare variants under fixed-effect models. Those methods will lose power if the genetic effects are heterogeneous among participating studies. We propose novel multivariate random-effect models which allow the effects of multiple variants within a gene to vary among participating studies and develop the corresponding meta-analysis methods to perform gene-level association tests. Our methods take score statistics as input and thus can accommodate any study designs and any phenotypes. We produce the random-effect versions of all commonly used fixed-effect gene-level association tests, including burden tests, variable-threshold (VT) tests and sequence-kernel association tests (SKAT). We demonstrate through extensive simulation studies that the new methods are substantially more powerful than existing ones in the presence of moderate and high heterogeneity and maintain similar power when the heterogeneity is low. An application to the NHLBI Exome Sequencing Project data led to the discoveries of several genes associated with blood pressures that were previously undetected by fixed-effect methods. The relevant software is freely available.

1785T

A mixed model using both principal components and top markers corrects for population stratification and improves power. *G. Tucker, B. Berger.* Mathematics and CSAIL, MIT, Cambridge, MA.

In recent years, mixed models have been used extensively to calculate GWAS association statistics. Mixed models implicitly assume an infinitesimal genetic architecture, i.e., one in which all SNPs have a small nonzero effect; however, it is widely believed that disease phenotypes do not follow an infinitesimal model and that modeling a non-infinitesimal architecture would increase power. As a step in that direction, Lippert et al. (Sci Rep 2013) developed the state-of-the-art FaST-LMM Select method, which constructs the genetic relationship matrix (GRM) from a subset of the SNPs that are likely to be causal. FaST-LMM Select improves power over standard mixed models, especially when the phenotype is caused by few SNPs. However, limiting the GRM to a subset of the SNPs can result in insufficient correction for population stratification. As a solution to this problem, we propose a novel approach that preprocesses the data using principal components (PCs) and feeds the results to a program such as FaST-LMM Select, thereby leveraging the benefits of FaST-LMM Select, while properly correcting for population stratification.

Through simulation, we show that our framework corrects for population stratification without compromising power. When stratification exists, we observe that Fast-LMM Select inflates statistics on null SNPS (λ_{GC} = 1.32±0.04) whereas our method is properly calibrated (λ_{GC} = 1.00±0.01). Moreover, our method improves power over FaST-LMM Select (as measured by the mean ψ^2-statistic on causal SNPs; 11.4±0.1 versus 10.6±0.2), likely because the PCs reduce noise in selecting subsets of SNPs for the GRM. In addition, our method selects fewer SNPs to include in the GRM, yielding computational savings. In simulations without stratification, both methods perform nearly identically; thus our method retains the advantages of FaST-LMM Select over standard mixed models, such as improving power and correcting for confounding due to rare variants in spatially structured populations (Lippert et al., Sci Rep 2013; Listgarten et al., Nat Genet 2013).

1786F

Use of P-values to evaluate the probability of a genuine finding in a large-scale genetic association studies. *O. Vsevolozhskaya[1], CL. Kuo[2], D. Zaykin[2].* 1) Mathematical Sciences, Montana State University, Bozeman, MT; 2) National Institute of Environmental Health Sciences, National Institutes of Health, USA.

To claim the existence of an association in modern genome-wide association studies (GWAS), a nominal P-value has to exceed a stringent Bonferroni-adjusted significance level. Despite strictness of the correction, a significant P-value does not indicate high probability that the claimed association is genuine. A simple Bayesian solution -- the False Positive Report Probability (FPRP) -- was previously proposed to convert the observed P-value to the corresponding probability of no true association. Although the FPRP solution is highly popular, it does not reflect probability that a particular finding is false. Here, we offer a simple POFIG method -- a Probability that a Finding is Genuine. POFIG enables one effectively to convert a P-value to the probability that a particular association with the trait is genuine. The validity of POFIG is supported by the results of a simulation study and the potential utility of our approach is discussed with reference to future GWAS discoveries.

1787W

A correction strategy for imputation across genotyping arrays. *Y. Xie[1], J. Rice[1], L. Bierut[1], R. Culverhouse[2], N. Saccone[3], E. Johnson[4], D. Hancock[4], COGEND collaborators.* 1) Department of Psychiatry, Washington University in St. Louis, St. Louis, MO 63110; 2) Division of General Medical Sciences, Washington University in St. Louis, St. Louis, MO 63110; 3) Department of Genetics, Washington University in St. Louis, St. Louis, MO 63110; 4) Behavioral Health Epidemiology Program, RTI International, 3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709-12194.

Genotype imputation methods are widely used to extend the utility of genome-wide association study (GWAS) data. Genotype imputation is a powerful approach because it can potentially identify causal SNPs that are untyped in the study data and can provide higher resolution for associated regions. However, bias can exist if different genotyping arrays are used and are unbalanced for case versus control subjects. Two currently used strategies of imputation across genotyping arrays are: (1) imputation based on the union of genotyped SNPs (i.e., SNPs on one of the two arrays); (2) imputation based on the intersection of genotyped SNPs (i.e., SNPs available on both arrays). Both of these strategies have weaknesses that are addressed by a third strategy we introduce here. We used data from chromosome 22 genotyped on Illumina 2.5M (32,903 SNPs) and 1M (14,071 SNPs) chips to illustrate properties of the three strategies. We demonstrate that the union strategy introduces bias (false positives) for SNPs available only on one array, presumably because the genotypes of some subjects are experimentally determined while others were imputed. The intersection strategy does not introduce this bias, but loses power for SNPs present only on one chip, due to the exclusion of the experimentally determined genotypes. We propose a hybrid approach that utilizes the genotypes for SNPs present only on one chip whenever the SNP has high imputation quality (defined by the Imputation Quality Score (IQS) ≥ 0.9). The IQS is a score we previously introduced based on Cohen's kappa rater agreement statistic. Using this approach led to good results with few spurious associations (5 false positives among 134,293 SNPs). In contrast, replacement of SNPs having high scores according to the IMPUTE2 internal quality metric gave rise to a high proportion of spurious associations (126 false positives among 131,554 SNPs). We compare the increase in power for our new approach to results from imputation based on the intersection of SNPs from the two arrays.

1788T

Testing Genetic Association at Untyped Rare Variants, an alternative to imputation based two-step approach. *K. Ye.* Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY., USA.

In recent years, we saw three important developments that empower us to investigate genetic association of rare variants (whose allele frequencies are below 5%) in genome-wide association studies. The first is the rise of large consortia, in which researches pool their samples together for meta-analysis. The second is the near completion of the 1000 Genome Project that provides detailed population genetic information on those rare variants and produces reference panels of haplotypes for a number of populations. The third, which links the first two, is the development of software tools such as Beagle, IMPUTE and MACH, that use a reference panel of haplotypes to impute untyped variants based on genotyped markers. While imputation provides convenience and flexibility to test association at untyped variants, a large proportion of imputed variants, mostly rare variants, are often excluded in subsequent association test because of poor imputation quality. In addition, the imputations are perform without considering disease status of subject, resulting biases at the risk alleles in disease populations, subsequently lower the statistical power of the genetic association tests at the imputed variants. Here, we present a novel association test that retains statistical powers especially at those rare variants with poor imputation quality. The new approach can be viewed as a special case of haplotype association tests, but it uses the information from haplotype reference panels, the same as those used for imputation. In short, for an untyped risk variant and a set of haplotypes at that locus, the expected frequencies of the haplotypes in a disease population are functions of (i) the relative risk of the risk allele, (ii) the proportion of risk allele on each haplotype. Therefore, once the proportion of a risk allele on each haplotype is known, one can obtain a maximum likelihood estimate of the relative risk, and perform the likelihood ratio test for the association. Recursive tree models are applied on a reference panel to select such set of haplotypes at each untyped locus and to estimate the conditional allele frequency. We will show that compared to the association tests on the imputed genotypes, this approach retains much higher statistical power at variants of low imputation quality. This is because, not only we use the information from a reference panel, the difference between disease population and normal population is also considered.

1789F

The more you test, the more you find: massive multiple testing does not promote spurious findings among top hits of association studies. *D.V. Zaykin[1], C.L. Kuo[1], O.A. Vsevolozhskaya[2].* 1) Biostatistics Branch, NIEHS/NIH, Research Triangle Park, NC; 2) Department of Epidemiology and Biostatistics, Michigan State University.

High throughput whole-genome sequencing technology dramatically increased the multiple testing burden of association studies. High level of multiple testing is thought to promote false positives. To keep spurious findings at bay, it is common practice to require that findings should pass a significance threshold adjusted for the number of tests. As the number of tested variants now approached tens of millions, unrealistically large sample sizes would be needed for associations of modest magnitude to reach statistical significance. Paradoxically, in discovery studies where genetic variants are tested in an agnostic manner, "top hits" of a study become increasingly more likely to be true signals as more tests are performed, without any correction for multiple testing. We show by both statistical theory and simulations that when the rate of occurrence of true signals does not decrease as more tests are performed, findings with the smallest P-values are more likely to be genuine in studies with more tests. Our findings are supported by prior observations that GWA studies with many tests enjoy a lower rate of false claims than traditional epidemiological and candidate gene association studies with fewer tests. This discrepancy has been attributed to the adoption of stringent significance thresholds and replication practices (Ioannidis et al., 2011). Replication is extremely important for eliminating erroneous claims due to study biases and errors unrelated to statistical chance. However, our results demonstrate that type-I error rates among best-ranking results of a study in fact diminish as more tests are performed. Our conclusions will be applicable to a wide variety of high-dimensional testing problems and implications are broad. Epidemiologists who tested effects of one hundred random exposures on disease and reported the smallest of one hundred P-values are more likely to be correct in identifying a truly associated effect than their colleagues who tested only ten exposures. With regard to GWAS and sequencing studies, our findings suggest that for a given sample size, researchers should expect the top hits to be increasingly enriched with genuine associations as more variants are tested.

1790W

Efficient Algorithms for Multivariate Linear Mixed Models in Genome-wide Association Studies. *X. Zhou[1], M. Stephens[1,2].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Statistics, University of Chicago, Chicago, IL.

Multivariate linear mixed models (mvLMMs) have been widely used in many areas of genetics, and have attracted considerable recent interest in genome-wide association studies (GWASs). However, existing methods for calculating the likelihood ratio test statistics in mvLMMs are time consuming, and, without approximations, cannot be directly applied to analyze even two traits jointly in a typical-size GWAS. Here, we present a novel algorithm for computing parameter estimates and test statistics (Likelihood ratio and Wald) in mvLMMs that i) reduces per-iteration optimization complexity from cubic to linear in the number of samples; and ii) in GWAS analyses, reduces per-marker complexity from cubic to approximately quadratic (or linear if the relatedness matrix is of low rank) in the number of samples. The new method effectively generalizes both the EMMA (Efficient Mixed Model Association) algorithm and the GEMMA (Genome-wide EMMA) algorithm to the multivariate case, making the likelihood ratio tests in GWASs with mvLMM possible, for the first time, for tens of thousands of samples and a moderate number of phenotypes (<10). With real examples, we show that, as expected, the new method is orders of magnitude faster than competing methods in both variance component estimation in a single mvLMM, and in GWAS applications. The method is implemented in the GEMMA software package, freely available at http://stephenslab.uchicago.edu/software.html .

1791T

Combined association and admixture mapping in Latinos. *A. Brisbin[1], A. Boyd[1], S. Rachid[2], G. Lei[1], A.C. Pereira[3], J.E. Krieger[3], M. de Andrade[4].* 1) UW-Eau Claire, Eau Claire, WI; 2) Carleton College, Northfield, MN; 3) Heart Institute, University of Sao Paulo Medical School, Sao Paulo, Brazil; 4) Biomedical Statistics and Informatics, Mayo Clinic, MN.

The use of ancestry information has the potential to enhance association mapping in admixed individuals, such as Latinos. Most existing methods of combining association and admixture mapping have been developed for African Americans, who typically have ancestry from two populations. In this work, we explore a range of possible approaches for incorporating ancestry information from three populations into association mapping. We compare our methods on simulated data, and find that incorporating either genome-wide ancestry proportions or local ancestry provides a similar increase in performance over methods which ignore ancestry. Finally, we apply our methods to identify candidate loci for systolic blood pressure in Brazilians.

1792F

Detecting epistasis effects via sliced inverse modeling. *J.S. Liu, B. Jiang, Y. Liu, M. Hu, C. Ye.* Statistics, Harvard, Cambridge, MA.

Expression quantitative trait loci (eQTLs) are genomic loci that regulate expression levels of genes. By assaying gene expression and genetic variation simultaneously on a genome-wide basis, scientists wish to discover groups of genomic loci that can affect the expressions of a subset of genes. The problem can be viewed as a multivariate 'regression' with variable selection on both responses (gene expression) and covariates (genetic markers), including also multi-way interactions (epistasis effects) among covariates. Instead of learning a predictive model of expression levels given combinations of genetic markers, we start with an inverse modeling perspective. By conditioning on gene expression levels, we model the genetic markers via a sliced inverse model. An efficient dynamic programming algorithm is developed to determine the optimal slicing scheme. The inverse modeling approach can be effectively used for both independent screening and joint modeling of interactive genetic markers. Through simulation studies and real data examples in multiple tissues, we demonstrate how the proposed method achieves significantly improved power in detecting tissue-common and tissue-specific eQTLs compared to traditional approaches such as stepwise regression methods.

1793W

Admixed-MASTOR: Mixed-Model Association Mapping of Quantitative Traits in Genetically Admixed Samples with Related Individuals. *T. Thornton[1], M.S. McPeek[2].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Statistics and Human Genetics, University of Chicago, Chicago, IL.

While genetic association studies for complex trait mapping have primarily focused on populations of European descent, more recent studies involve populations with admixed ancestry, such as African Americans and Hispanics. Genetic association studies in ancestrally admixed populations offer exciting opportunities for identifying variants that underlie phenotypic diversity. At the same time, the heterogeneous genetic background and dependencies among sample individuals from admixed populations, including population structure and relatedness, pose special challenges for trait mapping. In these circumstances, it is necessary to devise statistical methods for association mapping that account for the diverse genomes of the sample individuals and are robust in the presence of a variety of complex sample structure settings. We propose ADMIXED-MASTOR, a mixed-model, retrospective score test for genetic association with a quantitative trait in the presence of ancestry admixture and relatedness. ADMIXED-MASTOR appropriately accounts for ancestry admixture by incorporating individual-specific allele frequencies in the mixed-model that are calculated on the basis of ancestry derived from whole-genome analysis. We demonstrate that ADMIXED-MASTOR can provide a substantial improvement over existing association methods, such as EIGENSTRAT (Price et al., 2007) and EMMAX (Kang et al., 2010), in terms of power and type 1 error in admixed samples with related individuals. We further demonstrate the utility of ADMIXED-MASTOR with an application to the minority cohort of more than 12,000 African Americans and Hispanics from the Women's Health Initiative study for the identification of genetic variants for a variety of clinical outcomes and quantitative traits.

1794T

Detecting local haplotype sharing and haplotype-phenotype association. *Y. Guan[1], H. Xu[1,2].* 1) Pediatrics and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Southeast Univeristy, Nanjing, China.

We present a statistical method to detect association between local haplotypes and phenotypes. The method relies on a two-layer hidden Markov model (HMM) developed previously to detect structure of local haplotypes. Briefly, we compute probabilities of (unspecified) haplotypes descending from a set of ancestral haplotype clusters, from which we quantify genetic distance between diploid individuals using local haplotype sharing -- the probabilities of two hapotypes descending from same haplotype clusters. We devise a novel model fitting method for the two-layer HMM, reducing the complexity from quadratic to linear (in the number of haplotype clusters) for diploid individuals; thus, it is feasible to apply our method to current genome-wide association studies (GWAS). We build a random effect model to link the genetic distance and phenotypes to test for association. Because the genetic distance is quantified by local haplotype sharing, the associations detected are between phenotypes and (unspecified) haplotypes. Compare to traditional haplotype association methods, our method integrates out phasing uncertainty and avoids arbitrariness in specifying haplotypes. We recast the random effect model to a fix effect model through eigen decomposition of pairwise local haplotype sharing matrix; this allows us to take advantage of existing framework for association testing. We demonstrate the usefulness of our method by analyzing GWAS data sets from Wellocme Trust Case Control Consortium.

1795F

JEPEG: software for testing the joint effect on phenotype of eQTLs in a gene. *D. Lee, V. Williamson, T. Bigdeli, V. Vladimirov, S. Bacanu.* Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.

Expression of genes is known to affect many phenotypes/diseases. Given that gene expression is perturbed by nearby genetic variants commonly known as expression quantitative trait loci (eQTL), these variants are likely to affect many phenotypes in a multivariate manner. While modelling multivariately the action of eQTL/functional SNPs on phenotype is likely to increase detecting power for many traits, it is impeded by i) not measuring all functional SNPs in a gene and ii) the lack of access to individual genotypes. We overcome these obstacles by proposing JEPEG, a novel method/software testing for the joint effect of eQTLs in a gene by using only reported summary statistics. To achieve its objective, JEPEG i) directly imputes the summary statistics at unmeasured eQTLs and ii) tests for the joint effect of measured and imputed eQTLs. The direct imputation of summary statistics is achieved by employing the conditional expectation formula for multivariate normal variates. The joint testing is achieved by using summary statistics of measured and imputed functional SNPs and their genotype correlation matrix, as estimated from a relevant reference population (e.g.1000 Genomes). To decrease the number of degrees of freedom and, thus, increase power, before testing for the multivariate effect JEPEG first pools the genotypic information within functional categories using weights based on biologically relevant eQTL measures stored in a continuously updated database of human genes and eQTLs/functional SNPs affecting their expression. To decrease computational burden, the software implementing the method is written in C++ and is easily compilable to run on a variety of operating systems such as Linux, Windows and Mac. We used JEPEG to analyze summary statistics from the Psychiatric GWAS Consortium (PGC) stage 1 (PGC1). Besides genes reported to harbor significant signals in PGC1, we detected multiple significant genes not harboring significant signals in PGC1, but subsequently reported to harbor significant findings in PGC stage 2. These results strongly suggest that JEPEG can improve the power to detect trait-associated genes by aggregating the functional information from a gene.

1796W

Detection Boundary and Higher Criticism Approach for Rare and Weak Genetic Effects. *Z. Wu[1], Y. Sun[1], S. He[1], J. Cho[2], H. Zhao[3], J. Jin[4].* 1) Mathematical Sciences, WPI, Worcester, MA; 2) Departments of Genetics and of Pediatrics, Yale, New Haven, CT; 3) Departments of Genetics and of Biostatistics, Yale, New Haven, CT; 4) Department of Statistics, CMU, Pittsburgh, PA.

Genome-wide association studies (GWAS) have identified many genetic factors underlying complex human traits. However, these factors only explained a small fraction of genetic contributions to these traits. It was argued that there are many more genetic factors remain undiscovered. These factors, each is individually weak at the population level, distribute sparsely across the genome. In this paper, we adapt the recent innovations on Tukey's Higher Criticism to SNP-set analysis of GWAS and sequencing studies, and develop new theoretic framework in large-scale inference to assess the joint significances of such rare and weak effects for a quantitative trait. In the core of our theory is the so-called "detection boundary", a curve in the two-dimensional phase space that quantifies the effect rarity and effect strength. Above the detection boundary, the overall effects of genetic factors are strong enough and allow for reliable detection. Below the detection boundary, the genetic factors are simply too rare and too weak for reliable detection by any statistical methods. We show that the HC-type methods are optimal in that they reliably yield detection once the parameters of the genetic effects fall above the detection boundary, and that many commonly used SNP-set methods are not optimal. The superior performance of the HC-type approach is demonstrated through simulations and the analysis of a GWAS data set of Crohn's disease.

1797T

A novel method utilizing GWAS data identifies SNPs in the imprinted gene KCNK9 exhibiting parent-of origin effects on BMI. *C.J. Hoggart[1], J.H. Zhao[2], J. Luan[2], F. Gomez[3], G.B. Ehret[4], M. Mangino[5], A. Teumer[6], N. Tšernikova[7], T. Winkler[8], W. Zhang[9], B. Benyamin[10,11], D. Evans[12], S. Vedantam[13], M.R. Jarvelin[14,15], A. Scherag[16], C. Rivolta[17], I. Borecki[3], J. Hirschhorn[18], R. Loos[19], T. Frayling[20], Z. Kutalik[17,21,22].* 1) Department of Genomics of Common Disease, Imperial College London, London, UK; 2) MRC Epidemiology Unit, Institute of Metabolic Science, Box, Addenbrooke's Hospital, Hills Road, Cambridge, UK; 3) Division of Biostatistics and Statistical Genomics, Washington University School of Medicine, St. Louis, MO; 4) Geneva University Hospital, Genève, Switzerland; 5) King's College London, UK; 6) Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; 7) Institute of Molecular and Cell Biology, Tartu, Estonia; 8) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 9) Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK; 10) The University of Queensland, Queensland Brain Institute, St Lucia, Queensland, Australia; 11) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 12) School of Social and Community Medicine, University of Bristol, UK; 13) Endocrine Division, Children's Hospital, Harvard Medical School, MA; 14) Department of Epidemiology and Biostatistics, Imperial College, London, UK; 15) Institute of Health Sciences and Biocenter Oulu, University of Oulu, Finland; 16) Institut für Medizinische Informatik, Biometrie und Epidemiologie, Essen, Germnay; 17) Department of Medical Genetics, University of Lausanne, Switzerland; 18) Broad Institute, Cambridge, MA; 19) Mount Sinai Hospital, New York, NY; 20) Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter; 21) CHUV University Hospital and Faculty of Biology and Medicine, Lausanne, Switzerland; 22) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

It has been hypothesized that some genetic variants exert different effects on certain phenotypes depending on parental origin. Parent-of-origin effects may be caused by epigenetic factors such as methylation and histone-modification which occur in genomic imprinting regions. Only a few family based linkage studies have explored whether SNPs exhibit parent of origin effects on traits such as body mass index (BMI) and alcohol intake. Many of the findings from these studies remain un-replicated and the identified regions are often too large to provide meaningful interpretation. The only large-scale, genome-wide parent-of-origin study by Kong et al. 2009 focused primarily on type 2 diabetes and identified four loci, however, the discovery methodology required a combination of genealogy information and long-range phasing. Here, we present a novel method that is able to detect parent-of-origin effects using genome-wide genotype data of unrelated individuals. We demonstrate that parent-of-origin effects can be identified by increased phenotypic variance in the heterozygous (het) genotype group relative to that in the homozygous (hom) groups. Our rationale is that in the heterozygous group half of the population is mat-A/pat-B, and the other half is pat-A/mat-B, increasing the phenotypic variance in that group in the presence of a parent-of-origin effect. Sixteen GWAS cohorts participated in our discovery analysis totalling ~48,000 individuals. We tested all SNPs in known imprinted regions for parent-of-origin effects on BMI. One SNP reached significance after adjusting for multiple testing (rs2471083 [T/C], variance (het vs. hom): 1.058 vs. 0.963, P = 1.07×10-6). This SNP is located 100kb upstream of the gene KCNK9. Mutations in this potassium channel gene cause Birk-Barel syndrome. SNPs within 2kb have been shown to be associated with HDL cholesterol, adiponectin, and creatinine levels. We replicated the parent-of-origin association in four family-based studies; the combined analysis (of 3,016 heterozygous individuals) confirmed that individuals who carry the C allele paternally have 0.09 (SD unit) higher BMI on average than those carrying the maternal C-allele (P=0.0031). Currently gene expression experiments are underway to investigate whether the variant may influence KCNK9 expression in a parent-of-origin fashion. Our method opens new avenues to exploit GWAS data of unrelated individuals in order to identify parent-of-origin effects.

1798F

Genome-wide imputation in multiple cohorts to study susceptibility to sepsis in Europe. *T.C. Mills[1], A. Rautanen[1], E. Davenport[1], P. Hutton[2], J. Knight[1], T. Meitinger[3], C. Garrard[2], F. Stueber[4], C. Hinds[5], A.V.S. Hill[1], The GenOSept Investigators.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Adult Intensive Care Unit, John Radcliffe Hospital, Oxford; 3) Institute of Human Genetics, Neuherberg, Germany; 4) Berne University Hospital, Berne, Switzerland; 5) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, London.

Sepsis is a major burden to health care worldwide, and severe sepsis still carries a mortality rate of more than 30% in developed countries. In the USA, the number of hospitalisations with severe sepsis more than doubled between 2000 and 2007. In order to better understand the genetic factors which influence susceptibility to sepsis, we have performed a large scale genome-wide association study (GWAS). Community-acquired pneumonia (CAP) sepsis cases were collected and genotyped as part of the GenOSept (European cohort) and GAinS (United Kingdom cohort) projects. These cases were compared to six healthy control GWAS datasets from across Europe. Following stringent quality control, 950 CAP sepsis cases and 5241 healthy controls were included in an association analysis performed using plink. Only those cases and controls which clustered together in multidimensional scaling, and therefore represented similar population ancestry, were included in the final analysis. The value of lambda for this analysis was 1.038 prior to imputation (for an analysis of 36,805 SNPs genotyped in all cohorts), but no SNPs passed the genome-wide significance threshold of 5×10-8. We therefore performed genome-wide imputation using ShapeIT and IMPUTE2, followed by analysis using SNPTEST2. SNPs were removed from the analysis when the minor allele frequency overall or in an individual cohort was less than 2%, and when the info score from imputation in an individual cohort or overall as defined by SNPTEST2 was less than 0.8. Over 5 million SNPs remained following this QC, but the lambda value following imputation was inflated, as imputation across multiple cohorts can generate bias within the data. Only SNPs that showed similar allele frequencies between the two case cohorts were considered to be reliable. This approach enabled us to rule out strong associations that were due to imputation bias rather than true association, highlighting the importance of the second case cohort in this analysis, which could be used as an internal control. Further cleaning of the data is required to better determine which of the associating regions are likely to be real. Replication will be performed using an additional cohort of 564 GAinS CAP sepsis samples to confirm our findings. Our imputation-led approach has allowed multiple cohorts of European ancestry to be analysed and provide what is to our knowledge the first GWAS and the largest study of susceptibility to severe sepsis due to CAP.

1799W

LD Friends: A Method to Disentangle Polygenicity from Population Stratification. *B.K. Bulik-Sullivan[1,2], N. Patterson[1], A. Price[1,4,5], M. Daly[1,2], B. Neale[1,2,3].* 1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 5) Department of Biostatistics, Harvard School of Public Health, Boston, MA.

Recent large-scale genome-wide association studies (GWAS) such as current meta-analyses of height, schizophrenia and Crohn's Disease have reported inflated distributions of SNP association test statistics. A central question in these studies is the extent to which this inflation is driven by confounding bias such as population stratification versus polygenic inheritance. We propose a novel approach to distinguish between bias and polygenicity by leveraging linkage disequilibrium (LD) and the distribution of test statistics. To assess the quality of the method, we simulate phenotypes with varying levels of population stratification, cryptic relatedness and true polygenic association using genotypes from ~2500 controls from a Swedish schizophrenia cohort. We propose the *LD friends* score, defined for a SNP x as $l_x := 1 + \Sigma\, r^2_{xy}$, where the sum is taken over all SNPs y within 500kb of x. Our simulations show that there is a strong linear relationship between *LD friends* and mean ψ^2-statistic in the context of polygenic association, but that *LD friends* and ψ^2-statistic are uncorrelated in the context of population stratification. Furthermore, we show that in the presence of both true polygenic association and population stratification, the intercept from the linear regression of ψ^2-statistic on *LD friends* is informative about the level of bias remaining after correction for principal components. We apply this method to schizophrenia data from the Psychiatric Genetics Consortium and height data from the GIANT Consortium show that the relationship between ψ^2-statistic and *LD friends* is consistent with highly polygenic genetic architectures for schizophrenia and height, and inconsistent with uncorrected confounding from population structure or cryptic relatedness. In addition, we discuss methods for estimating *LD friends* in the context of multi-ethnic association studies with imputed genotypes, as well as the robustness of our method to violations of our model's assumptions and the applicability of our method to mixed linear model association.

1800T

Extended Mantel test for comparing differences in linkage disequilibrium between population. *N. Tanaka.* Biostatistics Section, National Center for Global Health and Medicine, Tokyo, Tokyo, Japan.

Checking the equality of population linkage disequilibrium (LD) structures is an important and often substantial in genetic association studies, especially if the target region includes disease susceptible genetic variants. This problem can be generalized as comparison of correlation structure between two populations. To solve this problem, the likelihood ratio test is commonly used especially in low-dimensional and regularity settings. In genetics, many researchers have been applying statistical procedures in which only elements of correlation matrices are used to calculate model-free test statistics, because the raw data which is used to estimate LD structures cannot be obtained in some cases. One of these is Mantel test (1967), which is originally proposed to compare distance or dissimilarity matrices for solving the problems inherent in explaining species-environment relationships in ecological studies. Mantel statistic is based on a simple cross-product term of each element of matrices, and the p-value is calculated based on the randomized distribution of 'elements'. Thus this statistics represents not the relationship between populations but distance measures (matrices). However, Mantel test is still used even when investigators have the raw data to calculate correlation structure. Then, extended Mantel test statistic is proposed to compare two population structures, not the matrices. To address the variability of each element of matrices estimated in each population, the p-value for testing the null hypothesis, the correlation structure between two populations is equal, is calculated based on the sampling distribution of the weighted Mantel statistic via permutation of the observation. Proposed test is compared with Mantel test and the bootstrap based test which was proposed by Hrafnkelsson et al (2010). It is shown that this extended test statistic has broad utility and would derive a conclusion that agrees with that from other multivariate methods and intuitive insight from visualized method, such as Haploview, through several real examples.

1801F

Bivariate joint phenotype-endophenotype measured genotype association testing in complex pedigrees. *J.W. Kent, T.D. Dyer, M. Almeida, V.P. Diego, J. Blangero.* Dept Genetics, Tx Biomed Res Inst, San Antonio, TX.

Multivariate approaches to genetic analysis of complex diseases offer the promise of increased power to detect contributing genetic factors, as when affection status is combined with quantitative measures, or endophenotypes, of disease. In addition, analysis of endophenotypes may provide both a more focused search as well as insight into the physiological pathways and processes underlying disease. Bivariate joint association testing allows analysis of the (possibly) pleiotropic effect of a genetic variant while explicitly modeling the correlation between traits. However, (1) joint associations may be dominated by the univariate association of one of the correlated traits with the variant; (2) a test of joint association is often underpowered relative to univariate tests because it has two degrees of freedom; and (3) test results may pose problems of causal interpretation. Here we use simulation in SOLAR of genotypes and phenotypes on pedigree data to distinguish three possible causal relationships between a variant V and two correlated traits A, B: (a) $V \to A \to B$; (b) $V \to B \to A$; (c) $V \to A$ and $V \to B$; where (c) represents a non-pleiotropic relationship between the variant and traits correlated via other effectors. (a) and (b) represent a variant that acts indirectly on one trait via its direct effect on an intermediate; since these are equivalent to instrumental variable (IV) models employed in 'Mendelian randomization', our results offer an efficient method for IV model estimation in complex pedigrees as well as estimates of its Type I and Type II error rates.

1802W

A Multi-Sample U-Statistic for Family-based Association Studies. *M. Li[1], Z. He[2], D. Schaid[3], M. Cleves[1], T. Nick[1], Q. Lu[4].* 1) Pediatrics, Univ Arkansas Medical Sciences, Little Rock, AR; 2) Biostatistics, University of Michigan, Ann Arbor, MI; 3) Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 4) Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.

Family-based study design is one of the most commonly used study designs in genetic research. It has many ideal features, including being robust to population stratification (PS). With the advance of high-throughput technologies and ever-decreasing genotyping cost, the analysis of high-dimensional genetic data has become common for family studies. The yield from the analysis of these high-dimensional family-based genetic data can be enhanced by adopting computationally efficient and powerful statistical methods. In this article, we propose a general framework of family-based U-statistic, referred to as family-U, for family-based association studies. Similar to existing family-based methods, it can offer robust protection against PS. In the absence of PS, it can also utilize additional information (i.e., between-family information) for power improvement. The proposed family-U method makes no assumption of the underlying disease models, and is applicable for different phenotypes (e.g. binary and quantitative phenotypes), and various pedigree structures (e.g., nuclear families and extended pedigrees). Through simulations, we demonstrated that family-U attained higher power over a commonly used method, FBAT, under various disease scenarios. We further illustrated the new method with an application to a large-scale family data from Framingham Heart Study. By utilizing additional information (i.e., between-family information), family-U confirmed a previous association of CHRNA5 with nicotine dependence.

1803T

Robust and powerful sibpair test for rare variant association. *K. Lin[1], S. Zöllner[1,2].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychiatry, University of Michigan, Ann Arbor, MI.

Background Modern sequencing technology allows investigating the impact of rare variant on complex disease. However, using a conventional case-control design, large samples are needed to achieve sufficient power for testing association of suspected loci to complex disease. In such large samples, population structure and batch effects can cause spurious signals. One approach to overcome stratification is family-based study design. Traditional methods such as the TDT and the FBAT are not especially powerful when applied to rare variants. Thus, there is a need to develop a method to efficiently use of family data to discover association between rare variants and disease phenotypes. **Method** We propose a novel test of a sample of affected sibpairs. In each sibpair, we estimate number of chromosomes shared IBD at the locus of interest. Based on this estimate, we assess the number of minor alleles located on shared chromosomes and the number of minor alleles on non-shared chromosomes. Shared and non-shared chromosomes are equally likely to contain variants that do not affect disease risk, while shared chromosomes are more likely to carry risk variants. Hence we test for an excess of rare variants on shared chromosomes. This test is robust to stratification as siblings have matched ancestries. Moreover, this design can correct for genotyping error and batch effects. As some regions are sequenced multiple times, sequencing error can be estimated by examining how often a sibpair who share two IBD chromosomes has inconsistent genotype calls and adjust the test to take into account the genotype error probabilities. We evaluate the power of this approach analytically as well as with computer simulations using a general model for the effect size of rare risk variants and considering different models of interaction between the locus of interest and the remaining genome. **Result** For most models with cumulative risk allele frequency <0.05, the proposed design shows superior power over the conventional case-control study given the same number of sequenced samples. **Conclusion** We introduce a new method for analyzing rare variants in families that is robust to population stratification and can incorporate other confounders such as genotyping error and batch effects. This new method is more powerful than a standard case-control design under wide-range of scenarios.

1804F

Analysis of Ultra High-Dimensional Data in Imaging Genetics Studies.
J. Kang, Y.J. Hu. Emory University, Atlanta, GA.

Many neurodegenerative diseases such as Alzheimer's Disease (AD) are highly heritable and have been linked to brain regions of cognitive performance. This has spurred major interest in investigation of such diseases based on genetic and neuroimaging biomarkers. Recent advances in neuroimaging and genetics allow imaging genetics studies to collect both highly detailed brain images (>300,000 voxels) and genome-wide genotype information (>12 million variants). However, there is lack of statistical powerful methods and computationally efficient tools to analyze such very high-dimensional data, which has greatly hindered the impact of imaging genetics studies. In this work, we develop methods for genetic dissection of neuroimaging phenotypes. We build a statistical model that assesses the association between genetic variants and neuroimaging traits. Our method is tailored to the brain-wide, genome-wide association discovery, where both the entire genome and the entire brain are searched for dependence patterns. We also build a regression model that encompasses both genetic and imaging markers to predict the overall risk of the disease. A variable selection procedure is incorporated to search over the whole genome and the whole brain for variants that should be included in the regression model. We implement our methods in a user-friendly, computationally efficient, publicly available and well-maintained software package. We adopt the new computer technology, the General Purpose Computing on Graphical Processing Unites (GPGPU), to make our analysis possible even on a personal computer (PC). We illustrate our method and software on simulation studies and analysis of the imaging genetics data from Alzheimer's Disease Neuroimaging Initiative (ADNI).

1805W

A general regression framework for a secondary outcome in case-control studies.
e. Tchetgen Tchetgen. biostatistics, Harvard UBoston, MA.

Modern case-control genome-wide association studies typically involve in addition to the disease outcome and the genetic information, the collection of data on a large number of outcomes, often at considerable logistical and monetary expense. These data are of potentially great value to subsequent researchers, who, although not necessarily concerned with the disease that defined the case series in the original study, may want to use the available information for a regression analysis relating the genotype data to a secondary outcome. Because cases and controls are selected with unequal probability, regression analysis involving a secondary outcome generally must acknowledge the sampling design. In this work, the author presents a new framework for the analysis of secondary outcomes in case-control studies. The approach is based on a careful re-parametrization of the conditional model for the secondary outcome given the case-control outcome and regression covariates, in terms of (a) the population regression of interest of the secondary outcome given covariates, and (b) the population regression of the case-control outcome on covariates. The error distribution for the secondary outcome given covariates and case-control status is otherwise unrestricted. For a continuous outcome, the approach sometimes reduces to extending model (a) by including a residual of (b) as a covariate. However, the framework is general in the sense that models (a) and (b) can take any functional form, and the methodology allows for an identity, log or logit link function for model (a). The approach is illustrated with an extensive simulation study and a genomewide association study of body mass index using data from a diabetes case-control study. The methods are found to be considerably more efficient than inverse-probability-weighting, and generally less biased than an estimator that stratifies or conditions on case-control status using a standard regression parametrization.

1806T

Mining the Human Phenome Using Allelic Scores that Index Biological Intermediates.
D.M. Evans[1], M.J. Brion[1,2], L. Paternoster[1], J.P. Kemp[1], G. McMahon[1], M. Munafo[3], N.J. Timpson[1], B. St Pourcain[1], D.A. Lawlor[1], A. Dehghan[4], J. Hirschhorn[5,6], G.D. Smith[1]. 1) Soc Med, Univeristy Bristol, Bristol, United Kingdom; 2) Broad Institute, MIT and Harvard, Cambridge, Massachusetts; 3) School of Experimental Psychology, University of Bristol, Bristol, UK; 4) Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands; 5) 6Division of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts; 6) Department of Genetics, Harvard Medical School, Boston, Massachusetts.

It is common practice in genome-wide association studies (GWAS) and their meta-analyses to focus on the relationship between disease risk and genetic variants one marker at a time. When relevant genes are identified it is often possible to implicate biological intermediates and pathways likely to be involved in disease aetiology. However, single genetic variants typically explain small amounts of disease risk. Our idea is to construct allelic scores that explain greater proportions of the variance in biological intermediates, and subsequently use these scores to data mine GWAS. To investigate the approach's properties, we indexed three biological intermediates where the results of large GWAS meta-analyses were available: body mass index, C-reactive protein and low density lipoprotein levels. We generated allelic scores in the Avon Longitudinal Study of Parents and Children, and in publicly available data from the first Wellcome Trust Case Control Consortium. We compared the explanatory ability of allelic scores in terms of their capacity to proxy for the intermediate of interest, and the extent to which they associated with disease. We found that allelic scores derived from known variants and allelic scores derived from hundreds of thousands of genetic markers explained significant portions of the variance in biological intermediates of interest, and many of these scores showed expected correlations with disease. Power calculations confirm the feasibility of extending our strategy to the analysis of tens of thousands of molecular phenotypes in large genome-wide meta-analyses. We conclude that our method represents a simple way in which potentially tens of thousands of molecular phenotypes could be screened for causal relationships with disease without having to expensively measure these variables in individual disease collections.

1807F

Evolutionary Triangulation: A Novel Approach for Filtering Association Results.
M. Huang[1], C. Amos[2], L. Muglia[3], S. Williams[1]. 1) Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH; 2) Department of Family and Community Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH; 3) Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

GWAS analyses have been successful in identifying many genetic associations with common complex diseases. Despite the successes, much of the heritability of complex disease remains unexplained. One possible explanation is that GWAS analyses suffer from an inflated type II error rate due to overly conservative correction for multiple testing. Thus, it is important to develop methods that can minimize this type of error while still controlling type I error. Our hypothesis is that the distribution of genetic variants that associate with a disease will mirror the distribution of the disease among populations, and therefore, we can use the population genetic structure as a filter to reduce type II error. We propose a method to salvage real or likely associations in GWAS by applying evolutionary differentiation as a metric. This method, Evolutionary Triangulation (ET), requires at least three populations with known disease prevalence data such that two have similar prevalence and the third is disparate. We use overlapping patterns of similarity and differences in F_{ST} to identify putatively associating genes. We tested ET by using three HapMap populations (YRI, CEU and GIH), because two are known to have similar prevalence for some cancers (e.g., in GIH and YRI melanoma and ovarian cancer are rare) but the third (CEU) is distinct and has a higher prevalence. Under stringent F_{ST} thresholds we first identified 33 ET SNPs by overlaps of high F_{ST} between CEU and GIH, high F_{ST} between CEU and YRI, and low F_{ST} between GIH and YRI. Then by mapping SNPs back to genes, we identified 35 genes; five of which associated with melanoma, two with skin and hair color, and one with preterm birth. These diseases and traits are concordant with the discrepancy of phenotype distribution among these populations. This ability to identify genes with prior validated associations for disease provides a proof of principle that distribution of genetic variation can serve as an effective filter to find genes that associate with certain traits or diseases. Using a melanoma GWAS dataset and the previously mentioned three HapMap populations, we showed that an ET SNP subset was significantly enriched for GWAS 'hits' ($p<0.05$) compared to a subset of SNPs generated through random sampling, showcasing ET's ability to identify putative GWAS associations. Our method, which is unbiased in terms of gene function, provides a potential means to address inflated type II error.

1808W

Detecting a weak association by testing its multiple perturbations.
M.T. Lo[1,2], W.C. Lee[1,2]. 1) College of Public Health, Graduate Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan; 2) Research Center for Genes, Environment and Human Health, National Taiwan University, Taipei, Taiwan.

Genome-wide association studies (GWAS) have successfully identified several genetic variants associated with complex traits and diseases. Theses variants only account for a small proportion of heritability for traits and have minuscule effects. One therefore needs a study with a very large sample size (the number of subjects, *n*) to detect these weak associations. We refer the method of increasing sample size for power as the 'n-based method'. The n of a study can of course be increased but only to an extent. We propose a novel approach that aims for increasing the sample size in a different direction: the number of variables (*p*), which is denoted as the 'p-based method'. For a putative genetic variant, the p-based method integrates the interactions (or more aptly, the 'perturbations') between the variant and other variables. In contrast to other dimension reduction methods, here we openly take advantage of the very large *p*. Theoretical power calculations and computer simulations show that this p-based method outperforms the traditional n-based method by a wide margin, when *p* can be made very large, say, to the thousands or millions. The method is also applied to a real data from a GWAS of age-related macular degeneration (AMD). The method detects a number of novel genetic variants associated with AMD.

1809T

A Bayesian approach to Detect Differentially Methylated Loci with Next Generation Sequencing using Integrated Nested Laplace Approximation.
L. Shuang, R. Podolsky, D. Ryu, V. George, H. Xu. Georgia Regents University, Augusta, GA.

DNA methylation at CpG loci is an important biomedical process involved in many complex diseases including cancer. In recent years, the development of next-generation sequencing (NGS) has been yielding large amount of sequencing data, which makes the NGS platforms useful for many applications. In this high-throughput sequencing approach, DNA samples are treated with bisulfate, which converts unmethylated cytosines to uracils and leaves methylated cytosines intact. The bisulfate-treated samples are then sequenced with NGS platforms. As a consequence, the NGS data include the counts of methylated molecules or unmethylated molecules at each CpG site for each individual. We introduce a robust Bayesian approach for differential methylation analysis for NGS data, using Integrated Nested Laplace Approximation (INLA) for deriving posterior distributions. We performed extensive simulations to compare our proposed method to existing alternate methods. The simulation results illustrate that our proposed approach can detect more true positive differential methylation sites and less false positives. Additionally, our approach allows us to compute approximations to the posterior marginal directly by using Integrated Nested Laplace Approximation (INLA). Compared to Markov Chain Monte Carlo (MCMC) for generating posterior samples, our approach using INLA is computationally faster, easier to implement and overcomes potential convergence issues associated with MCMC.

1810F

Methodology for mitochondrial phenome-wide association studies.
J. Hall[1], W. Bush[1], S. Pendergrass[2], R. Goodloe[1], J. Boston[1], E. Farber-Eger[1], D. Crawford[1], S. Mitchell[1]. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Center for Systems Genomics, The Pennsylvania State University, University Park, PA.

In this pilot study, we evaluate methods for the analysis of mitochondrial variation on multiple phenotypic outcomes~HDL-C, LDL-C, triglyceride levels, total cholesterol, type 2 diabetes, systolic and diastolic blood pressure, and mean corpuscular hemoglobin. This is a novel application of a phenome-wide association study (PheWAS), a recent innovation in study design where phenotypically rich datasets are examined for relatively few genetic variables. These studies are ideal for detecting pleiotropic effects for susceptibility alleles, which data suggest are prominent for mitochondrial variants. Mitochondrial genetic variation, typically represented as haplogroups, has been associated to a wide variety of traits, though is commonly ignored in genome-wide association studies. Given these associations, and the general biological implications of mitochondrial dysfunction, mitochondrial variants are prime candidates for PheWAS analysis. For this analysis, we accessed data from the Epidemiologic Architecture of Genes Linked to Environment (EAGLE) BioVU study, a subset of the Vanderbilt University biorepository consisting of ~15,000 DNA samples from non-European Americans representing African Americans (n=11,521), Hispanics (n=1,714), and Asians (n= 1,122). Restricting analyses to African Americans only, we estimated the effect of 130 mitochondrial variants captured by the Illumina Metabochip. Using mixed linear models, we calculated the variance explained by mitochondrial genetic variation for each of these phenotypes. We compared models with only nuclear principal components (estimating genetic ancestry) to models containing both nuclear and mitochondrial principal components (representing maternal lineage), additionally adjusting for both age and sex. Addition of mitochondrial principal components did not significantly influence model fits, indicating that adjustment with nuclear principal components is sufficient. Our initial analysis suggests that 2-3% of variance in HDL-C and triglyceride levels may be explained by mitochondrial genetic variation, but that less of an effect is observed for the other phenotypes assessed. In future studies, we will refine the analysis approach and apply it to the full multi-ethnic dataset, for several hundred phenotypes.

1811W

Mechanistic Phenotypes: An aggregative phenotyping strategy to identify disease mechanisms using GWAS data. *J.D. Mosley[1], S.L. Van Driest[2], E.K. Larkin[1], P.E. Weeke[1], J.S. Witte[3], Q.S. Wells[1], J.H. Karnes[1], L. Bastarache[4], L.M. Olson[5], C.A. McCarty[6], J.A. Pacheco[7], G.P. Jarvik[8], E.B. Larson[9], D.R. Crosslin[10], I.J. Kullo[11], G. Tromp[12], H. Kuivaniemi[12], D.J. Carey[12], M.D. Ritchie[13], R. Li[14], J.C. Denny[1,4], D.M Roden[1].* 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University, Nashville, TN; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; 4) Biomedical Informatics, Vanderbilt University, Nashville, TN; 5) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 6) Essentia Institute of Rural Health, Duluth, MN; 7) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 8) Department of Genome Sciences, University of Washington, Seattle, WA; 9) Group Health Research Institute, Seattle, WA; 10) Biomedical and Health Informatics, Seattle, WA; 11) Division of Cardiovascular Diseases, Mayo Clinic, Rochester MN; 12) The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA; 13) Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA, USA; 14) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

A single mutation can alter homeostatic mechanisms and give rise to multiple clinical diseases. Such variants may be missed in association studies focused on single disease, and we test here the hypothesis that broad disease mechanisms can be identified by detecting associations between genes containing low minor allele frequency (MAF) SNPs and 'mechanistic phenotypes', comprised of collections of related diagnoses. Our overall approach is to rank associations between non-synonymous SNPs (nsSNPs) represented on GWAS platforms and mechanistic phenotypes ascertained from electronic medical records (EMRs), and then seek enrichment in functional ontologies across the top-ranked associations. We studied two mechanistic phenotypes: (1) thrombosis, evaluated in a population of 1,655 African Americans; (2) five groupings of cancer diagnoses were evaluated in 3,009 European Americans. We used a two-step analytic approach whereby nsSNPs were first sorted by the strength of their association with a phenotype. The nsSNPs evaluated were selected solely for having a MAF<0.1. Three separate association approaches that focused on minor allele homozygotes were evaluated: a recessive genetic model and 2 models that assigned p-values based on the probability that the prevalence of the phenotype among the homozygotes for the minor allele (HZMAs) exceeded prevalence values derived from genotype-permuted data. The first utilized diagnoses shared by 2 or more HZMAs while the second utilized diagnoses overrepresented among HZMAs with Fisher's p<0.1. In the second step, the top-ranked SNPs for each model, selected using p-value thresholds of 0.01 and 0.05, were tested for enrichment in functional ontologies. For the thrombosis phenotype, 452 nsSNPs met the initial selection criteria, and the top-ranked nsSNPs were enriched in ontologies related to blood coagulation (>10 fold enrichment, Fisher's p<0.002). Enrichment was driven by the F5, P2RY12, PTAFR, F2RL2, HLA-DP1 and LEP genes. A total of 833 nsSNPs were evaluated for the tumorigenesis phenotypes. Ontological enrichment was only observed with the reverse genetics models. The enriched ontologies were related to DNA repair (9.9 fold, p=9.6×10-6) (FANCI, SLX4/FANCP, XRCC1, BRCA1, FANCA, CHD1L) and microtubule processes (5.1 fold, p=1.6×10-4) (NIN, BRCA1, KIF25, DNAH3, MC1R, SFI1, DST). Mechanism-oriented phenotyping using collections of EMR-derived diagnoses can identify fundamental disease mechanisms.

1812T

Pathway-based analysis for GWAS using the extended propensity score method. *U. Lee, SJ. Finch.* Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY.

Many complex diseases are influenced by genetic variations in multiple genes and non-genetic factors. In order to find the association between SNPs and disease, an extension of genomic propensity score (eGPS) (Zhao et al., 2012) was used to correct for bias due to both genetic and non-genetic factors. Pathway analysis, which identifies biological pathway associated with disease outcome, was also used here to improve the power of eGPS. I hypothesize that the type I error rate of this approach will be closer to its nominal value over a wide range of null conditions and that its power will be greater than the sum statistic and principle component analysis based on the simulation study reported here.

1813F

Accounting for gene-by-sex interactions in genome-wide association studies using random effects meta-analysis. *N. Furlotte[1], EY. Kang[1], B. Han[1,2], JW. Joo[1], E. Eskin[1,3].* 1) University of California Los Angeles, Department of Computer Science, Los Angeles CA; 2) Broad Institute, Cambridge MA; 3) University of California Los Angeles, Department of Human Genetics, Los Angeles, CA.

The prevalence of sex-specific effects complicates the analysis of association studies consisting of both males and females. The traditional approach to address this issue in genome wide association is to regress out sex as a covariate from the phenotype. Unfortunately, this approach ignores the potential for a difference in effect sizes between the two sexes and thus may lead to a loss in power. Furthermore, genome-wide gene-by-sex effects introduce additional background phenotypic variance, leading to both power loss and the introduction of false positive associations. In this paper, we present a novel meta-analytic approach for the analysis of genome-wide association studies consisting of both males and females. In our approach, males and females are analyzed separately and the results are combined using a random effects meta-analysis approach allowing for a difference in effect sizes between sexes. We show that by analyzing males and females separately, our method reduces the overall variance in each study leading to an increase in statistical power when applying meta-analysis. One significant challenge encountered in meta-analysis is the potential for between study population structure, resulting in inflation and spurious associations. To combat this, we propose a novel method to correct between study population structure. We apply our method to the Northern Finland Birth Cohort and the Wellcome Trust Case Control Consortium data and for both studies, we show that our method has increased power over a traditional approach while controlling for false positives.

1814W

Increasing the Power of Association: POPFAM Combines Arbitrary Affected Families, Unrelated Cases, Controls, and Reference Samples. *W. Stewart[1,2], M. Monti[3].* 1) Nationwide Children's, Columbus, OH; 2) The Ohio State University; 3) University of Pavia.

We have extended our genetic association software: POPFAM so that in addition to case-parent triads and unrelated controls, the genotypes of large affected families, unrelated cases, and (optionally) reference individuals (e.g. HapMap samples) can easily be incorporated into a single, more powerful, test of association. Furthermore, in order to maintain POPFAM's speed and flexibility, we have adapted our most recent release: POPFAM-v2.0_beta to a parallel computing environment. In this setting, workloads are distributed across multiple cores of a single machine and/or across multiple nodes of a cluster. Overall, by combining data across heterogeneous designs (e.g. ascertained affected family studies with variable pedigree structures, case-control studies with unequal sample sizes, and unphenotyped individuals), POPFAM-v2.0_beta gives researchers the ability to increase power for detecting genetic associations. Our software is distributed as a part of the larger package EAGLET, which is freely available from the worldwide web at: http://www.mathmed.org/wclstewart/SOFT/soft.html. Based on simulations, we find that POPFAM-v2.0_beta increases power relative to several commonly used methods, and that it compares favorably with WQLS (Bourgain et al. 2003).

1815T

Modified Random Forest Algorithms For Analysis of Matched Case Control Data or Case-parent Trio Data. *Q. Li, J. Bailey-Wilson.* Human Genome Research Institute, NIH, Baltimore, MD.

Random forests (RF) is a machine-learning method useful to detect complex interactions among genetic markers related to a disease trait based on case-control samples. We propose a new modification of the RF algorithm for matched case-control, or family based (trio) data analysis. RF is an ensemble method, which analyzes data and summarizes results using a large number of classification trees. During the procedure, each classification tree uses a proportion of samples and a subset of predictors. An R package, rpart, has functions implementing classification tree analysis and it can be modified to accommodate different study designs by substituting its functions of classification based on a novel criterion. For ease of implementation, our method utilizes the rpart package to conduct classification tree analysis on a subset of the samples and predictors. Then our ensemble code, also written in R, summarizes results from all trees. For matched case-control, or case-parent trio data, we sample the set of samples (in a matched set, or matched case, pseudo-controls set) to be fit to each classification tree. Different classification criteria are also proposed to accommodate the matched study design. To evaluate our method, we simulated matched case-control, and case-parent trio data, and applied our method to select the top 1% most important predictors. The results are compared with other machine-learning methods applicable for matched case-control data, including RF++, MDR, and trio Logic Regression.

1816F

Exploiting network methodology for rare variant association analysis. *H. Fier[1], J. Hecker[1], D. Prokopenko[1], P. Costa[1], S. Won[2,3], K. Ludwig[4,5], E. Silverman[6], M. Mattheisen[7], M.M. Noethen[4,5,8], E. Mangold[4,5], C. Lange[1,6,8,9].* 1) Dept. of Genomic Mathematics, University of Bonn, Bonn, Germany; 2) Department of Applied Statistics, Chung-Ang University, Seoul, Korea; 3) The Research Center for Data Science, Chung-Ang University, Seoul, Korea; 4) Department of Genomics, Life & Brain Center, University of Bonn, Bonn; 5) Institute of Human Genetics, University of Bonn, Bonn; 6) Channing Laboratory, Brigham and Women's Hospital, Boston; 7) Department of Biomedicine, Aarhus University, Denmark; 8) German Center for Neurodegenerative Diseases (DZNE), Bonn; 9) Department of Biostatistics, Harvard School of Public Health, Boston.

In this communication, we propose a new analysis approach for rare variant analysis that exploits network methodology. The approach is designed for large genomic regions, varying effect directions, etc. and enables the analysis of the relationships between the variants. The benefits of incorporating network methodology in genetic research has been shown by a number of current works, e.g. for prioritizing candidate disease genes (Akula et al. 2011, Kacprowski et al. 2013) or protein-protein networks (Guney and Oliva 2012). First, we construct networks based on the aggregated allele counts of rare variants for cases and controls. Subsequently, we introduce a statistic that tests the centralization tendency of the derived networks. In a simulation study, we assess the power of the methodology and compare it with standard rare variant association approaches. We illustrate the practical applicability of the approach by an application to a sequencing study for nonsyndromic cleft lip with or without cleft palate.

1817W

Joint Association Analysis for Family-based Sequencing Data Using a Family-Genetic Random Field Method. *Z. He[1], M. Li[2], M. Zhang[1], X. Zhan[1], Q. Lu[3].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR; 3) Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.

Advance in high-throughput sequencing technologies has revolutionized genetic study of complex diseases, especially the discovery of associated rare genetic variants. However, the high-dimensional feature of the sequencing data and the low frequency of the rare variants limit the power of detecting rare variants in population-based sequencing studies. As rare variants likely aggregate in pedigrees, family-based sequencing studies have substantial potential to enhance the power of detecting rare variants, remaining the advantage of being robust to population stratification (PS). In this paper, we propose a family-genetic random field method (F-GenRF) to test for association of rare and common variants in a region with complex diseases in family while adjusting for covariates. The new method is computationally efficient with following advantages: 1. It accommodates to a variety of disease phenotypes (e.g., quantitative, dichotomized or count phenotypes), and various pedigree structures (e.g., nuclear families and extended pedigrees); 2. It offers robust protection against PS using within-family information, and is able to utilize between-family information for power improvement in the absence of PS; and 3. It is robust to the misspecification of within-family correlation structure. Through extensive simulations, we demonstrated these advantages by comparing the proposed F-GenRF with several existing association methods in family-based sequencing study.

1818T

Simulation study for rare variants approaches in family-based and case-control data. *C. Herold[1,2], H. Fier[3], J. Hecker[3], D. Prokopenko[3], C. Lange[1,3].* 1) Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA; 2) German Center for Neurodegenerative Diseases (DZNE), D-53175 Bonn, Germany; 3) Department of Genomic Mathematics, University of Bonn, 53127, Germany.

After the great success of genome-wide association studies (GWAS) the current research focus now on the development of high-throughput sequencing technology to address the "missing heritability" problem. The majority of these high-throughput sequencing loci are rare variants i.e. loci with a minor allele frequency of less than 1% and therefore most of the single marker analyses for genetic association, e.g. Armitrage-trend test, do not provide sufficient power [Lange and Laird, 2002]. For this reason, several new approaches for rare variants have been developed like the burden tests Combined Multivariate and Collapsing (CMC) test [Li and Leal, 2008], weighted sum approaches [Madsen and Browning, 2009; Price et al., 2010] and a replication-based strategy [Ionita-Laza et al., 2011] or non-burden tests like SKAT [Wu et al., 2011] and C-Alpha [Neale et al., 2011], but mostly for unrelated individuals. In our study we adopted these methods to family-based data, including a spatial approach which is taking in account the physical location of the variants [Fier et al., 2012; Bonetti and Pagano, 2005]. The advantages of family-based data are the robustness against population structure and they tend to be more powerful due to the increased likelihood of affected relatives to share the same rare disease variants [Laird and Lange, 2010]. Furthermore, the selection appropriate controls for rare variants in case-control studies can be a challenging task because of the very low frequencies. We will present the results of our simulation study for different rare variant approaches in family-based and case-control data which show that the power can be increased using family-based instead of case-control data.

1819F

Finding co-regulated transcripts associated to cooperating eSNPs. *A. Kreimer[1,2], I. Pe'er[3].* 1) Department of Biomedical Informatics, Columbia University, 622 west 168th St. New York, NY 10032, USA; 2) Center of Computational Biology and Bioinformatics, Columbia University, New York, NY 10032, USA; 3) Department of Computer Science, Columbia University, 500 west 120th St. New York, NY 10027, USA.

Associations between the level of single transcripts and single corresponding genetic variants, eSNPs, have been extensively studied and reported. However, most expression traits are complex, involving the cooperative action of multiple SNPs at different loci affecting multiple genes. Finding these cooperating eSNPs by exhaustive search has proven to be statistically difficult. In this paper we utilized sequencing data with transcriptional profiles in the same cohort to identify two kinds of usual suspects: eSNPs that alter coding regions or eSNPs within the span of transcription factors (TFs). We devised a computational framework for examining pairs of such cooperating source eSNPs that are both associated with the same two target transcripts. We characterize such quartets through their genomic, topological and functional properties. We establish that this regulatory structure of quartets is frequent in real data, but is rarely observed in permutations. eSNP sources are mostly located on different chromosomes and away from their targets. In the majority of quartets, SNPs affect the expression of the two gene targets independently of one another, suggesting a mutually independent rather than a directionally dependent effect. Furthermore, the directions of effects within quartets are consistent, so that the two source eSNPs either having the same or opposite effect on the target transcripts. Same-effect eSNPs are observed more often than expected by chance. Overall, our analysis offers insights concerning the fine structure of human regulatory networks.

1820W

Genome-wide association detection power increased by observation of cis-acting mRNA. *A. Renwick[1], J.W. Belmont[2], C.A. Shaw[2].* 1) SCBMB, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Discovery of genetic loci causally associated with a trait of interest is hindered by the small effect typically attributable to any one locus. For alleles whose effect is mediated by a change in level of expression, mRNA offers an intermediate observation which closes the gap between allele status and trait. For such alleles, including mRNA data together with genotype and trait increases the power to detect true association.

1821T

Is the Tie-Corrected Mann-Whitney-Wilcoxon Test a Promising Alternative to the Cochran-Armitage Test in the Analysis of Genetic Association Studies? *S. Wellek*[1,2]. 1) Dept of Biostatistics, CIMH Mannheim/Univ of Heidelberg, Mannheim, Germany; 2) Dept of Medical Biostatistics, Epidemiology and Informatics, Univ of Mainz, Germany.

The most widely used procedure of testing for association between a SNP and disease status in genetic association studies is the Cochran-Armitage (CA) test. As argued by Wellek & Ziegler (2012) (Hum Hered 73:14-17), a conceptually more appealing though asymptotically equivalent variant of the CA test is Wald's maximum likelihood test about the regression coefficient associated with the SNP under analysis. Since none of these tests is uniformly most powerful, the question of possible competitors providing better prower against certain alternatives arises. With a view to ease of implementation in the analysis of GWAS, a promising alternative choice is the classical Mann-Whitney-Wilcoxon (MWW) test corrected for the large amount of ties occurring in SNP data taking on only three possible values (0,1,or 2). The power of both the CA and the MMW test can be computed exactly for sample sizes of up to 50 per group with reasonable execution time of the programs. For larger sample sizes, one has to rely on Monte-Carlo simulation. The objective of this contribution is to identify alternatives against which the MWW test provides higher power as compared with the CA test.

1822F

Urinary monocyte chemoattractant protein-1 (MCP-1/CCL2) and its association with renal allograft rejection. *A. Raza*[1], *S. Firasat*[1], *A. Abid*[1], *T. Aziz*[2], *M. Mubarak*[3], *S.A.A. Naqvi*[2], *S.A.H. Rizvi*[2], *S.Q. Mehdi*[1], *S. Khaliq*[1,4]. 1) Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 2) Department of Urology, (SIUT), Karachi, Pakistan; 3) Department of Histopathology, (SIUT), Karachi, Pakistan; 4) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan.

Abstract: Immune regulatory molecules are major players in renal allograft rejection. Among these, the chemokine, monocytes chemoattractant protein 1 (MCP-1/CCL2), has been reported as an independent predictor for late renal allograft loss. In this study, an attempt to device a non invasive marker, urinary MCP-1/CCL2 levels and their association with rejection episodes were determined in renal transplant patients from Pakistan. A total of 440 urine samples were selected for this study that include biopsy proven cases of rejection (187), non-rejection (48, No abnormality detected; NAD) and interstitial fibrosis and tubular atrophy; IFTA (47). Additionally, stable graft (42) and healthy controls (116) were also analyzed. The study was approved by the Institutional Ethical Review Committee and conformed to the Tenets of the Declaration of Helsinki. The quantification of urinary MCP-1/CCL2 levels was measured using the Human MCP-1/CCL2 Quantikine ELISA kit. The data were analyzed using the Statistical Package for Social Sciences (SPSS) and Med Calc softwares. The mean values of urinary MCP-1/CCL2 levels among rejection, IFTA, stable graft, NAD and healthy controls were 926.2±65.8pg/mL, 306.5±87.8, 253.8±57.3, 297±39.2 and 48.2±20pg/mL respectively. MCP-1/CCL2 levels were significantly different between the rejection vs. controls (p=<0.0001), rejection vs. stable graft (p=0.0001), rejection vs. NAD (p=<0.0001) and rejection vs. IFTA (p=<0.0001). Significant difference were also found among IFTA vs. control (p=<0.0001) and IFTA vs. NAD (p=0.024). The ROC curve analysis also showed significant differences among these groups. The area under curve (AUC) for rejection vs. control was 0.97. At a cutoff value 94.5pg/mL, the sensitivity and specificity were 95% and 91%. However, in the rejection vs. NAD, rejection vs. stable graft and rejection vs. IFTA, the AUC were 0.76,0.82 and 0.82 with a sensitivity of 55%.82% and 78% a specificity of 85.4%,74% and79% respectively. The results showed increased levels of urinary MCP-1/CCL2 in the allograft rejection patients. This non-invasive investigation of MCP-1/CCL2 levels in urine, with others markers, may help in monitoring early rejection episodes.

1823W

A nonparametric Bayesian model for clustering time course gene expression profiles. *D. Manandhar*[1], *B. Engelhardt*[1,2,3]. 1) Institute for Genome Sciences and Policy, Duke University, NC 27708, USA; 2) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27708, USA; 3) Department of Statistical Science, Duke University, Durham, NC 27708, USA.

Time course gene expression profiles have been used to understand numerous dynamic biological processes involving the cell-cycle, gene interactions, and genomic response to specific biochemical stimuli. In order to analyze these data, we have developed a nonparametric Bayesian mixture model that clusters the gene expression profiles using a Gaussian process (GP) to model gene expression trajectories. Numerous models -- such as mixed-effects model with B-splines, hidden Markov models, and models that treat gene expression values as samples from multivariate normal or t-distributions - have been developed for time-course gene expression data. Each of these models has analytic limitations that our GP-based nonparametric Bayesian mixture model handles efficiently: a) we use a GP prior as the base distribution for Dirichlet process mixture model, where each time series is a continuous distribution instead of discrete time points; b) we let the number of clusters be determined by the data; c) the model is robust to missing or noisy time point measurements; d) GPs naturally capture equally spaced time interval experiments or unequal intervals. This model assumes gene expression levels at adjacent time points are correlated; moreover, we allow cluster-specific models of time-point correlation to capture the dynamic level of interdependence between time-points in different clusters. Each cluster of genes can thus be interpreted, identifying gene products that respond to a stimulus early and fast, and others that respond slowly or not at all. Our model has a natural Bayesian test to quantify whether a specific gene expression trajectory belongs to a cluster, and also implicitly estimates a similarity matrix between genes that can be used in downstream analyses, including gene network models. When we applied this model to gene expression profiles capturing the response of human fibroblast cells to serum, we found dozens of distinct, interpretable gene expression patterns, versus ten clusters identified in the original analysis. We validate the robustness of clusters produced by out-of-sample prediction rates, by looking for enriched biological processes within cluster members, and by analyzing the mechanistic interpretation of each cluster trajectory. The larger number of distinct patterns of gene expression response suggests additional modes of transcriptional regulation in serum-exposed fibroblast cells, which we aim to study further.

1824T

Enrichment of functional information (545 annotation tracks) in GWAS hits. *S.A. Gagliano*[1,2], *M.R. Barnes*[3], *M. Weale*[4], *J. Knight*[1,2]. 1) Centre for Addiction and Mental Health, 250 College Street, Toronto, Ontario M5T 1R8 Canada; 2) Institute of Medical Science, University of Toronto, 1 King's College Circle, Toronto, Ontario, M5S 1A8 Canada; 3) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ, UK; 4) Dept. of Medical & Molecular Genetics, King's College London, Guy's Hospital, London SE1 9RT, UK.

There are limits to genetic discovery based on association information alone: the logical next step is to combine sources of information in our search for novel causal variants. We assessed the frequency of Genome-Wide Association Study (GWAS) hits in SNPs on common arrays which directly, or via linkage disequilibrium (r^2>0.8), overlap with functional information to determine whether combining functional information with GWAS data is informative. We looked at 1 936 864 SNPs from the 22 autosomes, which was the total number of SNPs obtained upon the union of three common GWAS panels. Linkage disequilibrium proxies were determined using data from the (N=379) Europeans (Phase I, version 3) in the 1000 Genomes Project. We defined GWAS hit SNPs as those in the Catalogue of Published Genome-wide Association Studies from the National Human Genome Research Institute (Hindorff et al., 2009). After filtering out hit SNPs that were not on the original GWAS panels, we were left with 8405 hit SNPs. Annotations came from several sources. The ENCODE Project data investigated for this project include: transcription factor binding sites, three histone modifications (H3K4Me1, H3K4Me3, H3K27Ac), DNase I hypersensitive sites, and two conservation scores (PhyloP and PhastCons). When applicable, peak data, based on regions of statistically significant enrichment, were used. Expression quantitative trait loci data for cis-acting variants (+/-1Mb of relevant gene) were also used. These data were from the NIH GTEx Roadmap Project and the UK Brain Expression Consortium. Additional information, regarding whether a variant is non-synonymous, within a splice site, or within a gene was also assessed. GWAS SNPs were categorized as annotated or not, except for the conservation scores, which were left as quantitative measures. All functional SNPs showed statistically significant enrichment for hit SNPs, except splice sites, which is likely due to an overall low frequency of splice sites compared to the other annotations. These results are being used to refine a methodology to identify SNPs that are likely to be causal based on their functional attributes, rather than relying solely on statistical results from GWAS.

1825F

Use of complementary expression-based data provides improved gene-based prioritization of Crohn's disease associated loci. *K. Ning, K. Gettler, J.H. Cho, NIDDK Inflammatory Bowel Disease Genetics Consortium.* Digestive Diseases, Department of Internal Medicine, Yale University, New Haven, CT, USA.

Purpose: We have recently identified 144 Crohn's disease (CD) loci containing 1,404 candidate genes in a large case-control meta-analysis. Bioinformatic analyses and gene co-expression networks identified a striking overlap between CD loci and genes implicated in mycobacterial susceptibility and differentiation of macrophage subsets (M1, M2). We sought to improve prioritization of CD-associated genes by integrating GWAS results with a variety of gene features. **Methods:** We developed a reference list of genes with strong evidence of involvement in CD pathology and tested for enrichment of a variety of different gene features, including expression level in the intestine, disease-based differential gene expression, eQTL data, association with other auto-immune diseases and gene-level CD association p-value. We then built a logistic regression model to optimize weights of gene features for prioritizing CD-associated genes. Genes having strong evidence of involvement in CD pathology were used to label the dependent variable and all features under study were used as predictors. Step backward model selection was then carried out to keep the most important predictive features in the model. **Results:** In our final model, six features enriched in the reference genes were kept after model selection and their weights were estimated. After building the regression model, we assessed its performance regarding its statistical fitness and its capacity to provide biological insight: (1) Analyses of area under the model's receiver operator curve with permutations showed that our model fit the data well statistically (AUC = 0.9), (2) Comparison of top genes prioritized by our model and by GWAS-based model showed that, unlike genes implicated solely by association evidence, genes specifically prioritized by our model were more enriched in CD-relevant pathways, notably IFNγ signaling (FDR adjusted p-value<1exp-7), and (3) Genes specifically prioritized in our model were significantly more likely to be differentially expressed between M1-M2 macrophage subsets, (which is modulated by IFNγ) compared with genes implicated by GWAS alone (p-value=2.4exp-5). Taken together, our model successfully integrated GWAS data with other genomic data to prioritize genes involved in CD pathology: by prioritizing genes within genome-wide significant loci and by salvaging genes lacking genome-wide significant association, but still having important roles in disease pathogenesis.

1826W

The PhenX Toolkit: discovering and promoting opportunities for cross-study analyses. *W. Huggins[1], H. Pan[1], D.M. Nettles[1], E. Eubanks[1], T. Hendershot[1], J.G. Pratt[1], D. Maiese[1], W.R. Harlan[2], J. Haines[3], H.A. Junkins[4], E.M. Ramos[4], C.M. Hamilton[1]*. 1) RTI International, Research Triangle Park, NC; 2) Retired, Associate Director for Disease Prevention, Office of the Director, National Institutes of Health, Bethesda, MD; 3) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 4) National Human Genome Research Institute, Bethesda, MD.

The PhenX (consensus measures for **Phen**otypes and e**X**posures) Toolkit (https://www.phenxtoolkit.org/) is an online catalog of 339 measures of phenotypes and exposures for use in genomic and epidemiologic research. The PhenX Toolkit has 1,094 Registered Users and has been accessed by more than 100,000 visitors from 150 countries. Covering a broad scope of 21 research domains (Demographics, Cardiovascular, Psychosocial), PhenX measures can be used to expand subject assessment and improve opportunities for cross-study collaboration. Investigators can find measures of interest by browsing or searching the Toolkit using the Smart Query Tool. For each measure, the Toolkit provides a description of the measure, the rationale for its inclusion, detailed protocol(s) for collecting the data, and supporting documentation. The Toolkit also provides custom data collection worksheets to support data collection and custom data dictionaries that support data submission to the database of Genotypes and Phenotypes (dbGaP). To support investigators who want to collect data via the Web, initial efforts have been made to develop Web-based versions of PhenX protocols. To promote cross-study collaborations, the Toolkit includes a new 'Register Your Study' feature allowing registered Users to browse basic information about each registered study (PI, research focus, number of subjects, study design) and a list of the PhenX measures being used. To discover related variables among all completed NIH funded GWAS, PhenX measures have been mapped to 19 studies in dbGaP, and we plan to map to the remaining studies. The dbGaP advanced search tool includes 'PhenX' as a filter option and the dbGaP variable descriptions present dbGaP-PhenX variable classifications (identical, comparable or related). To help researchers find and share data in other resources, PhenX concepts are included in the National Center for Biomedical Ontologies (NCBO) BioPortal. Using BioPortal tools, researchers can search for and identify relationships between PhenX terms and terms from over 200 ontologies, automatically annotate textual descriptions of data with PhenX concepts, and use PhenX terms to simultaneously search across 38 data repositories. PhenX measures are complementary to other ongoing standard measures initiatives and a comparison of available resources that provide researchers with recommended measures and protocols will be presented. Funding provided by NHGRI 5U01HG004597 and 3U01HG004597-03S3.

1827T

Kernel Machine Methods for Joint Testing and Integrative Analysis of Genome Wide Methylation and Genotyping Studies. *N. Zhao, M. Wu.* Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.,USA.

Comprehensive understanding of complex trait etiology requires examination of multiple sources of genomic variability. Integrative analysis of these data sources promises elucidation of the biological processes underlying particular phenotypes. Consequently, many large GWAS consortia are expanding to simultaneously examine the joint role of DNA methylation. Two practical challenges have arisen for researchers interested in joint analysis of GWAS and methylation studies of the same subjects. First, it is unclear how to leverage both data types to determine if particular genetic regions are related to traits of interest. Second, it is of considerable interest to understand the relative roles of different sources of genomic variability in complex trait etiology, e.g. whether epigenetics mediates genetic effects, etc. Therefore, we propose to use the powerful kernel machine framework for first testing the cumulative effect of both epigenetic and genetic variability on a trait, and for subsequent mediation analysis to understand the mechanisms by which the genomic data types influence the trait. In particular, we develop an approach that works at the gene/region level (to allow for a common unit of analysis across data types). Then we compare pair-wise similarity in the trait values between individuals to pairwise similarity in methylation and genotype values for a particular gene, with correspondence suggestive of association. Similarity in methylation and genotype is found by constructing an optimally weighted average of the similarities in methylation and genotype. For a significant gene/region, we then develop a causal steps approach to mediation analysis at the gene/region level which enables elucidation of the manner in which the different data types work, or do not work, together. We demonstrate through simulations and real data applications that our proposed testing approach often improves power to detect trait associated genes, while protecting type I error, and that our mediation analysis framework can often correctly elucidate the mechanisms by which genetic and epigenetic variability influences traits. A key feature of our approach is that it falls within the kernel machine testing framework which allows for heterogeneity in effect sizes, nonlinear and interactive effects, and rapid p-value computation. Additionally, the approach can be easily applied to analysis of rare variants and sequencing studies.

1828F

Integrating Multiple Correlated Phenotypes for Genetic Association Analysis Through Heritability. *J. Zhou[1], M. Cho[2,3,4], E. Silverman[2,3,4], N. Laird[1]*. 1) Harvard Sch Public Health, Boston, MA; 2) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA.

In genetic studies of complex diseases, many correlated disease variables may be analyzed in the hope of increasing power to detect causal genetic variants. A common statistical approach involves assessing the relationship between each phenotype and each single nucleotide polymorphism (SNP) individually and using a Bonferroni correction for the effective number of tests conducted. Alternatively, one can apply a multivariate regression or a dimension reduction technique, such as principal components analysis, and test for the association with the principal components of the phenotypes rather than the individual phenotypes. Other previous approaches have developed methods for combining phenotypes to maximize heritability at individual SNPs. These approaches are not practical for population sample with genome-wide scans. In this paper, by taking advantage of the estimated heritability and co-heritability, we construct a maximally heritable phenotype which is a linear combination of the various phenotypes. Our approach estimates heritability globally and is therefore applicable to genome-wide scans. Theoretically, and through simulations, we compare our approach with commonly used methods and assess both the heritability of the overall phenotype and the power. Moreover we provide a guideline of how to choose the phenotypes for combination. Applications of our approach to a COPD genome-wide association study show the practical relevance.

1829W

Analyses assessing enrichment of GWAS variants for non-coding annotations in the genome are upwardly biased. *G. Trynka[1,2], B. Han[1,2], K. Slowikowski[1,2], H. Xu[3], XS. Liu[3], S. Raychaudhuri[1,2,4]*. 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA, USA; 4) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

ENCODE and other efforts are annotating the non-coding genome, e.g with DNase hypersensitivity (DHS) sites. Investigators now commonly assess whether these annotations might link trait-associated variants (e.g from GWAS) to functional, non-coding regions. A ubiquitously applied approach is to quantify the overlap of GWAS SNPs (or SNPs in tight LD) with specific annotations, and assess if it exceeds chance. Standard practice is to assess statistical significance by comparison to randomly sampled SNPs, matched on minor allele frequency (MAF) or distance to transcription start site (dTSS). But, since GWAS is biased towards genomic regions with higher gene density and LD, standard enrichment analysis for annotations that cluster around genes might result in inflated statistics. To investigate this possibility, we compiled publicly available DHS regions from 217 samples collectively spanning 16.4% of the genome. Then to quantify type I error under different models, we simulated six different sets of GWAS catalogs, with all causal variants drawn from a single genomic structure (promoter, coding, intronic, 5' UTR, 3' UTR, or intergenic). Of these, only promoters should overlap with DHS peaks. We then tested significance of enrichment by randomly drawing 1000 SNP sets matched for different parameters: 1) MAF, 2) dTSS, 3) number of LD partners ($r2 > 0.8$). Surprisingly, matching null SNPs only on MAF and dTSS resulted in $p<0.001$ in 100% of instances. We observed that only by additionally matching on LD partners we were able to achieve appropriate type 1 error rates. As an alternative, we also defined null distributions by shifting DHS annotations locally by a random value between 5kb to +5kb. Using this approach, we observed close to the expected (5%) false positive rate at $p<0.05$ within coding (9%), intronic (11%), 5' UTR (4%), 3' UTR (10%), and intergenic regions (7%); but we had high power to detect DHS enrichment for promoter catalogs (95%). Conclusion: SNP enrichment strategies can be upwardly biased to detect false associations. We present two strategies to address this: 1) match SNPs stringently on the dTSS, MAF and the number of LD partners or 2) shift annotations locally to define null distributions. It is critical that, as investigators assess whether trait associated SNPs are relevant to specific annotations, these approaches are used to avoid dramatically inflated statistics.

1830T

dbVÖR: An open source database system for managing phenotype and genotype information for complex trait studies. *R.V. Baron[1], Y.P. Conley[2], M.B. Gorin[3], D.E. Weeks[1,4]*. 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Health Promotion and Development, School of Nursing, University of Pittsburgh, Pittsburgh, PA, USA; 3) Department of Ophthalmology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; Jules Stein Eye Institute, Los Angeles, California, USA; 4) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

In genetic studies of complex traits, accurate data management is vitally important, yet often challenging. As technological advances permit genotyping at an ever increasing numbers of markers, handling such large scale data in a reasonable amount of time requires clever computational approaches. When samples are genotyped at the same marker multiple times, a consensus genotype for use during the analyses needs to be decided upon. Data cleaning will require samples or marker ranges to be discarded or re-measured. To address these issues, we have developed a portable open source database system, built using Python and MySQL, named dbVOR, after Vör, an inquiring Norse goddess of wisdom from whom nothing can be concealed. To handle large amounts of genome-wide genotype data, instead of storing genotypes individually, one by one, these data are split into chunks, which are then stored in blocks in the database. dbVOR also can store experiments done with small numbers of markers in a conventional way, an individual genotype at a time. To resolve multiple genotypes of samples, we generate agreement matrices displaying the genotype concordance, and either can output the consensus genotype if all genotypes agree, or the user can specify which genotype is more trustworthy and to be used. dbVOR is run from the command line, controlled by a configuration file. dbVOR has facilities for importing data in Illumina and Affymetrix and tabular formats, for storing build-specific genetic and physical maps, and for handling marker and person aliases (e.g., to help resolve multiple different ID systems). dbVOR outputs in a variety of formats, including PLINK binary format and Mega2 annotated format. It also supports filtering of families, individuals, chromosomes, chromosome ranges, markers and traits, so a desired subset of the data can easily be extracted. Our age-related maculopathy exome chip data has 247,519 SNPs stored in blocks on 1,058 people. It took 470 seconds (s) and 3.22 GB of memory on a 3.33 GHz Core 2 Mac to export 25,175 SNPs on chr. 1 to a binary PLINK file. There are 39 SNPs in both the blocked and conventional data: extracting from blocked only took 10 s, conventional 75 s, and extracting from both and resolving conflicts took 100 s. dbVOR is freely available from our http://watson.hgen.pitt.edu/register web site. This work was supported by NIH grants R01 EY009859 (PI: Gorin), R01 GM076667 (PI: Weeks), and ARRA supplement R01 EY009859-S14 (PI: Gorin).

1831F

Mega2: enhanced data-handling for facilitating genetic linkage and association analyses. *C.P. Kollar[1], R.V. Baron[1], N. Mukhopadhyay[2], D.E. Weeks[1,3]*. 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 3) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

It is common in a genetic study of a complex disease to use a variety of different analysis programs. Invariably, this will require reformatting the data into the precise input format required by each of the analysis programs used. Our Mega2 software seeks to facilitate this process by automating common data reformatting tasks, thereby markedly reducing the chance of data reformatting errors, making data setup much less error-prone, and saving the user the time of writing, debugging, and maintaining data conversion scripts. We have recently extended the capabilities of Mega2 in a number of ways. In addition to accepting LINKAGE-format input files, it now accepts PLINK-format files, both text-based and compact binary formats, as input. Mega2 is also more efficient than it was previously, enabling it to handle larger data sets and to perform its conversion operations faster. While continuing to support conversion to commonly used formats like Merlin, Mendel, SimWalk2, and SOLAR, Mega2 has now been extended to support data conversion to more analysis formats, including PLINK, Cranefoot, IQLS, FBAT, MORGAN, BEAGLE, Eigenstrat, and Structure. For some output options, Mega2 supports the generation of high-quality plots of the results using our nplplot R package, as well as generation of custom track files for use within the UCSC genome browser. It has also been extended to support organisms other than humans. Internally, Mega2 is now written in C++, and the use of object-oriented programming techniques has greatly simplified the process of adding support for new analysis programs. Mega2 is open source and is freely available, along with extensive documentation and a 'quick-start' tutorial, from our http://watson.hgen.pitt.edu/register web site. This work was supported by NIH grant R01 GM076667 (PI: Weeks). Earlier contributions to our initial code base were made by Lee Almasy, Mark Schroeder, and William P. Mulvihill.

1832W

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging:: Machine learning techniques for quality control for 100,000 subjects genotyped on the Affymetrix® Axiom® system. *M. Kvale[1], S. Hesselson[1], T. Hoffmann[1], D.K. Ranatunga[2], J. Gollub[3], T. Webster[3], Y. Zhan[3], Y. Lu[3], G. Mei[3], L. Walter[2], D. Ludwig[2], B. Dispensa[1], C. Schaefer[2], P.Y. Kwok[1], N. Risch[1].* 1) Inst Human Genetics, Univ California, San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA; 3) Affymetrix Inc., Santa Clara, CA.

The Research Program on Genes, Environment and Health (RPGEH) has assayed genotypes on at least 670,000+ SNPs for each of 100,000 subjects to deliver a clean, consistent dataset for use for use in research on genetic influences on a broad variety of health conditions. The number of subjects created challenges for both calling genotypes and for ensuring the highest possible quality across the whole dataset. In this presentation, we describe novel applications of machine learning to better assess sample and SNP assay quality. To determine SNP quality in the Axiom® genotyping assay, it is necessary to go beyond the confidence measures provided by the BRLMM-P genotyping engine. The gold standard for assessing SNP quality is manual inspection of fluorescent intensity profiles and genotype statistics of the population. The 100,000 samples were genotyped in 58 separate sets of samples that lead to more than 42 million intensity profiles and associated genotype inferences; manual assessment of these would take an impractically long time. To leverage the utility of manual assessment over all the genotypings, we created a support vector machine (SVM) classifier based on geometric and statistical measures of the intensity profiles, confidence distributions, and genotype statistics and trained it using manually graded SNPs. Optimization of the SVM hyperparameters was performed using cross-validation and an adaptive multigrid search. The resulting SVM classified SNP quality at a level comparable with manual grading and showed typical classification accuracy of 98-99%, with disagreements happening mainly in the marginal quality cases. A significant source of poor genotype calls came from SNP locations in which double deletions (a copy number of zero) produced a null cluster not accounted for by BRLMM-P. The SVM classifier was adapted to detect the presence of null clusters and allow for recovery of these copy number variants. An SVM classifier was also created to aid in gender identification of assayed samples. Gender inference from the assay was invaluable in detecting sample and plate mix-ups. Because of wide variance in X and Y intensities and potential large chromosomal deletions, a simple threshold test is not sufficient. An SVM was trained against survey gender to look for problem misclassifications. Of more than 100,000 samples, several hundred misclassifications were detected and removed from the dataset.

1833T

FBAP: A pipeline for family-based quality control of pedigree structures and dense genetic marker data. *A.Q. Nato[1], N.H. Chapman[1], C.Y.K. Cheung[2], Z. Brkanac[3], E.M. Wijsman[1,2,4].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 4) Department of Genome Sciences, University of Washington, Seattle, WA.

Inheritance vectors (IVs) specify the flow of founder alleles of each non-founder in a pedigree making them useful as a foundation for statistical genetic analyses, such as linkage analysis, association scans, and next-generation sequencing studies. Stringent quality control (QC) procedures for pedigrees ranging from small up to very large pedigrees and for genotype data of increasing resolution (multiallelic linkage marker panels, dense diallelic SNP marker panels, and high-throughput sequence data) are essential for accurate family-based analyses. We develop a family-based analysis pipeline (FBAP) that implements QC checks on genetic data of diverse breadth and depth, samples IVs based on validated pedigree structures and cleaned genotype data, and formats data for family-based analyses such as linkage analysis, association analysis, or imputation from sequence data. FBAP first identifies relationship/pedigree errors through the comparison of pairwise kinship coefficients based on pedigree structure vs. empirical kinship coefficients estimated using genotype data. Marker genetic locations (Kosambi) and physical positions are taken from the Rutgers Map and subsequently converted to their corresponding positions based on the Haldane function. An initial marker subpanel is selected from the dense panels by screening SNPs through LD-based SNP pruning, and by setting a minimum intermarker distance of 0.5 cM. Additional subpanels are obtained using different starting markers. The IVs for each initial subpanel are sampled by gl_auto of the MORGAN package conditional on validated pedigree structures, marker map, allele frequencies, and genotype data. Using each initial subpanel, Mendelian-consistent errors of remaining markers that are not on this subpanel are detected by GIGI, which flags markers with possible errors. This process is performed on each subpanel to ultimately generate a final framework panel (composed of cleaned markers) and its corresponding sampled IVs. FBAP therefore allows researchers to select markers without Mendelian consistent or inconsistent errors, sample IVs based on validated data, and subsequently perform different types of family-based analyses.

1834F

Comparing a few SNP calling algorithms using low-coverage sequencing data. *X. Yu[1], S. Sun[1,2].* 1) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Cancer Center, Case Western Reserve University, Cleveland, OH.

Several single nucleotide polymorphism (SNP) calling programs have been developed to identify novel SNPs and mutations in next generation sequencing (NGS) data. However, low sequencing coverage presents challenges to accurate SNP calling. Moreover, commonly used SNP callers usually include several metrics for each potential SNP in their output files. These metrics are highly correlated in complex patterns, making it extremely difficult to select SNPs to do any further experimental validation. To compare the performance of SNP callers in a low coverage sequencing dataset, we first compare the SNP calling results generated from four algorithms, SOAPsnp, Atlas-SNP2, samtools, and GATK, without any post-output filtering. We have a few findings. First, we find that SOAPsnp calls more SNPs than other algorithms since it has little internal filtering criteria. However, Atlas-SNP2 reports the least number of SNPs since it has stringent internal filtering criteria. Second, using several cutoff values for the sequencing coverage of called SNPs, we find that filtering the SNPs with a higher coverage threshold improves the agreement among the four algorithms. Third, we explore the values of a few key metrics in each algorithm, and use them as post-output filtering criteria to maximize the agreement of SNP findings among algorithms. Our exploratory results show that high coverage regions or bases tend to have high calling qualities. We recommend the users to employ more than one SNP calling algorithm, and use coverage and calling quality as filtering criteria for reliable SNP identification.

1835W

Influence of low level contamination on variant calling and filtration of NGS data and its quality control. *H. Ling, K. Hetrick, E. Pugh, J. Romm, K. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. In NGS labs, cross sample DNA contamination can be one of the major causes of genotyping errors and false positives. Contamination may be present in source DNA, introduced during library preparation or cross-talk during PCR amplification. To better understand its influence on data quality and QC data, we investigated 1) at what level variant calls start to be affected by contamination, 2) whether variant calling can be corrected by down-sampling alternate alleles to a fixed value or by providing an estimated contamination level, and 3) the upper limit of contamination at which the correction will still work properly to eliminate the influence. We chose a set of high quality samples and manually constructed contamination by proportionally merging BAM files from subjects with varying levels and types of combination (related/unrelated, ethnicities, 2-/3-way) at pre-determined level. Both contaminated and non-contaminated samples were called together by Unified Genotyper. A number of QC metrics (reproducibility, concordance and sensitivity to array data) were employed to evaluate the presence of contamination and quality of variant calls. VerifyBamID was used to estimate contamination levels to see how accurate the estimation is and the performance of down-sampling in correcting variant calls by feeding the estimated contamination level to caller. The contamination level was well estimated by VerifyBamID despite some variation across different types of contamination. Without any effort of correction, variant calls start to be affected when contamination reaches 3%. Down-sampling was effective in correcting moderate contamination when an accurate estimate of contamination level was provided during variant calling. However, when contamination level reached 10% or higher, down-sampling to remove contamination did not work as well and significant numbers of errors remained in the variant calls. Application and integration of VerifyBamID with variant calling allows us to detect contamination at early stages and significantly improves calling accuracy.

1836T

Handling high-dimensional genetics data using multilevel dimensionality reduction algorithms in genetic association studies. *K. Cho[1, 2], D.R. Gagnon[1, 3], H. Wu[1, 4].* 1) Massachusetts Veterans Epidemiology Research and Information Center; VA Boston Healthcare System, Boston, MA; 2) Division of Aging, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA; 3) Boston University School of Public Health, Boston, MA; 4) Computer Science and Networking, Wentworth Institute of Technology; Boston, MA.

In recent years with advancing genotyping and sequencing technologies, there are more genetic data available than one can thoroughly analyze with existing tools and methods. This has become one of the analytical challenges in the initial phases of big scale genetic studies. In whole-genome or genome-wide exploration of novel associations, without a priori knowledge of specific regions or genes of interest, researchers rely on the available techniques and methods that perform association tests based on a way of summarization or assuming independence. In addition, traditional statistical procedures present eminent challenges in using these data, where the number of parameters p is scalably larger than number of observations n. These limitations may compromise the level of false positive results and power to conduct studies with acceptable level of confidence. One of the ways to reduce the impact of these limitations is to first perform a data dimension reduction process and then evaluate the resulting panels of markers. We propose several algorithms using traditional PCA, PCA incorporating heritability, LASSO, and combinations of these to evaluate the impact of type one error. Using the Genetic Analysis Workshop 18 simulated data, we apply the proposed algorithms on selected null regions of chromosome 3 with regards to the diastolic blood pressure phenotype. Then we use SKAT as the evaluation tool to perform association analysis using the resulting panels from each algorithm. We then systematically compare type I error rates among different algorithms with respect to the baseline panel. Our preliminary work using a subset of 2000 SNPs (1723 polymorphic SNPs) in unrelated samples, shows a substantial reduction of the markers through applying proposed algorithms. Among these, the PCA-only algorithm resulted in 47 components comprised of 363 unique SNPs and PCA followed by LASSO algorithm resulted in a set of 135 SNPs. Using the LASSO-only and LASSO followed by PCA, there were 110 and 109 SNPs (32 components) left in the resulting panels, respectively. We are also investigating PCA incorporating heritability approaches using the family datasets. As the amount and dimensionality of the network of genomics data is only escalating in the coming years, a timely, robust and practical analytical pipeline through multilevel dimensionality reduction techniques provides an efficient approach for the initial screening tool.

1837F

Genotype imputation in the era of next-generation sequencing. *G. Pistis[1,2,3], E. Porcu[1,2,3], C. Sidore[3], F. Danjou[3], M. Steri[3], A. Mulas[3], M. Zoledziewska[3], F. Busonero[3], F. Reinier[4], R. Atzeni[4], M. Lobina[3], R. Pilu[4], M. Marcelli[4], B. Tarrier[5], H.M. Kang[1], A. Angius[3,4], C.M. Jones[4], D. Schlessinger[6], F. Cucca[2,3], G. Abecasis[1], S. Sanna[3].* 1) University of Michigan, School of Public Health, Ann Arbor, MI, 48109; 2) Università degli Studi di Sassari, Dipartimento Scienze Biomediche, Sassari, 07100, Italy; 3) Istituto di Ricerca Genetica e Biomedica-CNR, Monserrato (CA), 09042, Italy; 4) Center for Advanced Studies, Research and Development in Sardinia - CRS4, Pula, Italy; 5) DNA Sequencing Core, University of Michigan, Ann Arbor, MI, USA; 6) Laboratory of Genetics, NIA, Baltimore, MD, USA.

Genotype imputation is an essential tool to infer missing and untyped genotypes in genetic studies. Nevertheless, the performance of genotype imputation on rare (MAF <0.5%) and low frequency variants (MAF <1% and >0.5%) is still an unexplored field, and several important issues must be specifically addressed: choice of reference panel, quality of input genotypes/haplotypes and imputation quality for less common variants. Recently, we completed whole genome sequencing of 2,120 Sardinians using a low-pass approach (average 4x coverage) as well as genotyping of 6,600 individuals enrolled in the SardiNIA project, using four different Illumina Beadchip arrays: OmniExpress, Cardio-MetaboChip, ImmunoChip and ExomeChip. We used combinations of the first three arrays to perform genotype imputation using Sardinian sequencing data (SardSeq) as population-specific reference panel, and used the fourth array to evaluate accuracy as squared Pearson correlation R2 between dosages and the real genotypes. We also performed imputation using two 1000 Genomes Project (1000G) reference panels (ALL and EUR datasets). We observed that the use of SardSeq boosted imputation accuracy compared to the other external reference panels generated from geographically distant populations, with differences highly dependent on the particular genotype array used. For example Cardio-MetaboChip provided a more accurate imputation when using the SardSeq panel (mean R2 0.7) compared to 1000G panels (mean R2 0.3). The benefit was even more striking for low frequency and rare variants when imputation was performed based on the OmniExpress array: the mean R2 were, for the first type of variants, 0.91, 0.57 and 0.52 when using SardSeq, 1000G-ALL and 1000G-EUR reference panels, and 0.69, 0.40 and 0.35 for the second. Furthermore, we found that the imputation accuracy using the SardSeq panel on Europeans was on average similar to that reached using 1000G panels on the Sardinian population, for any combination of the same arrays, in both real and simulated datasets. Interestingly, a combined panel of SardSeq and 1000G haplotypes was more beneficial in Europeans than in Sardinians. Our results suggest that using a population-specific reference panel rather than public references (especially for different ancestral or geographically distant populations) yielded better imputation quality and, particularly for less common imputed variants, might reduce false positive signals in association analyses.

1838W

Impact of Quality Control on the Heritability Analyses for Qualitative Traits. *J. Liu[1,3,4], T. Hoffmann[1,3], E. Jorgenson[5], J. Witte[1,2,3,4].* 1) Department of Epidemiology & Biostatistics; 2) Department of Urology; 3) Institute for Human Genetics; 4) Diller Family Cancer Center, University of California, San Francisco; 5) Kaiser Permanente Northern California Division of Research, Oakland, CA.

While a single SNP accounts for only a small fraction of the genetic variation in complex traits, recent work has shown that a substantial heritability of complex quantitative traits can be explained by considering all SNPs simultaneously from genome-wide association studies. This approach has been extended for evaluating the chip heritability of qualitative disease traits. However, analyzing disease traits requires more rigorous quality control than for quantitative heritability~or for a conventional GWAS~because slight case-control differences can be magnified into false evidence of heritability. In this study, we investigate various factors that may affect the assessment of disease heritability, and give a detailed quality control framework for accurately estimating heritability. Using two GWAS of prostate cancer, we show that leaving out key steps can vastly over- or under-estimate heritability. In a GWAS of European Americans, the naïve prostate cancer heritability is an unreasonable 68%, whereas the 'cleaned' heritability is 59%. In a GWAS of African Americans, the naïve heritability is 22%, and the cleaned heritability is 32%. The key factors that lead to over and underestimates are (1) SNPs fail Hardy-Weinberg Equilibrium test and (2) individuals with genome-wide similarities greater than 0.025. Both of the 'cleaned' estimates are in line with what one would expect from previous heritability estimates for prostate cancer. These findings emphasize the need to use more stringent quality control criteria for evaluating chip-heritability in disease traits than one might expect based on heritability for quantitative traits or GWAS.

1839T

Visualization software for the efficient review of alternate genotype calls. *Z. Xu, N. Pankratz.* University of Minnesota, Minneapolis, MN.

The Illumina HumanExome BeadChip (the 'exome chip') has unique challenges over traditional GWAS arrays, as rare variants are often not clustered well using the default GenomeStudio GenCall/GenCall2 algorithms. New algorithms have been developed to address this, such as zCall and optiCall, however no known software can visually compare genotype calls from different algorithms overlaid on raw intensity data in an integrated and efficient manner. We have developed platform-independent Java software to plot a marker's probe intensities for all individuals and to color-code any user-defined variable (e.g., gender, study, DNA source, genotyping batch, in addition to the genotype itself) thus allowing investigators to look for batch effects, detect if a SNP association is due to artifact, or to see if rare variants are being called accurately. We have also implemented an annotation system whereby the user can tag a marker (e.g., 'unusable', 'copy number variant', 'pseudo-autosomal chrY marker', 'extra heterozygote clusters', 'okay') using a single key stroke or mouse click which can also auto-advance to the next questionable marker. Lists of questionable markers can be generated automatically using over a dozen criteria, including low call rate, excess/low heterozygosity, mean theta values deviating from expected, etc. In addition, alternate genotype calls can be imported and customizable colors/symbols can be used to denote points that are discordant for the various calling algorithms (e.g., samples called as missing by GenCall and as heterozygotes by zCall could be orange triangles). The user can then use the software to draw a square around these points and set them to what the user believes to be the proper genotype call. Once this is done for a series of markers, the reclustered data can be exported into the user's preferred format (e.g., PLINK). The software is highly responsive in loading (<1s) and advancing markers (instantly), and when applied to exome chip data for 7676 individuals from the Minnesota Twin Family Study, all 24,898 flagged markers could be reviewed and recalled by one technician over the course of four days. The ability to plot disparate genotype calls and to recluster them is not known to be available in any existing software, commercial or otherwise. This software is publicly available as free and open source and will allow researchers to triage, annotate, and recall markers from any array as quickly and as efficiently as possible.

1840F

PhenoMan: Phenotypic data exploration, selection, management and quality control for association studies of rare and common variants. *B. Li, G. Wang, S.M. Leal.* Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Next generation sequencing and other high-throughput technology advances have promoted great interest in detecting associations between complex traits and genetic variants. Phenotype selection, quality control (QC) and control of confounders are crucial and can have a great impact on the ability to detect associations. Although there are available programs to perform association analyses, e.g. PLINK, they cannot be used for comprehensive management and QC of phenotype data. To address this need PhenoMan was developed to 1.) select individuals based upon multiple phenotype criteria, 2.) replace missing covariate data, 3.) remove duplicates samples and individuals who are related, belong to another population or have incorrect sex specification, 4.) recode primary traits and covariates, 5.) transform and visualize quantitative traits, 6.) remove or windsorize outliers, 7.) select covariates for analysis, 8.) create residuals and 9.) provide summary statistics. A report is generated for each data set in order that the same protocol can be used to ensure consistency and harmonization between analyses: PhenoMan is a user friendly interactive program that integrates data exploration, management and QC using a unified platform. Proper QC of phenotypes before proceeding to association analyses is critical to ensure control of type I and II errors, reliable effect estimates and consistent results between studies. PhenoMan is highly beneficial for preparation of case, control and quantitative trait data for association studies using new data sets as well as those obtained from public repositories. The PhenoMan program and documentation are available at https://code.google.com/phenoman.

1841W

Forensic Inference via a Genome Parade. *Y. Chen[1,2], R. Xia[1,2], F.A. San Lucas[4,2], S. Vattathil[3,2], P. Scheet[1,2,3,4].* 1) Division of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, TX; 2) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX; 3) Program in Human and Molecular Genetics, The University of Texas at Houston Graduate School of Biomedical Sciences; 4) Program in Biomathematics and Biostatistics, The University of Texas at Houston Graduate School of Biomedical Sciences.

Contamination of DNA may present problems or clues in studies using high-throughput next-generation sequencing data. Methods such as ContEst (Cibulskis et al, 2011, Bioinf., 27:2601) and ContaminationDetection (Jun et al, 2012, AJHG, 91:839) have been shown to provide accurate estimates of contamination levels using complementary microarray data to look for deviations from expected allele-specific read counts. Here we present a statistical model that incorporates the dependence of these population allele frequencies, or linkage disequilibrium (LD), to improve component identification when encountering mixtures of two diploid genomes in next-generation sequencing data. Our method, which we call Genome Parade, uses a hidden Markov model similar to fastPHASE (Scheet & Stephens, 2006, AJHG, 78:629) to accommodate dependence of alleles potentially from a contaminator's DNA, essentially presenting pairs of haplotypes 'in turn' from a model for genetic variation in front of the observed data for comparison (analogous to an 'identity parade'). Using simulated and real data, we explore certain forensic applications, as existing methods offer limited power to identify the genotypes of a contaminating individual. For example, compared with an approach that ignores LD, our method increases the squared correlation between estimated genotypes and true genotypes of the contaminating individual from .57 to .82 in a scenario of 2% contamination and sequencing depth of 100. This is essentially an application of low-coverage LD-based imputation in the unusually difficult setting where we are interested in the rare source of DNA in a mixture. This may be useful when attempting to identify a contaminating individual in a subtle mixture of DNA, when plausible candidates are not readily available, or to aid in reducing a search space for brute-force methods that use large databases. In a related application, Genome Parade may be useful when indeed there exists a suspected source of contamination but when the alternatives do not come from an identifiable list. In such cases, the probability of the data may be easily calculated for the suspected source, but to quantify this evidence a comparison must be made to haplotypes drawn from the general population. Finally, incorporation of LD may aid in quantifying contamination in applications of targeted sequencing, since ignoring the dependence in the data indicates a false sense of precision when data are limited.

1842T

Discovery and replication of genetic interactions for quantitative lipid traits. *E.R. Holzinger[1], M. Farrall[5], F. Drenos[7], C.B. Moore[3], I.B.C. Lipid Working Group[8], S. Setia[4], H. Watkins[5], F.W. Asselbergs[6], B.J. Keating[2], M.D. Ritchie[4].* 1) National Institutes of Health, National Human Genome Research Institute, Inherited Disease Research Branch, Baltimore, MD; 2) Center for Applied Genomics, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA; 3) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 4) Center for Systems Genomics, The Pennsylvania State University, University Park, PA 16870; 5) The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 6) Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands; 7) Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK; 8) IBC Lipid Working Group.

The genetic etiology of human lipid quantitative traits may be further elucidated by considering interactions between variants. In this study, we perform a genome-wide interaction study (GWIS) for four different lipid traits - low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (TC), and triglycerides (TG). Our analysis consisted of a discovery phase using individuals from five different cohorts (ARIC, CARIDA, CHS, FHS, and MESA; n~14,000) and a replication phase in a subset of the PROCARDIS cohort (n~6400). Currently, there is no gold standard for performing a powerful GWIS. Due to the computational and multiple testing burden of exhaustively testing all pairwise interactions, filters are often applied before interaction testing. In this analysis, we used two different filters: 1. Main effects filter (MEF) - exhaustively test for pairwise interactions between SNPs that pass a main effect threshold of $p<0.001$. 2. Biological filter (BF)- SNP-SNP models generated with prior biological evidence supporting a potential interaction using Biofilter 2.0. We tested interactions using a linear regression model with a multiplicative interaction term for the two model SNPs. We selected SNP pairs from the discovery analysis with interaction $p<0.001$ and identified all SNP-SNP models that represent the same signal based on LD. We then tested these models in the replication dataset. We corrected for the number of interaction signals tested in the replication phase using a Bonferroni correction (MEF: $p < 0.0003$; BF: $p<0.002$). In the MEF analysis, 15 interaction signals replicated for HDL-C, 4 for LDL-C, and 12 for TG. No models replicated for TC. For the BF analysis, the TG trait analysis resulted in 4 interaction signals that replicated. No interactions replicated for the other lipid traits. Interestingly, the same SNP-SNP interaction replicated in both the MEF and BF analyses for TG (rs1263173 and rs12225230; $p=4.8\times10^{-5}$). Currently, more replication efforts are being conducted in other independent datasets to further assess these results. In this study, we identify potential interaction signals that may be contributing to the genetic etiology of various lipid traits. We also developed a pipeline for performing a computationally efficient interaction analysis. Future work will investigate the molecular mechanisms that may underlie statistically validated interactions.

1843F

Association analysis of gene-environment interactions in lipid profile using exome sequence data. *Z. He, G. Wang, S. Leal.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Virtually all common diseases are a result of the complex interplay of genetic and environmental factors. Among the environmental factors, cigarette smoking is a well-established environmental risk factor contributing to multiple complex phenotypes, for example lipid profile which includes plasma high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), and triglycerides levels. However, little is understood about the potential interactions between genetic variations and smoking that are associated with lipid profile, mainly because the lower statistical power to detect an interaction than detecting a main effect. A number of methods have been developed to enhance the power of detecting gene-environment interaction. Among these methods, set-based approaches along with correlation-based screening have been shown to have superior performances, by aggregating signals within a set and reducing the multiple-testing burden in genome-wide gene-environment interaction studies. Set-based methods include a number of recently proposed rare variant association methods, for example the burden tests and variance component tests. In this study the relative performance of each of these methods was evaluated using simulated data under a wide range of scenarios. We also applied these approaches to analyze potential gene-environmental interactions that are associated with lipid profile, using sequencing data from NHLBI-ESP (Exome Sequence Project). Additional information about the interactions between gene and cigarette smoking that are associated lipid profile may yield secure insights into its role in cardiovascular disease.

1844W

Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4,965 mice. *E. Kang[1], B. Han[2,3,4], N. Furlotte[1], J. Joo[5], D. Shih[6], R. Davis[6], A. Lusis[6,7], E. Eskin[1,7].* 1) Department of Computer Science, University of California, Los Angeles, CA, USA; 2) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA; 3) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Interdepartmental Program in Bioinformatics, University of California Los Angeles, CA, USA; 6) Department of Medicine, University of California, Los Angeles, CA, USA; 7) Department of Human Genetics, University of California, Los Angeles, CA, USA.

Identifying environmentally specific genetic effects is a key challenge in understanding the structure of complex traits. Model organisms play a crucial role in the identification of gene-by-environment interactions, as a result of the unique ability to observe genetically similar individuals across multiple distinct environments. A large number of model organism studies with varying environmental conditions measure the same traits. For example, knockout or diet-controlled studies are often used to examine plasma cholesterol levels in mice. These studies when examined in aggregate provide an opportunity to identify genomic loci bearing environment-dependent effects. However, straightforward application of traditional methodologies to aggregate datasets suffers from several problems. First, environmental conditions are often variable and do not fit the standard univariate model for interactions. Additionally, applying a multivariate model results in increased degrees of freedom and low statistical power. In this paper, we propose a meta-analytic approach based on a random effects model to identify loci involved in gene-by-environment interactions by jointly analyzing multiple studies with varying environmental conditions. Our approach is motivated by the observation that methods for discovering gene-by-environment interactions are closely related to random effect models for meta-analysis. We show that the interactions can be interpreted as heterogeneity and can be detected without utilizing the traditional uni- or multi-variate models for discovery of gene-by-environment interactions. We apply our method to combine 17 mouse-studies containing a total of 4,965 animals. Using these studies, we identify 26 significant loci involved in plasma HDL cholesterol levels. Among these, 23 loci are previously confirmed to have an effect on HDL cholesterol or closely related lipid levels in the blood, while 3 loci are novel. Several of these loci show significant evidence of gene-by-environment interactions including a locus on chromosome 8 showing strong evidence of sex by knockout-driven LDL-level interaction, which simultaneously affects HDL cholesterol levels. A great advantage of our approach is that due to a various design strategy of combined study, it provides significantly higher power and improved resolution for identification of the underlying genes.

1845T

GMDR: A conceptual framework for detection of multifactor interactions underlying complex traits. *X. Lou.* Dept Biostatistics, Univ Alabama Birmingham, Birmingham, AL.

Biological outcomes are governed by interacting networks consisting of multiple genetic and environmental factors that jointly act in often unpredictable ways. Determining multifactor interactions underlying complex traits is the primary topic of interest in recent genetics studies but presents enormous statistical and mathematical challenges. The computationally efficient multifactor dimensionality reduction (MDR) approach, originally for a case-control study, has recently emerged as a promising tool for meeting these challenges. On the other hand, complex traits are expressed in various forms such as categorical, ordinal, continuous, count and time-to-event and have different data generation mechanisms that cannot be appropriately modeled by a dichotomous model; the subjects in a study may be recruited according to its own analytical goals, research strategies and resources available, not only unrelated individuals from a homogeneous population. Although several modifications and extensions of MDR have in part addressed the practical problems, they are still limited in statistical analyses of diverse phenotypes, multivariate phenotypes and correlated observations, correcting for potential population stratification, and unifying both unrelated and family samples into a more powerful analysis. I propose here a comprehensive conceptual framework, referred as to generalized MDR (GMDR), for a systematic extension of MDR. The proposed approach is quite versatile, not only allowing for covariate adjustment, being suitable for the analysis of almost any trait type, e.g., binary, count, continuous, polytomous, ordinal, time-to-onset, multivariate and others, as well as combinations of those, but also being applicable to various study designs including homogeneous and/or admixed unrelated-subject and family as well as mixtures of them. The proposed GMDR offers an important addition to the arsenal of analytical tools for identifying nonlinear multifactor interactions and unraveling the genetic architecture of complex biological traits. (This study is being supported by NIH grant DA025095.).

1846F

Novel statistical framework for gene-environment interaction. *A.R. Stefanescu, X. Wen.* Biostatistics, University of Michigan, Ann Arbor, MI.

Investigations by Voorman, et al. (2011), Zuk, et al. (2011), and Wen and Stephens (2011) have shown that model misspecification in the context of gene-environment interaction can lead to an array of serious analytical issues, including missing heritability and inflated type I error. Firstly, we demonstrate through simulation the detrimental effects of ignoring gene-environment interaction in cases when interaction is known to be present. Secondly, we address the weaknesses of the naïve model that ignores interaction by comparing it to a traditional interaction model and also proposing a more robust subgroup analysis framework. Finally, we intend to explore the potential applications of the new method by integrating tissue-specific eQTLs into GWAS analyses of complex traits.

Our preliminary results showed that when error variance was held constant and genetic effect size was varied by environmental category in simulated data sets, ignoring interaction explained little of the observed variance, while the interaction models explained a much greater proportion of observed variance as heterogeneity of genetic effect increased. Of the two interaction models, the subgroup analysis performed the best, showing consistently higher estimated heritability values than the traditional interaction model, regardless of levels of heterogeneity. Furthermore, when we allowed phenotype variance to differ by environmental category (a common phenomenon observed at the cellular level), the interaction models also performed substantially better than the model ignoring interaction. Of the two interaction models, the subgroup analysis performed the best. This result comes as no surprise since the traditional interaction model assumes constant error variance across environmental conditions.

The novel subgroup analysis structure is immediately applicable to studies in which the environment variable is discrete, such as in different tissue types (NIH GTEx Project, for example) and in threshold exposures. We further intend to generalize this model to the continuous environment case. The generalized model would be applicable to studies that use biomarkers such as hormone levels as environment variables.

1847W

Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. *J. Sul[1], W. Yang[1], E. Kostem[1], N. Furlotte[1], D. He[1], E. Eskin[1,2].* 1) Computer Sci, Univ California, Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, California, USA.

Genome-wide association studies (GWASs) attempt to identify genetic variants associated with complex traits by collecting a set of unrelated individuals. If individuals are related or from different populations, a phenomenon called population structure may cause spurious associations. GWASs on model organisms such as inbred mouse strains that are highly diverse in genetic background are readily susceptible to population structure. Several methods have been proposed to address this problem, and although GWASs have identified numerous variants associated with many traits using the methods, those variants explain only a small fraction of trait heritability.

Among several contributions that may account for a significant fraction of heritability, one is gene-by-environment interactions (GEIs). Discovering GEIs can provide insight into disease pathways, an understanding of the effect of environmental factors in disease, better risk prediction and personalized therapies. GEIs are often discovered in model organisms because of possibility to manipulate different environments.

In this paper, we first show analytically that for the same reasons that population structure causes spurious associations of genetic variants, it also causes spurious GEIs. We then use both simulation and inbred mouse strains termed Hybrid Mouse Diversity Panel (HMDP) to observe spurious GEIs. HMDP consists of 100 classical inbred and recombinant inbred (RI) strains, and they are genotyped through high density SNP typing. We analyze their lipid phenotypes, and the environment is a thioglycollate injection to recruit macrophages. We observe that test statistics for GEIs are inflated across most phenotypes, leading to spurious GEIs. Lastly, we propose a method to account for population structure for GEIs using mixed models. It has been shown that mixed models effectively control population structure on effects of genetic variants. We show that directly applying mixed models to GEIs, however, does not control population structure on GEIs because current mixed models are designed to control inflation of genetic effects. We extend mixed models to control population structure on both genetic and GEIs effects and show that our method removes the inflation of test statistics on simulation and the HDMP dataset.

1848T

Improved detection of variants with main or interaction effects using a robust location-scale testing framework. *D. Soave[1,2], A.D. Paterson[2,1], L.J. Strug[2,1], L. Sun[1,3].* 1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 2) Research Institute, The Hospital for Sick Children, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Toronto, Canada.

The most common approach in genetic association studies is to test for phenotypic mean (location) differences between genotypes. Complex genetic etiologies such as GxG and GxE interactions can result in phenotypic variance (scale) differences between genotypes for a SNP of interest. Formally modeling GxG and GxE effects can detect these interactions, but may not be practical due to issues such as multiple hypothesis testing or missing information concerning interacting exposures. As an alternative, Levene's test for equality of variance has been proposed (Pare et al., 2010). Not surprisingly, when the distribution of a quantitative trait per genotype differs mostly in variance, this scale-test has better power, but it is not robust for detection of differences in location. More recently, Aschard et. al. (2013) proposed a distributional test that compares the percentiles of phenotypic values between genotypes. It has the advantage of detecting either mean or variance differences, or both. While this approach comprehensively evaluates the phenotypic distribution between genotypes, the sample size required to differentiate distributions is much larger than for detecting mean or variance differences. Furthermore, the full information contained in a (approximate) normally distributed trait is well captured by the mean and variance. We propose a joint location-scale testing framework. We consider a direct likelihood approach that tests the null hypothesis of equal mean and equal variance between genotypes. We compare findings with alternative Fisher and minimum p-value methods that combine evidence from the complementary individual location and scale tests. The combined methods are more robust to model assumptions, and they allow for flexibility with the specific individual tests chosen. Extensive simulation studies confirm that the proposed location-scale testing framework is powerful for detection of differences in location or scale or both. Furthermore, in most of the simulation scenarios considered by Aschard et. al. (2013), the proposed method is equally or more powerful than the distributional test. Application to a candidate gene modifier study of cystic fibrosis and a genome-wide association study of type 1 diabetes complications show that the new testing framework preserves the priority of previously identified top ranked variants while pointing to new candidates.

1849F

Improved detection of genetic exposures with unspecified effect modifiers. *T.L. Edwards[1], C. Li[2].* 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Biostatistics, Vanderbilt University, Nashville, TN.

Complex phenotypes often result from interplay of multiple genetic and environmental factors. Association analyses can gain power by modeling interaction effects when they exist. However, existing methods require explicit specification of effect modifiers, and exhaustive pairwise scans of SNPs introduce well known computational, statistical, and logistical challenges. Knowledge of the loci that are contextually related to traits through effect modifiers may be crucial for understanding genetic relationships with factors such as lifestyle, race, and drug treatment effects. For continuous phenotypes, we propose a single-locus association method that accounts for interaction through modeling both marginal mean and variance as linear functions of genotype. We base this approach on the observation that when effect modification occurs, the difference between conditional means of the dependent variable at each level of the effect modifier varies by genotype, and that this influences the conditional variance for each genotype. We derive marginal mean and variance as functions of genotype and show that the marginal variance is a quadratic function of both genotype and strength of effect modification under commonly used interaction models. Using simulations we compare our method with a test of marginal genotypic effect on the mean with constant variance and with a method recently proposed by Aschard et al. (Genetic Epidemiology, 2013). The results show our method controls type I error rate and significantly improves power over both alternatives. For example, under one scenario, our method had 68% power at the GWAS significance level $p<5e-8$, while linear model had 39% and that of Aschard et al. had 45%; under another scenario, our method had 91% power versus 78% and 83% for the alternatives. This new method will help detect genes that affect phenotypes mostly through interactions with effect modifiers and may help explain some missing heritability for many complex phenotypes. This method could be applied to standing collections of genetic data to discover novel associations without the need for collecting and scanning candidate environmental effect modifiers, or the difficulty of coordinating pairwise SNP scans for consortium studies.

1850W

Genome Wide Interaction Study of Dengue Shock Syndrome.
L. GRANGE[1,2], J.F. BUREAU[1], S. SAKUNTABHAI[1,3]. 1) Functional Genetics of Infectious Diseases, Institut PASTEUR, PARIS, FRANCE; 2) Université PARIS Diderot, Paris VII, FRANCE; 3) Department of Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Although Genome-Wide Association Studies (GWAS) have been successful in identifying more than 500 single nucleotide polymorphisms (SNPs) associated with a broad range of diseases, these SNPs explain only a part of the genetic component of these diseases. Up to now, these GWAS have proceeded by testing one SNP at a time. Thus, only SNPs with a marginal effect have been detected, while other ones that act primarily through a complex mechanism involving interactions with other genetic variants and environmental factors have yet to be discovered. The identification of these interactions is very challenging and powerful statistical methods for conducting genome-wide interactions studies (GWIS) are needed. We compared power and false discovery rate among existing programs for detecting genetic interactions using simulation datasets with different genetic interaction models. Because of their respective low false discovery rate and high power, we chose to use Plink to perform exhaustive epistasis search on real data and Model-Based Multifactor Dimensionality Reduction (MBMDR) to validate the best signals obtained. Using the previously reported GWAS dataset of dengue shock syndrome (Khor ***et al.***, *Nature Genetics*, 2011), we performed an exhaustive epistasis search, using Plink's fast-epistasis mode, on 481 344 quality controlled SNPs and identified 2 couples of regions that contributed repeatedly to the best hits of this study. 3 SNPs located in the same region on chromosome 1 (between positions 77,814,710 and 78,041,795 - Reference Genome Build 36.3) gave strong interaction signals (*P* values < 10E-10) with 3 closely located SNPs on chromosome 18 (between positions 55,425,493 and 55,461,588). Similarly, 2 SNPs from the same region on chromosome 2 (between positions 124,921,887 and 124,962,562) interacted with 6 SNPs on chromosome 7 (between positions 26,688,535 and 26,863,964) (*P* values < 10E-10).Interestingly enough, these SNPs had no marginal effects (*P*>0.4 in GWAS). We used MBMDR on 200 SNP extracts from these regions to validate the significance of the interactive signals. Pairs from chromosomes 1 and 18 turned out not to be significant with MBMDR but pairs from chromosomes 2 and 7 reached p values of 10E-4 which is the lowest one can expect from the 10,000 permutations we ran. These preliminary results support a pure epistatic effect between *CNTNAP5* gene on chromosome 2 and *SKAP2* gene on chromosome 7. Replication and functional analysis are on the way.

1851T

Gene-Gene Interaction Detection Using a Two-stage Model.
Z. Wang[1], J. Sul[1], S. Snir[2], J. Lozano[3], E. Eskin[1]. 1) Department of Computer Science, Univ California, Los Angeles, Los Angeles, CA; 2) Department of Evolutionary and Environmental Biology, University of Haifa, Israel; 3) Department of Computer Science and Artificial Intelligence, University of the Basque Country, Donostia, Spain.

Genome wide association study (GWAS) has been used extensively to discover genetic variations associated with traits. Many studies have reported associations between a single nucleotide polymorphism (SNP) and traits. However, recent GWAS results show that complex traits are influenced by multiple SNPs and it is likely that these SNPs interact. To detect the interaction between SNPs, we need to perform a pairwise association test between a trait and a pair of SNPs. A brute force approach which calculates the association statistics between a trait and all pairs of SNPs is infeasible because of the large number of SNPs collected in current GWAS studies. We propose a two-stage model to reduce the number of tests needed while maintaining similar power to the brute force approach. In the first stage, our method performs the single marker test on all SNPs and selects a subset of SNPs that achieve a certain significance threshold. In the second stage, we perform a pairwise association test between traits and pairs of the SNPs selected from the first stage. We applied our approach to the Northern Finland Birth Cohort data and achieved 30 times speedup while maintaining 99% of the power of the brute force approach.

1852F

Long range phasing using consistency graphs of identity by descent.
I. Pe'er, H.R. Shokri, P.F. Palamara. Dept Computer Sci, Columbia University, New York, NY.

Long range haplotype phasing of heterozygous genotype calls had been proposed by considering segments of identity by descent to homozygous individuals. We hereby extend this framework theoretically by defining a special graph structure across a genomewide-typed or sequenced cohort. Formally, each segment of identity by descent is defined by a pair of typed individuals i,j and a genomic interval $(l..r)$. We define a graph G_i for each individual whose nodes are the segments $[(i,j),(l..r)]$ incident on individual i. Nodes $[(i,j),(l..r)]$ and $[(i,k),(l'..r')]$ are linked by an edge if their corresponding segments overlap $l'<r$ and $l<r'$. Such edges are weighted according to the consistency of the footprints of these segments along the genomes of individuals j and k, defined as log probability of their genotypes being identical by descent along the intersecting interval $(\max(l,l'),\min(r,r'))$. Maximum-likelihood phasing of i into her two parental haplotypes is equivalent to finding the maximum cut in G_i. This framework allows handling alleles of various frequencies, incorporating sequencing/genotyping errors, while properly weighting interval length and missing data. Computationally, we observe that G_i is an interval graph, for which the Max-Cut problem is fixed-parameter-tractable, i.e. polynomial given a bound on the clique number of G_i, using a dynamic program that progresses along the maximal cliques in G_i along the genome. This implies a polynomial solution for maximum-likelihood phasing of i. Practically, when G_i has large cliques, we observe that such cliques mean highly redundant phasing information, that we therefore omit. We further speedup performance by observing that most segments of identity by descent longer than a length threshold are only slightly longer than their defining cutoff. This means G_i is approximately a unit interval graph, where the Max-Cut problem admits a linear-time solution. To phase the entire cohort, we iterate phasing of each G_i, updating the information regarding identity-by-descent segments after phasing so that $[(i,j),(l..r)]$ is probabilistically assigned to the possible phases of i. When sufficiently many segments are available, these iterations quickly converge to the correct phasing.

1853W

The relationship between common environmental and genetic effects on human gene splicing and expression.
A. Battle[1], D. Knowles[1], S. Mostafavi[1], X. Zhu[2], J.B. Potash[3], M.M. Weissman[4], C. McCormick[5], C.D. Haudenschild[6], K.B. Beckman[7], J. Shi[8], R. Mei[9], A.E. Urban[2], D.F. Levinson[2], D. Koller[1,10], S.B. Montgomery[10,11]. 1) Computer Science, Stanford University, Stanford, CA; 2) Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA; 3) Department of Psychiatry, University of Iowa Hospitals & Clinics, Iowa City, IA; 4) 4 Department of Psychiatry, Columbia University and New York State Psychiatric Institute, New York, NY; 5) Illumina, Inc., La Jolla, CA; 6) Personalis, Inc., Menlo Park, CA; 7) Biomedical Genomics Center, University of Minnesota, Minneapolis, MN; 8) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 9) Centrillion Biosciences, Inc., Palo Alto, CA; 10) Department of Pathology, Stanford University, Stanford, CA; 11) Department of Genetics, Stanford University, Stanford, CA.

Environmental and behavioral factors including substance use have significant impact on public health and individual risk of many common diseases, but the cellular basis for such effects is not fully understood. Through sequencing RNA from primary tissue (whole blood) of 922 living individuals, we explore the transcriptional correlates of common environmental factors. The DGN cohort was interviewed extensively to record medical, behavioral, and environmental variables, offering an opportunity to study their effects on the cell at a large scale. We observe broad transcriptional changes correlated with time of day, substance use, and medication, including changes in pathways relevant to disease risk. Notably, differences are apparent not only in gene expression levels, but also in quantitative measures of alternative splicing. Further, by integrating genetic information into this analysis, we can explore the combined regulatory impact of genotype and environment, including GxE interactions affecting diverse transcriptional traits. In addition to evaluating environment-specific quantitative trait loci (eQTL), RNA-sequencing offers an additional avenue for exploring GxE interactions through a novel method to evaluate differential allele-specific expression (ASE). By employing Bayesian methods to identify the most likely environmental and technical covariates responsible for variation in gene expression, we have identified both differential expression and splicing, along with a number of cases where the effect of a regulatory variant appears to be modulated by specific environmental factors. For example, genes inferred to respond to smoking include pathways related to beta-cell regulation and diabetes. These data and analyses offer a window into the complex response of the cell to behavior and environmental factors that carry significant health consequences.

1854T

Within-gene interactions in GWAS identifies novel susceptibility loci - The WTCCC data revisited. *N. Sharaf Eldin[1], Q. Liu[1], S. Jabbari[1], L. Wang[2], C. Franco-Villalobos[1], S. Mahasirimongkol[3], H. Yanai[3, 4], L.J. Martin[1], K. Tokunaga[3], Y. Yasui[1].* 1) School of Public Health, University of Alberta, Edmonton, Alberta, Canada; 2) College of Agriculture & Biotechnology, Zhejiang University, China; 3) Department of Human Genetics, School of International Health, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 4) Fukujuji Hospital, Japan Anti-Tuberculosis Association, Kiyose, Japan.

Genome-wide association studies (GWAS) examine single nucleotide polymorphisms (SNPs) associated with disease risk. A Standard single-SNP analysis, however, ignores combined effects of multiple SNPs; and SNP-set interactions remain largely unexplored at the genome-wide level. Here we show how exploring interactions of all SNPs within each gene can identify appreciable numbers of novel susceptibility loci. We re-analyzed six diseases in The Wellcome-Trust-Case-Control-Consortium (WTCCC) data: bipolar disorder (BD), coronary artery disease (CAD), hypertension (HT), rheumatoid arthritis (RA), type 2 diabetes (T2D), and type 1 diabetes (T1D). We considered two forms of SNP-set interactions: SNP intersection and SNP union. SNP-set interactions within each gene were assessed using logic regression. The number of genes showing strong evidence of association was: 88 for BD, 81 for CAD, 88 for HT, 161 for RA, 78 for T2D and 192 for T1D. All strong single-SNP signals of WTCCC and around 80% of recent GWAS meta-analyses signals were confirmed. In addition, strong evidence emerged implicating a large number of new discoveries supported by apparent biologically plausible links to disease. Top significant genes were: *CBLN4* with BD, *P2RX4* with CAD, *BBOX1* with HT, *STAG3L4* with RA, and *RHOJ* with T2D. Novel disease susceptibility loci with biologically plausible links to the six diseases were detected in our interaction analysis. This emphasizes the importance of considering higher order SNP-set interactions in addition to the standard single-SNP analysis in GWAS.

1855F

Applications of Hidden Markov Models with Conditional Emission Probabilities to Identify Regions of Identity-By-Descent in Whole-Exome Sequencing Data. *M. Kimmel[1,2], S. Hicks[1], S.E. Plon[3].* 1) Department of Statistics, Rice University, Houston, TX; 2) Department of Bioengineering, Rice University, Houston, TX; 3) Departments of Pediatrics and Human and Molecular Genetics, Baylor College of Medicine, Houston, TX.

Identifying regions of identity-by-descent (IBD) using the observed identity-by-state (IBS) status is an effective approach in identifying disease-causing mutations in Mendelian disorders. A previously developed inhomogenous first-order hidden Markov model (HMM) was applied to genotype data produced from whole-exome sequencing (WES) data to identify chromosomal regions of IBD in siblings with an autosomal recessive disorder [Bioinformatics 27: 829-836, 2011]. Our approach redefines the observed and hidden state space from a binary status to the true IBD and IBS status and uses inhomogeneous transition probabilities to account for position, distance and sex-specific recombination rates. To improve prediction accuracy, we extend the HMM to incorporate conditional emission probabilities and show these conditional emission probabilities vary as a function of the minor allele frequency. This analysis suggests minor allele frequency should be included in the determination of IBD regions when using WES data. We evaluate the HMMs using simulated human WES data and real datasets to identify regions of IBD. Using the known IBD status from the simulated families, we compare the root mean squared error (RMSE) from each model averaged over a set of 100 simulated families. We show a first-order HMM with conditional emission probabilities using the redefined hidden IBD status has smaller root mean squared error (RMSE = 0.17) compared to the first-order HMM previously developed (RMSE = 0.22). Our IBD model which incorporates minor allele frequency provides researchers a tool to filter large portions of the exome more accurately when searching for the causal variant(s) associated with Mendelian disease. Supported by CPRIT grant RP101089, NCI grant CA155767, NCI T32 training grant CA096520 and NCN (Poland) grant 519579938.

1856W

Cryptic relatedness in epidemiologic collections accessed for genetic association studies. *J. Malinowski[1], R. Goodloe[1], K. Brown-Gentry[1], D.C. Crawford[1,2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Dept. of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

Over the last few years, epidemiologic collections have been a major resource for genotype-phenotype studies of complex disease given their large sample size, racial/ethnic diversity, and breadth and depth of phenotypes, traits and exposures. A major disadvantage of these collections is they often survey households and communities without collecting extensive pedigree data. Failure to account for substantial relatedness can lead to inflated estimates and spurious associations. To examine the extent of cryptic relatedness in an epidemiologic collection, we accessed the Third National Health and Nutrition Examination Survey (NHANES III), a population-based survey with DNA performed from 1991-1994 (n=7,159). Genome-wide genetic data is not available in NHANES III; though hundreds of SNPs genotyped in a variety of candidate genes are available for analysis. We performed identity-by-descent (IBD) estimates in three subpopulations of NHANES III: Hispanic (MEX, n=2073), non-Hispanic white (NHW, n=2631), and non-Hispanic black (NHB, n=2108) using PLINK to identify potential familial relationships from presumed unrelated subjects. After quality control, we calculated IBD with 784 SNPs in MEX, 721 SNPs in NHW, and 691 SNPs in NHB. In MEX, we identified two potential identical twin relationships ($z_2 \geq 0.90$, π-hat>0.95), 312 potential parent/child relationships ($z_1 \geq 0.95$, $0.48<\pi$-hat<0.53), and 98 potential full-sibling relationships ($0.40<z_1<0.60$, $0.15<z_0/z_2<0.35$, $0.40<\pi$-hat<0.60). In NHW we did not observe any potential identical twin relationships, however, we did observe 1833 potential parent/child relationships ($z_1 \geq >0.95$, $0.48<\pi$-hat<0.50) and 18 full-sibling relationships ($0.40<z_1<0.60$ and $0.15<z_0/z_2<0.35$, $0.40<\pi$-hat<0.60). We identified two potential identical twin relationships in NHB ($z_2 \geq 0.93$, π-hat>0.95), 376 potential parent/child relationships ($z_1 \geq 0.95$, $0.48<\pi$-hat<0.52), and 44 potential full-sibling relationships ($0.40<z_1<0.60$, $0.15<z_0/z_2<0.35$, $0.40<\pi$-hat<0.60). We did not observe any 1st degree cousin relationships ($0.70<z_0<0.80$, $0.20<z_1<0.30$) in any of the three subpopulations. We identified numerous potential 2nd degree relationships ($0.45<z_0/z_1<0.55$, $0.23<\pi$-hat<0.27) such as half-siblings, grandparent/grandchild, and uncle/nephew in each of the three subpopulations, though PLINK was unable to discriminate among these relationships. Despite the lack of genome-wide data, our results suggest substantial cryptic relatedness in this epidemiologic collection.

1857T

IBDLD-3: A fast and parallelized software package for IBD estimation from genomic data. *L. Han, M. Abney.* Human genetics, University of Chicago, Chicago, IL.

The estimation of identity by descent (IBD) from genomic data has gained much interest recently as a tool for a variety of genetic analyses. Here we present a new implementation of our IBD estimation method that is a dramatic improvement in both speed and accuracy over our original software package. Computation time is substantially reduced due to, in part, the use of OpenMP to parallelize execution over many processors or cores. We have also implemented methodological improvements in our hidden Markov model based approach both for estimating the parameters of the model and in the computation of posterior IBD probabilities that result in both higher accuracy in the estimates and computational speed-ups. These improvements apply for both the case where pedigree information is available and when it is not. Our simulation studies show that IBDLD3 is much more accurate at estimating the true IBD sharing than the original implementation with reduced noise giving substantially improved performance at both point-wise estimates of IBD and in the detection of segments. IBDLD3 is robust to genotype and pedigree error, or other forms of misspecified relationships, and is computationally fast enough to estimate all possible IBD sharing states at every SNP from a high-density genotyping array for hundreds of thousands of pairs of individuals. The open source C++ software package IBDLD v3.1 is freely available to be downloaded.

1858F

Combinatorial Conflicting Homozygosity (CCH) enables the rapid identification of genetic linkage in the presence of multiple phenocopies. *A.P. Levine[1], T.M. Connor[1], D.D. Oygar[2], G.H. Neild[1], P.H. Maxwell[3], A.W. Segal[1], D.P. Gale[1].* 1) Division of Medicine, University College London, London, United Kingdom; 2) Nicosia State Hospital, Burhan Nalbantoglu General Hospital, Nicosia, North Cyprus; 3) School of Clinical Medicine, Cambridge University, Cambridge, United Kingdom.

The analysis of multiplex kindreds has been successfully used to identify rare disease-linked variants; however genome-wide linkage analysis of such pedigrees can fail to identify linked loci in the presence of phenocopies, as can occur where the phenotype is either non-specific or incompletely ascertained. To overcome these challenges, we developed a method based on the principle of conflicting homozygosity (CH) in which runs above a certain length of consecutive biallelic markers containing no occurrences of homozygosity for both major and minor alleles across a set of affected individuals are inferred to result from identical by descent (IBD) inheritance. An equivalent approach was described by Leibon et al. and Thomas et al. (2008), referred to as Shared Genomic Segment Analysis. We add a combinatorial feature to our method, packaged in a Python software program, in which the analysis is automatically repeated on all subsets of affected individuals. To characterise the parameter thresholds that define a locus as one likely inherited IBD, we analysed subsets of unrelated HapMap individuals. Gene-dropping simulations in pedigrees defined the sensitivity and specificity of our approach. We applied our program to a large family in which microscopic haematuria, proteinuria or renal impairment segregated as an autosomal dominant trait in 17 members. The proband was wild type for a panel of candidate mutations. Linkage analysis (which took >3 days) failed to identify one or more loci with suggestive or significant linkage. Consistent with this, CH analysis of all 17 affecteds (which took <2 seconds) demonstrated no loci shared IBD. We hypothesised the presence of one or more phenocopies and sequentially analysed all possible combinations of subsets of progressively decreasing size of the 17 affected individuals. Analysis of all 12,376 combinations of 11 of the 17 affecteds (which took ~7 hours running serially) identified three loci, one of which contained the gene *COL4A3*, mutations in which are known to cause familial kidney disease. Sequencing identified the previously described pathogenic G871C mutation in all 11 affecteds carrying the haplotype, no unaffected individuals and none of the six affecteds who did not share the linked haplotype (including the proband). We propose that combinatorial CH can be used to identify loci shared IBD in the presence of phenocopies and may be of value in gene hunting.

1859W

Investigating the Importance of Disparate Genetic Influences across African and European Descent Populations. *T. de Candia[1,2], M.C. Keller[1,2].* 1) Psychology & Neuroscience Dept, University of Colorado, Boulder, CO; 2) Institute for Behavioral Genetics, University of Colorado, Boulder, CO.

The degree to which genetic influences overlap across ethnicity has been a matter of contention. Besides the possibility that causal variants (CVs) differ between ethnicities, traits may be differentially predicted by genes due to different SNP-CV linkage disequilibrium (LD) patterns as well as artifactual differences in plates or genotyping procedures. To systematically investigate the factors that can affect genetic correlations tagged by SNPs between ethnic populations, we simulated CVs in 10849 African descent (AD) and 10007 Europen descent (ED) genotypes from four datasets (ARIC, WHI, MESA, CARDIA). A random set of 2,000 SNPs that passed quality control procedures and that were largely independent of each other were chosen as CVs for each ethnicity. For each CV, we constructed a continuous phenotype with a heritability of 0.5, and used a bivariate mixed linear model (Lee et al, 2012) to estimate the heritabilities and correlation tagged by nearby SNPs. We then explored the extent to which LD and CV allele frequencies, as well as differences in these parameters across ethnicity, predicted these parameters. Across CVs, heritabilities were underestimated for both ADs (0.40, sd=0.11) and EDs (0.44, sd=0.12), and as expected, this appeared to be be explained by incomplete LD ($r2=.33$ in ADs and .17 in EDs, $p < 1.55e-06$). We found that the mean genetic correlation was 0.94, only slightly lower than its expected value of 1. Controlling for minor allele frequencies differences, differences in LD patterns across ethnicities lowered genetic correlations ($r=-0.25$, $p < 2e-16$). On the other hand, controlling for LD pattern differences, allele frequency differences did not predict differences in genetic correlations ($r=-0.01$, $p > 0.05$). Our findings suggest that genetic correlations will be high between ethnic groups when CVs are the same, despite differences between ethnic groups in SNP LD patterns and MAFs. Thus, low genetic correlations between ethnic groups might suggest different causal variants are influencing the phenotypes and/or technical aspects of the design artificially reduce genetic correlations. Of the two factors investigated here, differences in LD patterns across ethnicities appeared to have a greater influence on SNP-correlations than did differences in MAFs. We discuss the implications of these findings on genetic correlations we observe for height and BMI.

1860T

Employing Identity by Descent to Inform Rare Variation in Gene Based Tests. *D.S. Park, C. Gignoux, L. Fejerman, D.G. Torgerson, D. Hu, S. Huntsman, C.S. Eng, E.G. Burchard, E. Ziv, N. Zaitlen.* University of California San Francisco, San Francisco, CA.

Genetic variants genotyped by commercial genotyping platforms capture approximately half of the estimated heritability of the dozens of complex human phenotypes examined to date (Visscher et al. 2012). Rare variants may make up a substantial fraction of the missing heritability (Manolio et al. 2009). Recent advances in sequencing technologies have enabled deeper exploration of rare genetic variants and their role in disease susceptibility than was previously possible with array based genotyping (Eichler et al. 2010). Although sequencing costs are likely to continue to fall, the resources to achieve the sample sizes necessary to perform a well-powered rare variant analyses can be prohibitive (Liu et al. 2013). In this work, we examine the use of segments of pairwise identity by descent (IBD) computed from array based genotype data to serve as a proxy for rare variants, preventing the need for expensive sequencing and permitting rare variant analyses of existing data sets. Our assumption is that the majority of rare variants have occurred recently and therefore exist on the same IBD background. We apply existing methods (B. L. Browning and Browning 2011) (Gusev et al. 2009) to identify segments of IBD and a clustering approach to group individuals with pairwise IBD status commensurate with an untyped rare variant (Gusev et al. 2011). However, instead of testing clusters for association directly, we turn each cluster into a 'pseudo-SNP' which represents an untyped rare variant. Since traditional single-variant tests tend to be inadequately powered to detect associations with rare variants, various gene-based tests have been proposed to examine the combined effects of rare variants. We test genes for association using SKAT in combination with our IBD based pseudo-SNPs (Wu et al. 2011). We applied our method to two Latino cohorts, one consisting of asthmatics from the GALA cohort (Burchard et al. 2004) (Kumar et al. 2013) and the other consisting of breast cancer cases (Fejerman et al. 2008). We examine in detail the underlying assumption of our method by determining the fraction of rare variants captured as pseudo-SNPs in several exome and whole genome sequencing data sets.

1861F

Computationally-efficient long-range phasing with very large datasets. *M.J. Barber[1], R.E. Curtis[2], K. Noto[1], Y. Wang[1], J.M. Granka[1], N.M. Myres[2], J.K. Byrnes[1], C.A. Ball[1], K.G. Chahine[2].* 1) Ancestry.com, San Francisco, CA; 2) Ancestry.com, Provo, UT.

While computationally intensive, phasing a large set of genotypes (i.e. > 100,000 samples) into probable haplotypes presents the opportunity to leverage sample size to increase phasing accuracy for each and every sample. Phasing switch error rate can be greatly minimized when using genotypes with a specified parent-offspring relationship, but such datasets are not universally available. Recent advances in methodology have utilized the large number of identity-by-descent segments (IBD-SEGs) between a sample and the rest of the samples in the dataset to improve phasing accuracy. These 'long-ranged' phasing approaches use an IBD-SEG to help phase the matched IBD region. At AncestryDNA, we are applying the principle of long-range phasing to very large (and growing) datasets of genotyping data. Our approach assembles high-confidence IBD-SEGs to form an explicit surrogate parent for each sample. The assembly of an explicit surrogate parent is computationally efficient: the only requirement is assessing each IBD-SEG for quality separately, rather than performing a joint analysis. Given an explicit surrogate parent, phasing and IBD-SEGs can then be updated. Our approach has the added advantage of enabling updates to the phase estimates of full or partial genotypes when new high-quality IBD-SEGs are identified, as is common in constantly growing databases such as AncestryDNA's. We test the accuracy of our approach using simulated genotyping datasets and thousands of confirmed parent-offspring relationships from the AncestryDNA database. Our novel approach aims to efficiently and accurately phase large numbers of samples in a way that could be relevant and widely practical for a variety of applications, including datasets from genome-wide association studies that are being generated by the genetics community.

1862W

IBD Mapping of Autism Microarray Data. *G. Povysil, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria.

Identity by descent (IBD) between two individuals means that their alleles are identical because they were inherited from a common ancestor. This information can be used via IBD mapping to increase the power of association analysis by grouping single nucleotide variants (SNVs) based on IBD. The basic principle of IBD mapping is to look for segments of DNA that are shared identical by descent more often among cases than controls.

HapFABIA is a biclustering algorithm that was originally designed to extract short IBD segments that are present in multiple individuals from large sequencing data using only rare variants. In this setting we used HapFABIA on SNP microarray data from the Autism Genetics Resource Exchange (AGRE) to look for IBD segments that are shared more often by cases than controls.

We found several IBD segments that were almost exclusively shared by cases. Some of these map to genes that have been previously associated with autism. Further analyses are needed to confirm or dismiss these results.

1863T

Individual Perspective Measures of Population Structure. *X. Zheng, B. Weir.* Department of Biostatistics, University of Washington at Seattle, Seattle, WA.

The methods of principal component analysis (PCA) and hierarchical clustering on individual genotypic data have been widely used to detect population structure in genome-wide association studies. The principal component axes often represent perpendicular gradients in geographic space, and we provide an interpretation of PCA based on relatedness measures, which are described by the probability that sets of genes are identical-by-descent (IBD). An approximately linear transformation between the projection of individuals onto the principal components and allele admixture fractions assuming two or more ancestral populations is revealed. Furthermore, a measure of individual dissimilarity based on the coancestry coefficient was proposed for hierarchical clustering. Compared to other dissimilarity measures (such like identical-by-state methods), its expected value is directly related to the kinship coefficient without being confounded by allele frequency, and it is a moment estimator and suited for large-scale GWAS data. Both PCA and hierarchical clustering were applied to HapMap Phase II and III data. The population admixture proportions inferred by PCA are consistent with what HAPMIX and ADMIXTURE estimate. Hierarchical clustering successfully separates Chinese and Japanese samples in HapMap Phase II data based on their coancestry coefficients. Finally, a combination of PCA and hierarchical cluster analysis should help us better understand population structure for isolated and admixed populations.

1864F

Model-free Estimation of IBD Sharing Probabilities in Admixed Populations. *M. Conomos.* University of Washington Department of Biostatistics, Box 357232, Seattle, WA 98195.

A large number of genetic studies involve samples from structured populations, including those with ancestry admixture. Estimating measures of relatedness such as kinship coefficients and identity by descent (i.b.d.) sharing probabilities in these samples is a challenge. Commonly used relatedness estimators, such as those implemented in PLINK, the genetic relationship matrix (GRM), and the EM algorithm, assume population homogeneity and are biased in the presence of population structure. The REAP method (Thornton et al., 2012) addresses relatedness estimation in admixed populations by utilizing individual specific allele frequencies. In order to estimate individual specific allele frequencies, however, REAP relies on estimates of admixture proportions and ancestral allele frequencies from likelihood based methods such as ADMIXTURE (Alexander et al., 2009) or FRAPPE (Tang et al., 2005). As a consequence, REAP is inherently sensitive to mis-specification of model assumptions in these methods; mis-specification of the number of ancestral populations represented in the sample, or insufficient reference samples from them, may lead to inaccurate admixture proportions and/or biased allele frequency estimates. We propose a model-free method of estimating i.b.d. sharing probabilities using principal components without the requirement of external reference samples. We accomplish this through the use of a modified GRM, similar to that in REAP, along with an analogous novel estimator for the probability of sharing two alleles i.b.d. that relies on an alternative coding of genotype values. We demonstrate the utility of our method by estimating i.b.d. sharing probabilities of more than 12,000 African Americans and Hispanics from the Women's Health Initiative study.

1865W

Homozygosity mapping combined with linkage analysis in human families in the age of high-density DNA variants. *J. Ott[1,2], Y. Li[3], E. Engle[4], S. Shaaban[5].* 1) Institute of Psychology CAS, Beijing, Beijing, China; 2) Ott Lab, Rockefeller University, New York; 3) School of Statistics, Shanxi University of Finance & Economics, Taiyuan, Shanxi, China; 4) Departments of Neurology, Medicine, and Ophthalmology, Harvard Medical School, Howard Hughes Medical Institute, Boston, MA; 5) Department of Neurology, Harvard Medical School, Dubai Harvard Foundation for Medical Research, Boston, MA.

Exome sequence data are often obtained in small human families. Extracting SNPs from variant (vcf) files allows for linkage analysis but power is often low. Also, multipoint linkage analysis tends to be problematic due to correlations among closely spaced SNPs. Here we propose a modern, new approach: (1) At each SNP, only one lod score (at $\vartheta = 0.0001$) is calculated. (2) With SNPs being in chromosomal order, a run of positive lod scores (ROL) is defined as a set of contiguous SNPs all with lod scores > 0.01, delimited by SNPs with lod scores $\leq$ 0.01 or a chromosome end. ROLs are analogous to runs of homozygosity (ROH) in homozygosity mapping in unrelated individuals but are more realistic as they take into account allele frequencies and family relationships. (3) For two families with a given disease, we focus on long ROLs (eg. > 500 kb) and count the number N of ROLs between the two families that overlap by k base pairs, eg. $k = 1000$ bp. (4) We approximate the null distribution of N by computer simulation by randomly placing observed ROLs on chromosomes so as to obtain a p-value associated with N. We applied our new method to two 2-generation families with comitant strabismus, each with 1 or 2 affected children and two related parents. Traditional linkage analysis furnished max. lod scores < 2 (genome-wide p >> 0.05). Focusing on the 80 longest ROLs in each family, our approach identified $N = 17$ ROLs overlapping by at least 1 bp ($p = 0.1319$), and $N = 16$ ROLs overlapping by at least 1 Mb ($p = 0.0027$). Thus, this approach finds strong genome-wide significance where traditional linkage analysis is unsuccessful, and at the same time allows replicating linkage results in a second family, with both families being of only modest size. As negative controls, (1) we applied our method to two families with very different diseases, with resulting p-values >> 0.50; (2) also, when our eye disease families are analyzed under no linkage then no single overlap is observed. ROLs significantly overlapping between two families presumably harbor common susceptibility variants. Thus, by design, this procedure takes into account multiple disease genes, possibly located on different chromosomes. Our new method is expected to be more powerful for recessive than dominant acting variants but this is not a serious shortcoming as most mutations are recessive.

1866T

A statistical approach for variant calling using pedigree information and phase informative reads. *K. Kojima, N. Nariai, T. Mimori, M. Takahashi, Y. Yamaguchi-Kabata, Y. Sato, M. Nagasaki.* Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi Japan.

Variant detection from genome-wide sequencing data is essential for the analysis of disease causing mutations and elucidation of disease mechanisms. However, variant calling in low coverage regions is difficult due to sequence read errors and mapping errors. Hence, variant calling approaches that are robust to low coverage data are demanded. We propose a new variant calling approach that considers pedigree information and haplotyping based on sequence reads spanning two or more heterozygous positions termed phase informative reads. In our approach, genotyping and haplotyping by the assignment of each read to a haplotype based on phase informative reads are simultaneously performed. Therefore, positions with low evidence for heterozygosity are rescued by phase informative reads, and such rescued positions contribute to haplotyping in a synergistic way. In addition, pedigree information supports more accurate haplotyping as well as genotyping, especially in low coverage regions. Since undesirable influence from homozygous positions to the read assignment prevents accurate haplotyping, we introduce latent variables that determine zygosity at each position, and avoid the influence from homozygous positions by using the latent variables. In performance evaluation with a parent-offspring trio sequencing data,our approach outperformed existing approaches in accuracy on the agreement with SNP array genotyping results. Also, performance analysis considering distance between variants showed that the use of phase informative reads is effective for the accurate variant estimation, and further performance improvement is expected with longer sequencing data.

1867F

Mathematical model explains gender differences in intelligence by variation in mating preferences. *M. Nagel.* Molec Gen, Ctr Nephrology, Weisswasser, Germany.

Gender differences in variance in intelligence measures is a contentious issue. While some investigators insist that males show greater variance than females, other fervently disagree. By our study we show that at least part of the controversy may be explained by cultural and social differences in mating behavior. Though previous work already mentioned the tremendous effect that mating preferences can have, it did not include such possibilities into their models. The mathematical model proposed by this study allows to evaluate changes in variance in respect to mating preferences. The model presumes that most of the alleles determining human intelligence are located on the X chromosome. The model works no matter whether those effects on intelligence are positive or negative. Therefore this assumption is~given the considerable number of known X-linked intellectual deficiencies~quite plausible. The model follows the mathematical technique of folding two probability functions. The probability mass function of cognitive (dis)abilities, which is binomial, and the hypothetical probability density function of mating preferences, which can define any probability distribution, however skewed. The model predicts that with improved mating precision~that is partners mate more precisely in the same range of intelligence~the long disputed greater variance in intelligence measures among male individuals vanishes and becomes equal to that of women. It can be hypothesized that recent social-economical changes had exactly that effect. It can be further hypothesized, that the same effect has implications for the evolution of human intelligence.

1868W

Accounting for model structure uncertainty when predicting intervention effects from observational data. *M. Rantalainen, C.C. Holmes.* Dept Statistics, Univ Oxford, Oxford, United Kingdom.

Prediction of the consequences of intervention effects, such as gene knockdowns, from observational data is of central interest in biomedical sciences. Experimental perturbations are expensive and sometimes impossible due to ethical consideration. This motivates the development of statistical and causal methods to prioritize between experiments. Graphical models together with intervention calculus provide a formal framework for predicting bounds on intervention effects from observational data but only when the underlying graphical model is known. Recently developed methods have shown how the graph itself can be inferred from data, with the limitation that only bounds on the causal effects can be estimated (Maathuis, M. H. et al., Estimating high-dimensional intervention effects from observational data. Annals of statistics, 2009). These methods have been shown to outperform (non-causal) statistical methods in terms of predicting intervention effects, and have also been validated through perturbation experiments. These methods are, however, based on accurate inference of the equivalence class of a Directed Acyclical Graph (DAG) from data. Here we present an extension to previously reported methods that allow for uncertainty in the inferred DAG, combining ideas from the area of causal inference with a Bayesian treatment of the structure of the graphical model. Our objective is to properly account for the uncertainty associated with the structure of the graphical model that is inferred from molecular profiling data. The proposed method utilises Markov Chain Monte Carlo sampling over orders combined with Bayesian model averaging to approximate the posterior of intervention effects. We report simulation results that demonstrate under realistic conditions the proposed methods have considerable advantages over existing methods, as well as results from applications to real omics data in the context of cardiovascular disease.

1869T

A variance-component based approach to pathway analysis in expression data. *E.E. Quillen, V.P. Diego, E. Drigalenko, M.P. Johnson, J.W. MacCluer, T.D. Dyer, E.K. Moses, M.C. Mahaney, J.E. Curran, H.H.H. Göring, L. Almasy, J. Blangero.* Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX.

The application of pathway and gene-set based analyses to high throughput data is increasingly common and represents an effort to understand underlying biology where single-gene or single-marker analyses have failed. Many such analyses rely on the a priori identification of genes associated with the trait of interest. In contrast, this variance-component based approach creates a distance matrix between individuals based on the covariance in expression of genes within each pathway. Previously generated mRNA data from 1240 Mexican-American participants in the San Antonio Family Heart Study was considered in relation to circulating HDL levels. Previous single-gene and over-representation analyses of HDL in this sample identified one gene and four associated pathways. Transcript levels were ascertained from total RNA isolated from lymphocytes and analyzed with the Illumina Sentrix Human Whole Genome (WG-6) Series I BeadChips. All probes were standardized by z-scoring within individuals. 16,681 probes were detectable above baseline levels in the majority of individuals, have non-zero heritability, and can be annotated using ReMOAT mapping. Expression levels for genes composing 228 pathways drawn from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database were adjusted for age and sex before inverse normalizing. For each pathway, the normalized values were used to construct matrices of Malhanobis distances between each pair of individuals in R. Using SOLAR, a polygenic null model of HDL level was generated including sex and age as covariates. A model including each distance matrix in turn was compared to the null model to determine if the pathway is significantly associated with HDL. A final model was constructed by iteratively including each of the associated pathway matrices, in order of magnitude of effect, and removing matrices that do not explain additional variation, likely due to overlap in genes between the pathways. This method has a number of advantages when compared to single transcript and pathway over-representation analyses, including the ability to estimate the proportion of variation explained by each pathway and all pathways combined, as well as the logistical advantage of only calculating the distance matrices once for each mRNA data set regardless of the number of phenotypes. Most importantly, this method allows for the simultaneous consideration of multiple pathways to determine if they represent independent associations.

1870F

Genetic pathways enriched with type 1 diabetes suggest novel causal genes for type 1 diabetes. *M. Evangelou, D.J. Smith, O.S. Burren, N.M. Walker, J.A. Todd, C. Wallace.* JDRF/WT Diabetes and Inflammation Laboratory , Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom.

Pathway analysis can complement single-SNP analysis in exploring genomewide association (GWA) data to identify disease associated genes. We applied a gene-based competitive pathway analysis to previously published type 1 diabetes (T1D) GWA data covering a total of 5,916 T1D cases and 7,338 controls, all of white European origin. After mapping each SNP to a unique gene according to distance, we explored three different statistics that summarize the association of each gene (the minimum single SNP p-value, the mean -log p value, and the Fisher's product statistic) and two methods for summarizing the association across genes in a pathway (Fisher's product method and the adaptive rank truncated product method).

Of 314 BioCarta and 1,272 Reactome pathways, 30 and 101, respectively, were associated using at least one statistic, and just five (three BioCarta and two Reactome) were associated using all six statistics. The BioCarta pathways were 'IL-2 receptor beta chain in T-cell activation', 'Antigen dependent B-cell activation' and 'The Co-stimulatory signal during T-cell activation' and the Reactome pathways were 'GRB7 events in ERBB2 signaling' and 'Regulation of IFNG signaling'. This echoes much that is known about the etiology of type 1 diabetes.

We identified SNPs with small GWAS *p*-values near genes in associated pathways that have not been reported as associated with T1D. The genes *FASLG, RAF1* and *SOCS1* were identified as potential novel gene risks of T1D. The *FALSG* is associated with Celiac disease but has not reported as associated with T1D, whereas the *SOCS1* gene is associated with a number of autoimmune diseases including T1D. The *RAF1* gene has not been reported as associated with T1D or with any other autoimmune disease. SNPs near these three genes are currently being genotyped in a replication cohort. Such replication will prove the utility of pathway analysis not only for identifying pathways of interest, but for using this information to alter our prior belief of association and thus identify novel disease associated SNPs that do not reach stringent genomewide significance levels.

1871W

A Powerful Statistical Method for Genetic Pathway Analysis. *N. Liu, Q. Yan, N. Yi.* Dept Biostatistics, Univ Alabama at Birmingham, Birmingham, AL.

Genetic pathway represents certain biological mechanisms of diseases. Pathway analysis takes into account the biologically pathway-gene-marker hierarchical structure, while traditional analysis strategy for genome-wide association studies (GWAS) usually focuses on single marker analysis. Therefore pathway analysis may potentially complement single-marker analysis and provide additional insights for the genetic architecture of complex diseases, thus may lead to higher statistical power and more biologically meaningful interpretation. In this work, we proposed a new method for genetic pathway analysis. It is based on kernel machine testing and adaptive rank truncated product test. The genotype data of 2000 subjects from WTCCC Type I Diabetes were used for simulation studies. We compared the performance of our new method and several other methods. The preliminary results show that the new method has the highest power while maintaining correct Type I error rate in most of the scenarios we considered.

1872T

A supervised dimension reduction approach for pathway-based analysis in Genome-wide association study. *Z. Wei[1], J. Li[2], W. Wang[1], H. Hakonarson[2].* 1) New Jersey Institute of Technology, Newark, NJ; 2) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Pathway-based gene set enrichment analysis (GSEA) has been routinely conducted as a complementary approach for conventional single-SNP based association tests in genome wide association study (GWAS). Since the first simple extension from the GSEA for microarray gene expression data years ago, methodology research is undergoing rapid development for improving GSEA in GWAS. Quite a few new methods have been proposed. One common strategy is to utilize dimension reduction technique, hoping to improve power by reducing the unnecessary large degree of freedom. However, most of existing methods are unsupervised approaches, namely, not exploiting trait information, when reducing dimensionality. Here we employ a supervised strategy by applying the ridge regularized Kernel Sliced Inverse Regression (KSIR) to achieve dimension reduction. Not only can KSIR exploit trait information, but also it is capable of capturing non-linear effects by flexibly employing various non-linear kernel functions. Using simulation studies, we show that the KSIR method outperforms conventional unsupervised competing methods in terms of causal pathway ranking and the statistical power. We also demonstrate the superior performance of KSIR in analysis of a real dataset, the WTCCC Ulcerative Colitis dataset consisting of 1762 cases and 3773 controls as the discovery cohort, and 591 cases and 1639 controls as the replication cohort. We identify several immune and non-immune pathways relevant to Ulcerative Colitis, including some novel ones that may be worthy of further investigation for their roles in the development of Ulcerative Colitis.

1873F

Rank-based analysis of transcriptome data reveals biologically relevant atopic dermatitis genes and pathways. *TB. Mersha[1], D. Ghosh[2], GK. Hershey[1], JA. Bernstein[2].* 1) Division of Asthma Research, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH; 2) Immunology and Allergy, Department of Internal Medicine, University of Cincinnati, Cincinnati, OH.

Atopic Dermatitis (AD), a chronically relapsing inflammatory disorder of the skin, affects an estimated 15 to 30% of children and 2 to 10% of adults. While the exact causes of AD are not fully understood, genetic differences may also be responsible for AD susceptibility. Several independent groups have identified differentially expressed genes (DEGs) between healthy versus AD patients using microarray technology. However, there is little overlap in the DEG lists reported by different groups due to variation from random noise, biological and experimental differences, as well as differences in the extraction and handling of RNA samples. Therefore, a statistical rank-based analysis is necessary to identify a set of genes that are consistently dysregulated among multiple independent microarray studies. We obtained AD-related expression data from the GEO dataset (http://www.ncbi.nlm.nih.gov/geo/; accessed in March 2013) using 'human [organism] AND Atopic dermatitis' search terms. In this study we focused on six datasets after eliminating data obtained from non-human/ex-vivo cell-culture experiments. We, then, screened genes differentially expressed between skin samples from AD and healthy control subjects using classical t-test and p-values of <0.05 as significant threshold. We identified differentially expressed genes (with fold change >=1.5) consistently up/down-regulated in at least 5 out of 6 datasets and ranked them according to their average fold changes. The associations of particular gene across studies may represent a true association with AD and should be given highest priority. We identified 24 DEGs, out of which 8 were found to be consistently upregulated, while 16 genes were consistently downregulated in all the datasets analyzed. Top upregulated genes were KRT16 (related to Keratinocyte production) and LTF (related to the expression of lactoferritin, as anti-microbial protein), while the top down-regulated genes were LOR (loricrine), LCE2B (late cornified envelop 2B), FLG (fillaggrin) genes- mostly related to skin barrier function. Pathway analysis revealed that the DEGs are related to epidermal development or inflammatory pathways. Thus, rank-based analysis of publicly available data indicates biologically relevant genes and pathways thus providing mechanistic clues to observed gene expression patterns in AD.

1874W

Novel Top Down Dissection of Complex Traits Based on Germline Signatures. *H.K. Im[1], N.J. Cox[2].* 1) Department of Health Studies, University of Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL.

Genome-wide association studies and more recently whole genome/exome sequencing studies have identified thousands of genetic variants reproducibly associated with many complex diseases and traits. However, by and large the biological mechanisms underlying these discoveries are still lacking. In addition, it is becoming increasingly clear that a large number of variants with modest effect sizes are contributing a substantial proportion of the total heritability of complex traits. Thus methods that aggregate variants into biologically meaningful units and integrate functional and regulatory evidence are dearly needed. Several gene-based approaches have emerged in order to address this problem. We propose to go a step further with a top down systems level approach that dissects the anatomy of complex traits in terms of higher level biological processes and components. For this purpose, we are in the process of developing a catalog of germline signatures (GSigDB) using cellular, animal, and bioinformatic models as well as clinical and non-clinical human data. These signatures allow testing novel hypothesis regarding disease etiology, pleiotropy among traits, to name a few. These signatures are stored as optimal weights to be applied to whole genome data and generate polygenic scores or predictions of endophenotypes. The weights are computed integrating data from existing large scale meta analysis and regulatory and functional information. For example, for the insulin resistance signature we use the effect sizes from the meta analysis of ~50K individuals published by the MAGIC consortium. We applied our approach to the 7 diseases from Wellcome Trust Consortium. We found that both insulin resistance (p=0.0029) and beta cell function (p=0.053) have similar effect sizes on the risk to type 2 diabetes. Interestingly, we also found that type 1 diabetes risk shows a significant association with insulin resistance (p=0.054) and beta cell function (p=0.040) with similar effect sizes. BMI was not significantly associated with type I diabetes, as expected. For the 5 other diseases there was no significant association with insulin resistance or beta cell function except for a negative association between Crohn's disease and insulin resistance (p=0.0057) . This finding is intriguing but needs to be replicated in an independent dataset. In conclusion, our top down approach allows us to ask high level biological questions and generate new mechanistic hypothesis.

1875T

A systems-biology approach to identify and prioritize sub-networks of functionally-related genes for Alzheimer's disease and subsequent in vivo validation of candidate genes using a *C. elegans* model of Aβ toxicity. *S. Mukherjee[1], M. Kaeberlein[2], J. Kauwe[3], A. Naj[4], P. Crane[1], Alzheimer's Disease Genetics Consortium.* 1) Department of Medicine, University of Washington, Seattle, WA; 2) Department of Pathology, University of Washington, Seattle, WA; 3) Departments of Biology and Neuroscience, Brigham Young University; 4) Perelman School of Medicine, University of Pennsylvania.

Background: Recent genome-wide association studies (GWAS) have identified around 20 variants as late-onset Alzheimer's disease (LOAD) susceptibility loci in whites. In addition to these single loci tests, it is important to detect and understand combined effects of multiple associated genes on LOAD. We performed a preliminary network analysis incorporating human protein-protein interaction database from BioGRID to the HapMap2-imputed combined ADGC data set. Post-GWAS, this helps researchers to prioritize functionally related genes and networks that are of the highest biological relevance underlying the pathogenesis of LOAD. Methods: We combined HapMap2-imputed data sets from 15 studies after performing strict quality control. We performed a case-control association for LOAD adjusting for population sub-structure and study sites on a set of 19,692 unrelated individuals using PLINK and those results were used to perform a gene-wide analysis using VEGAS. The gene-wide association results were then integrated into the human protein-protein interaction network using a dense module searching (DMS) method to identify candidate genes or sub-networks for LOAD. We then attempted to functionally validate candidate genes from this network in vivo using a transgenic *C. elegans* model of $A\beta_{1-42}$ toxicity. Results: The network analysis identified several of the known LOAD risk loci as well as other genes such as *LPL, HSF1, EGFR STAT5B,* and *UBC* to be strongly associated with LOAD. RNAi knockdown of the *C. elegans* orthologs of *HSF1 (hsf-1)* or *UBC (ubq-1 or ubq-2)* significantly accelerated the age-associated onset of $A\beta_{1-42}$ toxicity. Conclusions: We were able to identify a set of significant modules and candidate genes, including some well-studied genes not detected in the single-marker analysis of GWA studies for LOAD, and to demonstrate a role for two of these genes as modifiers of Aβ toxicity in *C. elegans*. This approach provides complementary data to a GWAS of a complex disease phenotype by incorporating biological knowledge derived from protein-protein interactions and allows for initial functional validation *in vivo*. Further functional enrichment analysis is needed to determine whether these novel loci may provide targets for interventions to ameliorate LOAD.

1876F

Hierarchical learning of regulatory networks from RNA-sequencing of 35 human tissues. *E. Pierson, S. Mostafavi, A. Battle, D. Koller, The GTEx Consortium.* Stanford University.

Understanding the regulatory interactions between genes, and identifying those interactions that are shared across human tissues and those that are specific to each tissue, is essential for understanding regulatory mechanisms that govern similarities and differences between tissues and can help reveal tissue-specific responses to environmental or genetic perturbations. The GTEx project, with RNAseq data from more than 1,000 samples across 35 human tissues, provides an opportunity for such analysis. Here, we utilize the GTEx dataset to learn regulatory networks across these 35 tissues, where each network captures interactions between genes in a particular tissue. To do so, we present a method that makes use of a tissue hierarchy, enabling us to learn such networks even for tissues with very few samples. Our method consists of two components. First, we derive a tissue hierarchy that reflects the relationships between tissues based on expression data using hierarchical clustering, and is shown to be consistent with tissue lineage relationships. We then develop a fast algorithm that uses this hierarchy to learn sparse regulatory networks in each tissue, using a L2-regularization penalty to enforce the fact that tissues nearby in the hierarchy should have similar networks. This regularization increases accuracy by allowing for transfer learning from tissues for which we have many samples to tissues for which we have few. We show, using cross validation on three different gene sets, that this algorithm learns networks more accurately than either learning a single network for all tissues or learning a network for each tissue independently. We discuss the broad applicability of this algorithm to any set of hierarchically related networks, such as those from phylogenetic or cancer datasets. We further validate the accuracy of our learned networks by comparison to four external databases: COEXPRESdb, TS-CoExp, previously identified transcription factors, and Gene Ontology. Finally, we analyze hubs and clusters in the learned networks to identify conserved and distinct patterns across tissues. We find that tissue-specific transcription factors have fewer connections in our networks than the average gene, and that general transcription factors have more, along with network connectivity consistent with shared and tissue-specific functional relationships.

1877W

Seasonal changes in gene expression represent cell type composition in whole blood. *M. Neeleman[1], S. de Jong[2], J.J. Luykx[1,3], M.J. ten Berg[4], E. Strengman[5], H.H. ten Breeijen[6,7], L.C. Stijvers[4], J.E. Buizer-Voskamp[1], S.C. Bakker[1], R.S. Kahn[1], S. Horvath[8,9], W.W. van Solinge[4], R.A. Ophoff[2,8].* 1) Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA; 3) Department of Psychiatry, ZNA hospitals, Antwerp, Belgium; 4) Department of Clinical Chemistry & Haematology, Division of Laboratories & Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 6) Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences; Faculty of Science, Utrecht University, The Netherlands; 7) Department of Clinical Pharmacy, University Medical Center Utrecht, The Netherlands; 8) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA; 9) Department of Biostatistics, School of Public Health, University of California, Los Angeles, California 90095, USA.

Seasonal patterns in behavior and biological parameters are widespread. Here, we examined seasonal changes in whole blood gene expression profiles of 233 healthy subjects. Using weighted gene co-expression network analysis, we identified three co-expression modules showing circannual patterns. Enrichment analysis suggested that this signal stems primarily from red blood cells and blood platelets. Indeed, a large clinical database with 51,142 observations of blood cell counts over three years confirmed a corresponding seasonal pattern of counts of red blood cells, reticulocytes and platelets. We found no direct evidence that these changes are linked to genes known to be key players in regulating immune function or circadian rhythm. It is likely, however, that these seasonal changes in cell counts and gene expression profiles in whole blood represent biological and clinical relevant phenomena. Moreover, our findings highlight possible confounding factors relevant to the study of gene expression profiles in subjects collected at geographical locations with disparaging seasonality patterns.

1878T

Imputation Performance of ~4,000 genomes from the UK10K Project. *J. Huang[1], B. Howie[2], J. Marchini[3] on behalf of the UK10K Cohorts Consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Department of Human Genetics, University of Chicago, Chicago, Illinois, USA; 3) Department of Statistics, University of Oxford, Oxford, UK.

Imputation from reference haplotypes has proved instrumental to the success of genome-wide association studies. Imputation based on whole-genome sequencing (WGS) data is expected to further increase the utility of this approach to variants of lower minor allele frequency (MAF). To date, the 1000 Genomes project has made publically available data for a set of 1,092 genomes (4X) for individuals of multiple ethnicities. The UK10K Cohorts project has now generated a dataset of 3,781 genomes (6X), aiming to exhaustively characterize genetic variation down to 0.1% MAF in the British population.

We compared the UK10K reference panel with the 1000 genomes panel alone and the two combined. The combination of two WGS reference panels was carried out using a new functionality in IMPUTE2. We generated two pseudo-Illumina 610k bead chip datasets as imputation targets by sub-sampling WGS data from: (1) 10% samples of the UK10K project; (2) four European samples from Complete Genomics (80X). The imputation quality was measured by Pearson correlation coefficient (r^2) between the imputed genotype dosage and the masked sequences. For each of the above six scenarios (three reference panels, two SNP-arrays), we run two further analyses: (1) re-phasing the WGS data using a new algorithm implemented in SHAPEIT v2; (2) varying the -k_hap parameter in IMPUTE2 that controls the number of haplotypes used to impute each sample (500 vs. all haplotypes).

Using 3,781 UK10K genomes significantly improved imputation quality for variants with MAF between 0.1% and 5%. Increasing the number of haplotypes sampled and re-phasing the WGS reference panel provided further improvement. Adding 1000 Genomes data to UK10K data increased the number of variants imputed but not imputation accuracy. For variants with MAF of 0.1%, 0.5%, 1%, 2%, 5% respectively, the mean r^2 for using the optimal approach (two reference panels, rephased, all haplotypes sampled) is 0.531, 0.634, 0.709, 0.771, 0.856, compared to 0.253, 0.338, 0.511, 0.634, 0.780 by using the standard approach (1000 Genomes data alone, phased as-is, 500 haplotypes sampled).

In summary, we demonstrated improved imputation for European samples by using UK10K data and a new strategy to re-phasing and combining WGS reference panels.

1879F

Next generation sequencing and its application in clinical practice. *E. Klee[1], P.A. Decker[1], N. Jia[1], S.K. McDonnell[1], S.N. Thibodeau[2], M. de Andrade[1]*. 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Next generation sequencing technologies have rapidly advanced to the point that patient-specific whole-genome sequencing may soon be as routine as X-rays and cholesterol testing. However, the challenge in interpreting the vast amount of data generated by genomic sequencing and effectively using it to guide decisions about an individual's health care is significant. As part of Mayo Clinic's Center of Individualized Medicine we are developing a pipeline to analyze whole exome sequencing (WES) of small to large pedigrees and provide insights in genetic disease. In this abstract we focus on the challenge of SNV/SNP selection necessary to perform identity by descent (IBD) analysis for validating the correct relatedness between relatives. For SNV selection we selected samples with genotype data from both WES and SNP chip arrays to select sets of informative markers based on minor allele frequency and linkage disequilibrium. We used the maximum likelihood method available in PLINK and the maximum likelihood method available in PREST for IBD analysis and applied, them to pedigrees of different size. To determine possible bias in the computed IBD, we varied the number of related subjects, the type of relationship and number of variants used in the analysis. We observed that the IBD calculation in PREST is not as susceptible to biased results for close relationship as in PLINK , and including a sample of unrelated subjects helps stabilize the IBD calculation in PLINK and PREST.

1880W

A Weighted U statistic for Genetic Association Analyses of Sequencing Data. *C. Wei[1], M. Li[2], Z. He[3], Q. Lu[1]*. 1) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 2) Division of Biostatistics, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas; 3) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan.

Despite the recent success of genome-wide association studies, a large proportion of genetic variants predisposing to complex diseases remain uncovered. Evidence from genetic studies and evolutionary theory has suggested rare variants could play an important role in the biological pathways of complex diseases. The advance of next generation sequencing technology facilitate the generation of massive amount of genetic variants and offers great opportunity to investigate the role of millions rare variants in the genetic etiology of complex disease. Nevertheless, great challenge has also been posed to statistical analyses of high-dimensional sequencing data. The association analyses based on traditional statistical methods endure substantial power loss because of low frequency of genetic variants and extremely high dimensionality of the data. We therefore developed a weighted U statistic for high-dimensional association analysis of next-generation sequencing data. Based on the non-parametric U statistic, our method makes no assumption of the underlying disease model and can be applied to various types of phenotypes (e.g., binary and continuous phenotypes). Through simulation studies and an empirical study, we found our method outperformed a commonly used SKAT method when the underlying assumption is violated (e.g., the phenotype follows a heavily skewed distribution) and attained comparable performance to SKAT when underlying assumption is satisfied. In an empirical study of Dallas Heart Study (DHS) sequencing data, our method was also able to detect the association of ANGPTL 4 with very low density lipoprotein cholesterol.

1881T

Accurate Genotype Calling with Contaminated Sequencing Data. *M. Flickinger, G. Jun, G.R. Abecasis, M. Boehnke, H.M. Kang.* Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI.

Advances in next-generation sequencing have enabled a wide range of large-scale genetic studies. While the quality of the sequence data is generally improving, protocols are not perfect and inevitably sampling handling errors may occur. One common error is sample contamination through the mixing of two or more samples. We previously developed methods to detect contaminated samples (Jun et al. AJHG 91:839-848, 2013), so that these could be flagged and dropped from analysis. Here we describe new mixture-model-based methods to account for contamination in the genotype likelihoods which result in genotype calls for contaminated samples with accuracy approaching that for uncontaminated samples. To explore the operating characteristics of our method, we simulated contamination in silico by mixing reads from pairs of 198 European 1000 Genomes Project BAM files for whom both high-depth (50-150x) exome and low-depth (4-6x) whole genome sequencing data were available. We then compared the resulting calls to the available Omni 2.5M and Exomechip genotype arrays to evaluate success. In our first experiment, we constructed samples with an 85:15 mix corresponding to 15% contamination. In the high-depth exome data, the overall genotype discordance with the Exomechip decreased from 3.6% to 0.9% by modeling the contamination, compared to 0.3% for datasets with no contamination. For the low-depth data, we used our method to model contamination in the genotype likelihoods and then Beagle for linkage-disequilibrium-aware calling. Here we saw a drop in discordance with the Omni array data from 9.3% to 3.5% compared to 1.1% with no contamination. Furthermore, the false positive and false negative rate (% of incorrect and missed non-reference alleles) were also reduced by >50% for both high- and low-depth data. When the identity of the contaminating sample is known, our approach can incorporate this information resulting in further improvement: reducing discordance for high-depth data to 0.7%, approaching the 0.3% discordance observed when analyzing uncontaminated samples. In conclusion, our methods provide an effective alternative to expensive re-sequencing for accurately genotyping low to moderately contaminated samples.

1882F

A Rare Variant Selection Algorithm to Locate Susceptible Rare Variant from Sequencing Data. *S. Wang, H. Sun*. Dept Biostatistics, Columbia Univ, New York, NY.

Current association methods for sequencing data have been focused on aggregating rare variants across a gene or a genetic region due to the fact that analyzing individual rare variants is underpowered. To identify which rare variants in a gene or a genetic region out of all variants are associated with the outcomes is a natural next step. Here we propose a forward variable selection-based algorithm that is able to identify the locations of potentially susceptible rare variants that are associated with the outcomes with sequencing data. More specifically, we test associations of variants individually in the first step to select the most outcome-related variant that has the maximum test statistic. In step 2, weighted linear combinations of two variants with the selected variant from Step 1 and all other variants one by one are generated and tested to select the most outcome-related two-variant combination. Note that the most outcome-related variant selected in Step 1 could be either risk or protective. In Step 2, we also consider the combination of potentially protective variants through flipping of the variant coding. Similar to a forward variable selection, we keep adding one variant in each step until the test statistic of the linear combination of the combined variants is no longer increasing, and the combined variants are consider as the best combination of potentially causal variants. We evaluated the selection performance of the proposed algorithm through simulation studies where different effect sizes, sample sizes and directions of the effects of the individual rare variants were considered. The results demonstrated that the proposed algorithm is able to select subsets with most of the outcome related rare variants in all simulation scenarios considered. The proposed method was also applied to sequence data on the ANGPTL gene family from the Dallas Heart Study (DHS). The proposed method was implemented in an R package RVsel which will be freely downloaded at http://www.columbia.edu/~sw2206.

1883W

Functional Linear Model with both Functional Response and Functional Predictors for Genetic Studies of Temporal Quantitative Trait with both GWAS and Next-Generation Sequencing Data. *D. Lee, C. Hanis, M. Xiong.* University of Texas School of Public Health, Houston, TX., US.

Traditional quantitative genetics has primarily studied traits that are measured at a specified location or time. The traits are investigated as isolated and static variables. However, in real biologic world, many quantitative traits change over time. These quantitative traits are repeatedly measured as functions of time or complete curves. In other words, these traits are observed either as continuous random functions, or on a dense grid. It is now well documented that next-generation sequencing (NGS) can generate several millions or even dozens of millions of genetic variation data. As a consequence, these genetic variation data are so densely distributed across the genome that the genetic variation can be modeled as a function of genomic location. In summary, random functions appear as either responses or predictors, or both. Although the current widely used statistical methods for QTL analysis are based on multivariate analysis, the standard multivariate statistical analysis often fails with functional data. To overcome the limitations of the traditional QTL analysis, we propose to use functional linear model in which a trait curve is modeled as a response function, the genetic variation in a genomic region or gene is modeled as a functional predictor, and the genetic effects are modeled as a function of both time and genomic position for genetic analysis of a temporal quantitative trait with both GWAS and NGS data. By intensive simulations, we demonstrate that the functional linear model for genetic studies of the temporal quantitative traits has the correct type 1 error rates and much higher power to detect association than the current methods. The proposed method is applied to Starr County Sleep Studies where oxygen saturation of 767 individuals with 795,736 SNPs were measured for 380 minutes on average.

1884T

Weighted combination of truncated P-values for rare causal variants detection. *W.Y. Lin[1], X.Y. Lou[2], G. Gao[3], N. Liu[2].* 1) Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan; 2) Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama; 3) Department of Biostatistics, Virginia Commonwealth University, Richmond, Virginia.

With the development of next-generation sequencing technology, there is a great demand for powerful methods to detect rare causal variants (minor allele frequencies (MAFs) < 1%). Testing for each variant site individually is known to be underpowered, and many methods have been proposed to test for the association of a chromosomal region containing multiple variant sites with phenotypes. However, this pooling strategy inevitably leads to the inclusion of a large proportion of neutral variants, which may compromise power. Here, we have proposed a T-sigma-MidP method that combines per-site P-values with weights based on MAFs. Before combination, we first imposed a truncation threshold upon the per-site P-values to guard against noise caused by the inclusion of neutral variants. Simulations have shown that T-sigma-MidP outperforms competing tests in the majority of simulation settings. We applied T-sigma-MidP to the Dallas Heart Study data set, and obtained more significant results in testing for the association of the ANGPTL4 gene with triglyceride and very-low-density lipoprotein. The proposed T-sigma-MidP method is recommended for the easiness of its implementation, its satisfactory power to detect rare causal variants (population MAF < 1%) and/or uncommon causal variants (1% < population MAF < 3%), and its robustness to noise from neutral variants.

1885F

Detecting Genomic Clustering of Risk Variants from Sequence Data: Cases vs. Controls. *JP. Sinnwell[1], DJ. Schaid[1], SK. McDonnell[1], SN. Thibodeau[2].* 1) Dept Biostatistics, Mayo Clinic, Rochester, MN; 2) Dept. Lab Med Path, Mayo Clinic, Rochester, MN.

As the ability to measure dense genetic markers approaches the limit of the DNA sequence itself, taking advantage of possible clustering of genetic variants in, and around, a gene would benefit genetic association analyses, and likely provide biological insights. The greatest benefit might be realized when multiple rare variants cluster in a functional region. Several statistical tests have been developed, one of which is based on the popular Kulldorff scan statistic for spatial clustering of disease. We extended another popular spatial clustering method - Tango's statistic - to genomic sequence data. An advantage of Tango's method is that it is rapid to compute, and when single test statistic is computed, its distribution is well approximated by a scaled chi-square distribution, making computation of p-values very rapid. We compared the Type-I error rates and power of several clustering statistics, as well as the omnibus sequence kernel association test (SKAT). Our results showed that the Ionita-Laza version of Kulldorff's scan statistic had the greatest power over a range of clustering scenarios. Although our version of Tango's statistic, which we call 'Kernel Distance' statistic, had slightly less power than the scan statistic, its rapid computation time makes it an appealing competitor to the scan statistic.

1886W

GEE-based kernel association test in family based sequencing studies. *X. WANG[1,2], S. Lee[1], T. Cai[1], T.W. Yu[3], C.A. Walsh[3], X. Lin[1].* 1) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 2) Graduate Program in Public Health, Stony Brook University, Stony Brook, NY; 3) Division of Genetics, Children's Hospital Boston, Boston, MA.

Family-based designs in sequencing studies provide opportunities to detect genetic variants that complement studies of unrelated individuals. By exploiting co-segregation and Mendelian inheritance information, family based designs are more efficient at prioritizing casual variants, and which can improve genotyping error check and imputation of variants in unsequenced samples. In this study, we consider testing the effect of a SNP/SNV set in family studies. Specifically, we propose a Generalized Estimating Equations (GEE)-based kernel association test, a variance component-based testing method, to test for the association between a phenotype and multiple variants in a SNV set jointly using family samples. The proposed approach allows for both continuous and discrete traits, where the correlation among family members is taken into account through the use of an empirical covariance estimator. We derive the theoretical distribution of the proposed statistic under the null and develop analytical methods to calculate the p-values. We also propose an efficient resampling method for correcting for small sample size bias in family studies. The proposed method allows for easily incorporating covariates and SNP-SNP interactions. Simulation studies show that the proposed method properly controls for type-I error rates under both random and ascertained sampling schemes in family studies. We demonstrate through simulation studies that our approach has superior performance for association mapping compared to the single marker based minimum p-value GEE test for a SNP set effect over a wide range of scenarios. We illustrate the application of the proposed method using data from an exome sequencing study of autism.

1887T

Comparisons of the power functions and asymptotic relative efficiencies of empirical genetic relationship kernels in heritability and association testing. *V.P. Diego[1], M. Almeida[1], J. Peralta[1,2], T.D. Dyer[1], J.W. Kent, Jr.[1], J.T. Williams[1], L. Almasy[1], H.H.H. Göring[1], R. Duggirala[1], J. Blangero[1].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA; 2) Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, WA, Australia.

Empirical genetic relationship kernels (GRKs) have spurred on the development of powerful statistical genetic methods. Using the genome partitioning method, empirical GRKs can be used to estimate heritabilities based on the set of genes underlying specific metabolic pathways. Under our novel eigensimplification approach to likelihood analysis of variance component (VC) models, we have shown in theory how empirical GRKs can be used for rapid and efficient association analysis in samples of large extended families, where the speed of the analysis is commensurate with that for the same sample size of unrelated individuals. However, there are several competing GRK algorithms in the literature, and it is not obvious which of these is optimal for a given study design. To address this issue we studied the statistical properties of GRKs estimated over sequence data for 959 individuals from Genetic Analysis Workshop 18 by six well-known algorithms in relation to heritability and association testing. Named after their software packages, the six GRK algorithms are PLINK, GCTA, KING-homogeneous, KING-robust, REAP, and LDAK. The eigensimplification approach, which is based on a spectral representation of the likelihood, gives elegant expressions for the expected likelihood ratio test statistic (*ELRT*). For heritability, the *ELRT* is: $-\Sigma \ln[1+h^2(\lambda_{gi}-1)]$, where h^2 and λ_{gi} are the heritability and additive genetic eigenvalues, and the summation is taken over n individuals. For association, we have: $-n\ln(1-h^2_q)+\Sigma\ln[1+h^2_t(\lambda_{gi}-1)]-\Sigma\ln[1+h^2_r(\lambda_{gi}-1)]=$ *ELRT* [h^2_q>0:unrelateds]-(*ELRT* [h^2_t>0:pedigrees]-*ELRT* [h^2_r>0:pedigrees]), where h^2_t, h^2_q and h^2_r are the total, quantitative trait locus and residual heritabilities, respectively. Power analyses are made relatively easy because the *ELRT* well approximates the noncentrality parameter (NCP) in likelihood-based power functions. Moreover, precisely because the *ELRT* measures the NCP it can be used for asymptotic relative efficiency (ARE) analysis using the Pitman ARE. We discuss our results in terms of the observed relationships between study design and GRK. We also found that since certain GRK algorithms may produce GRKs that are not positive semidefinite (PSD), it is important to exercise care in this regard. If a GRK is non-PSD, then procedures to either constrain such kernels to be PSD or to correct them in the sense of making them PSD must be employed.

1888F

A DNA variant caller adapted to assess mitochondrial DNA variation in lymphocytes from 2,000 Sardinians. *J. Ding[1], C. Sidore[2,3], O. Meirelles[1], M.K. Wing[2], F. Busonero[3], R. Nagaraja[1], F. Cucca[3], G.R. Abecasis[2], D. Schlessinger[1].* 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

The degree to which mitochondrial DNA (mtDNA) varies heritably and somatically has been much discussed, but has not been systematically analyzed on a population basis. Further, accumulation of mutations in mitochondrial DNA has been suggested to play an important role in aging. To approach such an analysis of mtDNA variation, genotype calling and analytic programs developed for nuclear DNA must be modified, because each cell has 100-10,000 mtDNA copies that can vary at any site (i.e., heteroplasmy). We have developed an algorithm that is adapted to identify variants in mtDNA; it incorporates the sequencing error rate at each base in likelihood calculations and is flexible to allow for different allele fractions at a variant site across individuals. Our analysis procedure is also adapted to the circular mitochondrial genome, a key difference from the linear chromosomes assumed by most read mapping algorithms.

We assess homoplasmies and heteroplasmies in mtDNA sequences of lymphocytes from ~2,000 SardiNIA Project participants. The distribution of the number of homoplasmies per individual is bimodal, with an average value of 22. The number of heteroplasmies with a minor allele fraction threshold of 4% varies considerably among individuals, with most individuals displaying 3 or less heteroplasmies, but some showing considerably more. The overall heteroplasmy increases with age, but the slope is small, yielding an average increase of ~1 heteroplasmy between ages 20 and 80. As expected, mothers and their children share essentially all homoplasmies but a lesser proportion of heteroplasmies. Overall, it is clear that mitochondrial variants do accumulate in lymphocyte populations as a function of age, but the extent of accumulation remains hard to assess accurately because many variants are on par with sequencing error rate. Currently, attempts are being made to increase resolving power by using higher sequence coverage and by studying individuals from families.

The results to date provide information about mtDNA haplogroups and the inheritance of homo- and heteroplasmies in Sardinia. The algorithm can be further extended in several ways: for example, to study mtDNA from cloned normal cells in greater depth, and to investigate the nuclear DNA variability in cancer cells.

1889W

Understanding the limits of pooled next-generation sequencing to identify causal modifier variants in Cystic Fibrosis. *J. Gong[1], F. Lin[2], T. Chiang[1], K. Keenan[3], M. Miller[1], D. Soave[1,4], W. Li[1,4], L. Sun[4,5], J. Rommens[2,6], L. Strug[1,4].* 1) Program in Child Health Evaluative Sciences, the Hospital for Sick Children, Toronto, ON, Canada; 2) Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, Ontario, Canada; 3) Program in Physiology and Experimental Medicine, the Hospital for Sick Children, Toronto, Ontario, Canada; 4) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 5) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 6) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

Modifier genes contribute to variability in morbidity across multiple organs affected in cystic fibrosis (CF). SLC26A9, SLC9A3 and SLC6A14, were shown to contribute to CF intestinal obstruction (MI) in a hypothesis driven GWAS (Sun et al, (2012) Nat Genet 44:562-9). PCR-enriched overlapping segments of the three genes from 200 CF individuals, 100 with MI (cases), were sequenced using pooled next-generation methods (pNGS) by deCODE genetics. We sought to (1) identify variants for follow-up genotyping in a larger sample; and (2) determine whether pNGS can provide reliable association findings. We performed quality control on the FastQ reads followed by read alignment, variant call and filtration. For each variant, we estimated minor allele frequency (MAF) for cases and controls, using a pNGS-specific method (Wang et al, (2010) Genet Epidemiol, 34:492-501). We compared these estimates to those from genotype data on the same set of individuals. We then compared three association methods for pNGS data: simulation-based method (Wang et al, ibid), Fisher's exact test, and a permutation approach. Finally, we performed Sanger sequencing (SS) in one region of SLC6A14 selected by the association results to assess the reliability of our findings. For 54 common SNPs in both the pNGS and array-based genotype data, we found a high correlation between the estimated MAFs (ρ=0.95 for cases and 0.93 for controls). Comparison of the association methods showed no consistent rankings across the variants, although one region in intron 6 of SLC6A14 was highlighted by all three methods. Some of the sequencing variants in this region were in high LD with imputed SNPs, but in general this region was poorly imputed. pNGS identified all the variants detected by SS in this region. However, the local MAF estimates from both pNGS and imputation differed substantially from that provided by SS, resulting in discrepant association interpretation. Our experience suggested that pNGS is reliable for variant detection. Its estimation of MAF was reasonable on average, but a sub-gene region was observed where the MAF estimates (and thus association results) differed substantially from values obtained from SS. The discrepancies were functions of region-specific qualities of the sequence data and genetic architecture, neither of which can be known a priori. There was no general agreement among the tested pNGS association methods, limiting the usage of pNGS data for reliable association testing.

1890T

Accurate local ancestry inference in exome sequenced admixed individuals using off target sequence reads. *Y. Hu[1,2], C. Willer[3], X. Zhan[2], H. Kang[2], G. Abecasis[2].* 1) Integrative Biology, University of California, Berkeley, CA, 94720, US; 2) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109, US; 3) Department of Human Genetics, University of Michigan, School of Medicine, Ann Arbor, MI, 48109, US.

Estimates of the ancestry of specific chromosomal regions in admixed individuals are useful for studies of human evolutionary history and in disease gene mapping. Previously, this ancestry inference relied on high quality genotypes from genome-wide association study (GWAS) arrays. These high quality genotypes are not always available when samples are exome-sequenced, the strategy of choice for many ongoing genetic studies. Here we show that off-target reads generated during exome sequencing experiments can be combined with on-target reads to accurately estimate the ancestry of each chromosomal segment in an admixed individual. To reconstruct local ancestry, our method SEQMIX models aligned bases directly instead of relying on hard genotype calls. We evaluate the accuracy of our method through simulations and analysis of samples sequenced by the 1,000 Genomes Project and the NHLBI Grand Opportunity Exome Sequencing Project. In African-Americans, we show local ancestry estimates derived using our method are very similar to those derived using Illumina's Omni 2.5M genotyping array, and much improved in relation to estimates that use only exome genotypes and ignore off-target sequencing reads. Software implementing this method, SEQMIX, can be applied to analysis of human population history or for disease gene mapping studies in admixed individuals.

1891F

Flexible and robust methods for rare-variant testing of quantitative traits in pedigrees. *Y. Jiang[1,2], K.N. Conneely[1,2], M.P. Epstein[1,2].* 1) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA.

Rare-variant sequencing studies are increasingly popular strategies for investigating the missing heritability of complex human traits. Although a series of statistical methods have been developed to analyze rare variants, the majority of tests are restricted to case-control and population-based study designs. Few statistical methods have been developed to analyze rare variants in family-based studies. As more and more resequencing studies employ family-based designs to overcome potential bias caused by population stratification and to study cosegregation patterns of causal variants, the development of powerful methods for family-based analyses of rare variants is needed. With this in mind, we propose a rare-variant association test for quantitative traits in families that uses a kernel framework. Within a region of interest, the model partitions a family member's genotype at a rare variant into a within-family component (robust to population stratification but sensitive to genotype error) and an orthogonal between-family component (sensitive to population stratification but less sensitive to genotype error). For large-scale sequence data, our approach first constructs a kernel test using the between-family component as a screening tool to identify top hits. We then follow up these top hits using a kernel test based on the robust and independent within-family component. Unlike other methods, which assume that all variants in the tested region have the same direction of effect on the phenotype, our method remains powerful when causal variants in the test region may have differing direction of effect. Our method can also accommodate covariates and is flexible in that it can still perform well under partial missingness of parental genotypes. Finally, our method has the practical benefit of permitting efficient calculation of p-values based on asymptotics rather than requiring computationally-expensive permutations, which enables practical application to genome-wide data. Using simulated data, we have already shown that our method can avoid inflated false positive rates caused by population stratification. We will also use simulated data to explore power and further apply our method to the Sardinia sequencing dataset available through dbGaP for illustration purposes.

1892W

Study designs for next-generation sequencing: some empirical evidence. *P. Marjoram, D.G. Conti, M. Salomon, D.C. Thomas.* Dept Preventive Med, Univ Southern California, Los Angeles, CA.

GWAS aims to find SNPs associated with phenotypes of interest. Since detected SNPs are likely to be in linkage disequilibrium with the truly causal SNP, rather than being causal themselves, a common follow-up study will be to sequence an area around one or more regions containing such SNPs. Multiple non-trivial questions arise: What regions should we chose? How wide should those regions be? How deep should the coverage be? What study design should we use: two-phase case-control (Schaid et al., Genet Epidemiol 2013;37: 229-38.), family-based (Xu & Zhi, AJHG 2012;90:1028-45), etc.? How do we best analyze the resulting data? We will describe our efforts as part of NIH's "GWASeq" consortium, which aims to provide the empirical data that will allow the community to answer these questions. Our component of the consortium focuses on data from the Colorectal Cancer Family Registry (C-CFR) (Newcomb et al., CEBP 2007;16: 2331-43). We have sequenced around 4000 samples from the C-CFR, across 10 regions, at more than 50X coverage. Our data is a mixture of population-based and pedigree-based samples. We will describe the study, the data that results from it, and the conclusions that can be drawn regarding design of such studies in future and the implications regarding so-called "missing heritability.".

1893T

Are base qualities necessary in the context of high coverage sequencing? *M.W. Snyder, J. Shendure.* Genome Sciences, University of Washington, Seattle, WA.

Base quality scores from next-generation sequencing instruments are used in a variety of data processing steps. Some aligners use quality scores to guide read placement on a reference. During variant discovery and variant calling, probabilistic and likelihood-based models may make use of quality scores in genotype inference. Downstream, post-processing steps often include filtering variants on the basis of base quality score metrics. However, the storage and use of these quality scores has associated costs. First, intermediate file sizes are significantly increased: for reads in the FASTQ format, the stored base qualities nearly double the overall size of each file. These same qualities may then be propagated forward to SAM or BAM format, where file size is again inflated by the presence of base quality information. Additionally, the use of base qualities during likelihood calculations, read placement, or variant filtering imposes a computational penalty on data processing pipelines. Conventional wisdom suggests that these costs are justified by the concomitant increase in genotype accuracy to which they give rise. Here, we challenge the validity of this assumption in the context of very high coverage datasets generated on modern sequencing instruments. We attempt to use the counting statistics naturally arising from high coverage sequencing as surrogates for base qualities, and investigate the impact on genotype accuracy of processing sequencing data that has been stripped of base quality scores. To wit, we perform alignment, variant discovery, and genotyping without base quality scores, using a combination of conventional and custom software tools, and compare our results to variant calls from existing data analysis pipelines and gold standard genotypes. We apply our approach to a family trio (NA12877, NA12878, and NA12882) sequenced to 200X coverage on an Illumina HiSeq 2000 system and for whom gold standard genotypes are available from orthogonal methods. In preliminary investigations, we achieve approximately 99% non-reference genotype concordance with array-based genotypes. We downsample the original reads and evaluate genotype accuracy as a function of mean coverage. We additionally describe the cost savings in terms of reduced computational time and smaller storage footprint achieved through the application of this approach.

1894F

A UNIFIED SEQUENCE KERNEL ASSOCIATION TEST ALLOWING FOR ADMIXED SUBJECTS WITH ARBITRAY RELATEDNESS. *W. Ouyang[1,2], H. Deng[1,2], H. Qin[1,2].* 1) Department of Biostatistics and Bioinformatics Tulane University School of Public Health and Tropical Medicine.1440 Canal Street, New Orleans, LA 70112; 2) Center for Bioinformatics and Genomics, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street,New Orleans, LA 70112.

Rare genetic variants may potentially contribute to more missing heritability complex human diseases. Over the recent years, next generation sequencing studies (NGSs) have been conducted to provide more comprehensive and accurate description of rare and common variants, and a number of statistical methods have been developed for identifying sequence associations. Current prominent sequence association methods assume homogeneity and/or un-relatedness of study subjects. Such assumptions could be severely violated in the samples from admixed populations, e.g., African Americans and Hispanic Americans. Population stratification and cryptic relatedness are two important confounders in genetic association studies in admixed populations. In this paper, we have developed a unified sequence association test (AdmSKAT) to allow for the NGS data of admixed subjects with arbitrary pedigree structure and diverse directions of effect of genetic variants on phenotype. Our AdmSKAT approach extends the conventional population based sequence kernel association test (SKAT) and family-based SKAT by jointly modeling population stratification, relatedness and local ancestry-by-genotype interaction to ensure type I error control and to improve statistical power. We simulated extensive sequence data of admixed families for method comparisons. Under the null of no sequence association, our AdmSKAT and FamSKAT controlled type I error rates, whereas naïve application of the conventional SKAT and FLM lead to inflated type I error rates. Our AdmSKAT were more powerful than the FamSKAT when applied to all the simulated scenarios of admixed families. AdmSKAT has higher power than competing method in many different scenarios.We then apply our AdmSKAT to analyze the real and simulated sequence data on hypertension of Mexicans from GAW18.

1895W

Phasing and imputation of 12,000 1x coverage whole genome sequenced Chinese women. *W. Kretzschmar[1], J. Flint[1], J. Marchini[2], CONVERGE Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Statistics, University of Oxford, Oxford, United Kingdom.

The CONVERGE study of Major Depression has collected low-coverage (1x) sequencing data on 12,000 Chinese women. This study is one of the largest whole-genome sequencing studies currently underway. To accurately infer genotypes in these samples, LD-based genotype refinement methods are needed. However, the study size makes this a challenging proposition. We have investigated several computationally tractable strategies for calling genotypes. Current MCMC schemes for phasing and genotype calling do not explicitly encapsulate the local IBD structure between individuals. We have investigated new adaptive MCMC schemes to sample haplotypes and genotypes that underlie each sample that try to learn details about local haplotype sharing as the MCMC sampler evolves. We have also investigated whether current methods can be made to produce acceptable results within practical time frames.

We have applied these methods to genotype likelihoods at ~14 million 1000 Genomes Project Phase 1 (TGPP1) SNPs polymorphic in TGPP1 Asians. We were able to infer genotypes in the whole CONVERGE cohort at a computational cost equivalent to ~2,000 CPU days on a single core of a current Intel processor. We have measured the accuracy of our imputed genotypes using external validation genotypes on 16 individuals typed on the genome-wide Illumina Zhonghua-8 SNP chip with 890,371 SNPs. We find that the mean imputation r^2 at MAFs <0.5%, 0.5-1%, 1-2%, 2-5%, and >5% are 0.21, 0.57, 0.70, 0.82, and 0.92. Overall, these results illustrate the feasibility and accuracy of using very low-coverage sequencing in large studies.

1896T

Construction of an accurate haplotype reference panel that incorporates multi-allelic variants from sequencing data. *A. Menelaou[1], S.L. Pulit[1], L.C. Francioli[1], J. Marchini[2], P.I.W. de Bakker[1], Genome of the Netherlands consortium.* 1) Department of Biomedical Genetics, University Medical Centre Utrecht, Utrecht, Netherlands; 2) Department of Statistics, University of Oxford, United Kingdom.

Next-generation sequencing allows in-depth characterization of human polymorphism. Although imputation from sequencing panels can increase the effective coverage of GWAS, the imputation quality (and therefore power to discover novel loci) is dependent on accurately phased haplotypes, especially for lower-frequency variation. Furthermore, as the focus shifts increasingly beyond common biallelic SNPs, it is also necessary to incorporate multi-allelic variants into such panels.

To investigate the added value of sequencing trios for the construction of an imputation panel, we analyzed the Genome of the Netherlands (GoNL) Project data, which includes 250 parent-offspring trios sequenced at 14x using Illumina HiSeq. We constructed the haplotypes using MVNcall, a genotype calling and phasing method that accounts jointly for LD structure, sequencing information, and family relationships. We evaluated the imputation quality in 81 independent samples with Dutch ancestry sequenced with Complete Genomics (CG), where we obtained an aggregate imputation accuracy (r^2) of 0.63 for SNPs with frequency <0.5%. For those SNPs, we observed a 16% increase in accuracy when we phased the trios (0.63) jointly versus phasing only the (unrelated) parents (0.47), highlighting the impact of trio-based phasing. In the same Dutch samples, imputation accuracy for lower frequency variants <10% was increased for GoNL as compared to 1000 Genomes Phase 1 (1KG) panel.In independent CEU and TSI samples from 1KG with CG data, imputation quality with GoNL was better than that with 1KG-CEU and equivalent to that with 1KG-TSI, at the same sample size. A combined GoNL+1KG panel achieves the highest imputation accuracy (0.70 for SNPs with frequency <0.5%), motivating the creation of a worldwide sequencing-based panel.In parallel, we have extended MVNcall to handle imputation of multi-allelic variants including SNPs, indels, structural variants and microsatellites. Because MVNcall determines phase for each site independently onto an accurate pre-phased haplotype scaffold, there are several advantages to other existing methods. First, the imputation of nearby sites is not affected by the presence of complex variants (that are inherently more difficult to impute). Second, there is a considerable computational efficiency gain by considering all variants independently. We have applied MVNcall in STRs detected by 1KG Phase 3, and multi-allelic SNPs and indels discovered in GoNL.

1897F

Modeling gene expression and rare sequence variation to identify genes and subnetworks underlying autism risk. *L. Liu[1], J. Lei[1], S. Sanders[2], J. Willsey[2], Y. Kou[3,4], L. Klei[5], C. Lu[1], X. He[6], A. Ma'ayan[4], J. Noonan[2], N. Sestan[7], J. Buxbaum[3,4,8,9], M. State[2,10,11,12], B. Devlin[5], K. Roeder[1,6].* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; 2) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; 3) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 4) Department of Psychiatry, Icahn School of Medicine at Mount Sinai , New York, New York, USA; 5) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA; 6) Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University,Pittsburgh, Pennsylvania, USA; 7) Department of Neurobiology and Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06510, USA; 8) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 9) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 10) Program on Neurogenetics, Yale University School of Medicine, New Haven, Connecticut, USA; 11) Child Study Center, Yale University School of Medicine, New Haven, USA; 12) Department of Psychiatry, Yale University School of Medicine, New Haven, USA.

Rare variants, especially de novo loss of function (LoF) mutations, have proven effective at identifying genes affecting risk for autism spectrum disorders (ASD). When a de novo LoF mutation falls in a gene, and more than one ASD proband carries such a mutation, the gene is likely to be a risk gene. Based on the rate of de novo mutations in ASD probands versus their siblings, we infer that roughly half of the genes with LoF mutations observed only once per gene in the combined sample affect risk. To extract more information from Whole Exome Sequence (WES) data, He et al. develop a statistical model that integrates data from family and case-control studies to infer the likelihood a gene affects risk for ASD; this model, called TADA for Transmission And De novo Association, summarizes genetic evidence a gene affects risk for ASD. Still, given the limited WES data, can we garner more information regarding ASD risk? Relative to genes with strong genetic support for involvement in ASD, we hypothesize that genes expressed at the same developmental period and region of the brain, and with highly correlated expression, would themselves be more likely to affect risk for ASD because they are functionally interrelated. To find these sub-networks of co-expressed genes we jointly model two kinds of data: gene co-expression in specific brain regions and periods of development; and the TADA analysis of rare variants from recent sequencing studies; both from published studies. We model the ensemble data as a Hidden Markov Random Field (HMRF), with the graph structure determined by gene co-expression. An HMRF model is a natural framework for the interrelationships amongst nodes of a network and to combine such interrelationships with node-specific observations: gene identity, its expression, its associated genetic data, and whether it affects risk for ASD (yes/no), which will be estimated. The HMRF models the dependency in the network, which is inherent in the correlation of a gene's state, ASD risk gene or not, with its near neighbors states. The statistical analysis identifies over a hundred genes that plausibly affect risk, a large fraction of which are not known to affect risk for ASD. For those that are not novel, many were implicated on the basis of the statistical model, not the strength of the genetic evidence alone. A third of the genes with a single LoF de novo mutation were implicated. These results expand our understanding of the neurobiology of ASD.

1898W

Gene-based Generalized Functional Linear Models for Case-Control Association Studies. *R. Fan[1], Y. Wang[1], L. Lobach[2].* 1) Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD 20852; 2) he Memory and Aging Center, Uunversity of California, San Francisco, CA 94158.

Functional generalized linear models are developed for testing associations between dichotomous traits and multiple genetic variants in a genetic region. Although the observed genetic marker data are discrete, we view them as realizations of continuous genetic variant functions. We believe that the genetic variant functions have intrinsic functional structure. By using modern state-of-art functional data analysis technique, the observed high dimension genetic variant data are used to estimate the genetic variant functions based on B-spline or Fourier basis functions or functional principal component decompositions. Then, the estimated genetic variant functions are used in the logistic models to connect to phenotype adjusted for covariates. Both fixed and mixed effect functional generalized linear models are built to test the association between dichotomous traits and genetic variants adjusting for covariates. After extensive simulation analysis, it is shown that the Rao's efficient score tests or global score tests of the proposed fixed effect models have similar power as optimal sequence kernel association test (SKAT-O) for two scenarios: (1) the causal variants are all rare; (2) the causal variants are both rare and common. The proposed test statistics generate accurate type I errors in simulation studies. The methods can be used in both gene-based genome-wide/exome-wide association studies or candidate gene analysis.

1899T

Functional Linear Models for Association Analysis of Quantitative Traits. *Y. Wang[1], R. Fan[1], J.L. Mills[2], A.F. Wilson[3], J.E. Bailey-Wilson[3], M.M. Xiong[4].* 1) Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD; 2) Epidemiology Branch, Division of Epidemiology, Statistics and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD; 3) Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Human Genetics Center, University of Texas - Houston \ P.O. Box 20334, Houston, TX.

Functional linear models are developed in this paper for testing associations between quantitative traits and genetic variants, which can be rare variants or common variants or the combination of the two. By treating multiple genetic variants of an individual in a human population as a realization of a stochastic process, the genome of an individual in a chromosome region is a continuum of sequence data rather than discrete observations. The genome of an individual is viewed as a stochastic function which contains both linkage and linkage disequilibrium (LD) information of the genetic markers. By using techniques of functional data analysis, both fixed and mixed effect functional linear models are built to test the association between quantitative traits and genetic variants adjusting for covariates. After extensive simulation analysis, it is shown that the F-distributed tests of the proposed fixed effect models have higher power than that of sequence kernel association test (SKAT) for three scenarios: (1) the causal variants are all rare; (2) the causal variants are both rare and common; and (3) the causal variants are common. The superior performance of the fixed effect functional linear models is most likely due to its optimal utilization both genetic linkage and LD information of multiple genetic variants in a genome and similarity among different individuals, while SKAT only models the similarities and pairwise LD but does not model linkage and higher order LD information sufficiently. In addition, the proposed fixed effect models generate accurate type I errors in simulation studies. We also show that the functional kernel score tests of mixed effect models are preferable in candidate gene analysis and small sample problems. The methods are applied to analyze three biochemical traits in data from the Trinity Students Study.

1900F

Pooled Whole Exome Sequencing as a Strategy for Establishing Population Mutation Profiles. *E. Rinella[1], K. Upadhyay[2], C. Oddoux[3], H. Ostrer[2].* 1) New York University Medical Center, New York, NY; 2) Albert Einstein College of Medicine, Bronx, NY; 3) Montefiore Medical Center, Bronx, NY.

While the cost of next-generation sequencing continues to decline, it can still be prohibitive to conduct large cohort studies. One strategy that we propose is to pool individual samples prior to library preparation thereby significantly reducing per sample sequencing costs. We created a pool from 200 Ashkenazi Jewish (AJ) women and submitted two identical aliquots of the pool for whole exome sequencing. Preliminary analysis of the two 200-Pool aliquots shows an average read depth (RD) of 122x and 116x for 200-PoolA and 200-PoolB, respectively, yielding a total of 59,010 on-target variants. The allele counts were combined for the 2 aliquots and frequencies were compared to array data for 1574 SNPs previously genotyped for the individuals in this pool, revealing an accuracy of r2=0.90 at an on-target RD <100x and r2=0.98 at RD ≥100x. The 59,010 variants were also filtered against the Online Mendelian Inheritance in Man (OMIM) database yielding 278 (0.47%) mutations, of which 61 of the OMIM variants were classified as pathogenic mutations. Among the mutations previously described in this population were those for hereditary hemochromatosis (HFE p.H63D), Factor V Leiden (F5 p.R534Q), and Canavan disease (ASPA p.Y231X). Among the newly identified conditions for which mutations were identified were autosomal recessive hypoparathyroidism (PTH p.R83X), Fletcher factor deficiency (KLKB1 p.S143N), autosomal dominant deafness (MYO1A p.G662E), Warfarin sensitivity (CYP2C9 p.I359L), trimethylaminuria (FMO3 p.V257M), hyperglycinuria (SLC6A20 p.T199M), atransferrinemia (TF p.P589S), and autosomal recessive renal tubular acidosis (ATP6V0A4 p.M580T). Pooled exome sequencing provides a cost effective method for establishing population mutation profiles, such as the profile developed here for the Ashkenazi Jewish population, and if linear will be an important reference for specific disease association studies.

1901W

A hidden Markov random field approach to modeling genetic association based on graphs derived from gene expression. *J. Lei[1], L. Liu[1], B. Devlin[2], K. Roeder[1].* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA;; 2) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA;.

Gene-based association testing problems involve tens of thousands of hypothesis tests and it is well known that traditional family-wise error rate control will be too conservative for large multiple testing problems. Standard false discovery rate (FDR) control procedure improves power, but because this approach focuses solely on the p-values of individual hypotheses, this approach too can have suboptimal power by ignoring correlations between individual hypotheses. When substantial prior knowledge is available, however, power can be increased if the prior information and the data can be modeled appropriately. We propose a hierarchical hidden Markov random field (HMRF) model for gene-based association using a graph built from gene expression data. The prior information from the network can be applied to improve the power and reduce false discovery under the realistic assumption that a test for a given gene should depend on data from neighboring genes in the network. The parameters of the HMRF model determine the spatial clustering of the gene-based association statistics. By applying iterative conditional modes (ICM), the HMRF model parameters can be estimated efficiently. The multiple testing is then carried out under an empirical Bayes framework by controlling the local false discovery rate (lFDR). To further increase flexibility and computation efficiency, we propose a hierarchical approach that focuses on subnetworks of co-expressed genes, which vary in structure from loosely netted gene modules to very tightly clustered genes sets. Data from simulations and from an autism study of rare variants demonstrate that our proposed approach is much more powerful than the single feature analysis. Our procedure is computationally feasible with provable performance and optimality guarantees under realistic assumptions. The proposed method can be applied to a wide variety of multiple hypothesis and feature selection problems.

1902T

A New Statistical Method for Identifying Differentially Methylated Regions in Complex Diseases. *P. Liu[1,2], C. Chen[1,2], X. Hua[1,2], Y. Lu[1,2], M. Liang[1].* 1) Department of Physiology, Medical College of Winconsin, Milwaukee, WI; 2) Cancer Center, Medical College of Winconsin, Milwaukee, WI.

DNA methylation is an important epigenetic modification that regulates transcriptional expression in complex diseases, such as cancer. It occurs when a methyl group becomes covalently attached to the 5-carbon position of a cytosine residue at a ‘CpG’ site. Of particular interest are CpG sites located in CpG islands which are clusters of CpG sites showing dynamic variations of methylation levels and often colocalized with *cis* regulatory elements such as gene promoters. Increased DNA methylation in CpG islands within or close to gene promoters is typically associated with transcriptional repression or gene silencing. Genomic segments with enriched 5-Methylcytosine marks can be mapped to single-base resolution using next-generation sequencing. Such high-throughput base-resolution methylation data have unique features and hence require the development of new analytic approaches. These important features include dependence of methylation among CpG sites within a CpG island and the influence of depth of sequence coverage on measuring methylation rate at CpG sites. Here, we proposed a new statistical method for identifying differentially methylated CpG regions from base-resolution methylation sequencing Data. Our approach accounts for correlation structures among CpG sites and incorporates depth of sequence coverage as weights for measuring methylation rate in CpG sites. We evaluated the performance of our new approach and compare it to other tests including the most commonly-used t-test and Fisher’s exact through a wide range of simulations. The simulation results show that regular tests for analyzing methylation sequencing data dramatically inflate type I error rates in the presence of physical proximity of CpG sites. While our new method is a valid test and gains statistical powers when weighting methylation rate by depth of sequence coverage. To demonstrate its utility, we applied it to the analysis of methylation sequencing data from non-small cell lung cancers and revealed novel methylation patterns in lung cancer.

1903F

A nonparametric model for haplotypes in population bottlenecks. *L.T. Elliott[1], Y.W. Teh[2].* 1) Gatsby Computational Neuroscience Unit, University College London, London, UK; 2) Department of Statistics, University of Oxford, Oxford, UK.

Clustering methods based on hidden Markov models are useful for approximating the haplotype structure of genetic sequence data that have undergone a recent population bottleneck or admixture. In these models (such as fastPHASE, IMPUTE2 and BEAGLE), latent cluster indicators can specify the genetic founder or component from an admixed population from which a loci on a given sequence originates. In this work we present BNPPHASE (Bayesian nonparametric Phase), a nonparametric model of haplotype structure based on a hierarchical Dirichlet process hidden Markov model (HDP-HMM). Bayesian nonparametric models handle latent variable domains with unbounded dimensionality and in BNPPHASE the number of genetic founders is unknown but inferred concurrently with the rest of the model without a separate model selection phase. BNPPHASE describes genetic sequences in which the proportions describing the distribution of each one of an unknown number of genetic founders varies along the chromosome reflecting ancestral recombination events. BNPPHASE advances previous work in nonparametric genetic models by providing nonhomogeneous cluster proportions and by identifying haplotypes across loci through the use of the recently proposed 'sticky' HDP-HMM leading to more efficient inference and to posteriors that more closely resemble the genetic process. The popular fastPHASE model can be seen as a finite truncation of BNPPHASE. The BNPPHASE model can be described by a generative process in which the cluster indicators for each sequence either remain the same from one loci to the next or, with a probability given by a recombination rate, are updated according to local Dirichlet processes that are linked together using a standard hierarchical Dirichlet process model. In a series of experiments involving phased data simulated from a population bottleneck with K ancestral genetic founders, we hold out data and compare the imputation accuracy of BNPPHASE with that of its finite truncations. We show that the flexibility of BNPPHASE, which stems from its nonparametric nature, allows BNPPHASE to achieve higher imputation accuracy even after the capacity of a given finite model is exhausted. Furthermore, we show that BNPPHASE can be used to estimate the number of genetic founders in a sample and that it can discover situations in which the total number of ancestral genetic founders changes along the chromosome due to selection or genetic drift.

1904W

Using network methodology to infer population substructure. *D. Prokopenko[1], C. Lange[1,2,3,4], J. Hecker[1], P. Costa[1], E.K. Silverman[3], H. Loehlein Fier[1].* 1) Institute of Genomic Mathematics, University of Bonn, Bonn, Germany; 2) Department of Biostatistics, Harvard School of Public Health, Boston, United States; 3) Channing Laboratory, Brigham and Women's Hospital, Boston, United States; 4) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany.

One of the main caveats of association studies is the possible affection by bias due to population stratification. Existing methods rely on model-based approaches like structure (Pritchard et al. 2000) and ADMIXTURE(Alexander et al. 2009) or on principal component analysis like EIGENSTRAT(-Price et al. 2006, Patterson et al. 2006).

Here we describe the problem of population substructure from the graph-theoretical point of view. We group the sequenced individuals into triads, which depict the relational structure, utilizing a similarity measure, i.e. covariance matrix. We then merge the triads into a network. We apply community detection algorithms in order to identify homogeneous subgroups or communities, which can further be incorporated as covariates into a logistic regression. We apply our method to populations from different continents in the 1000 Genomes Project and evaluate the type 1 error based on the empirical p-values. Our simulation results suggest that the network approach provides a more precise information of population structure than existing methods.

1905T

eqtl-tools: a toolset for scalable eQTL mapping. *A. Di Narzo[1,2], K. Hao[1,2].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Institute of Genomic and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.

Motivation: Expression Quantitative Trait Loci (eQTL) mapping on modern datasets requires the evaluation of a massive number of correlations. Current tools are either too slow or have hard limits on the input/output data formats and memory requirements. They also lack the flexibility to accommodate sex chromosomes or nonparametric models. Implementation: eqtl-tools is a set of command line utilities for fast and comprehensive eQTL mapping which uses plain text, tab-delimited tables as the main data format, allows for easy control of memory usage, and handles missing data consistently. Conclusions: eqtl-tools significantly lowers the amount of RAM necessary to perform eQTL analysis, while incurring in moderate speed penalty w.r.t. the best alternative solution to date. It handles chromosome X and facilitates integration into custom data analysis pipelines. It also allows more sophisticated modeling, e.g. nonparametric methods, and calibrates empirical false discovery rate (FDR) through permutation. Availability: The package is easy to deploy, self-documenting, and released under the GPL-v3 license on eqtl-tools.googlecode.com. Contact: antonio.dinarzo@mssm.edu.

1906F

Trans-ethnic fine mapping of cis eQTLs across eight ancestry groups.
C. Grace[1], J.C. Whittaker[2], J. Huxley-Jones[2], A.P. Morris[1]. 1) Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, Oxfordshire, United Kingdom; 2) GlaxoSmithKline, Medicines Research Centre, Gunnels Wood Road, Stevenage, UK.

Trans-ethnic meta-analysis can increase power to detect expression quantitative trait loci (eQTLs) and improve fine-mapping. We considered lymphoblastoid cell line expression in Phase 3 HapMap populations from multiple ancestry groups: CEPH (CEU, European), Chinese (CHB, East Asian), Gujarati Indians (GIH, South Asian), Japanese (JPT, East Asian), Luhya (LWK, African), Mexican (MEX, Hispanic), Maasai (MKK, African) and Yoruba (YRI, African). Genotypes were available for all samples and were used as a scaffold for imputation with the Phase 1 1000 Genomes panel (June 2011). We tested for association with expression of 1639 probes within each study for cis eQTLs (1Mb upstream and downstream of gene), and then combined studies via trans-ethnic meta-analysis implemented in MANTRA, a Bayesian approach which allows for heterogeneity in allelic effects between groups. We first combined studies within the same ancestral group: African (LWK, MKK, YRI), East Asian (CHB, JPT) and European, South Asian and Hispanic (CEU, GIH, MEX). Subsequently, we combined the 'ancestry-specific' association statistics across ethnicities. With this data, we aimed to evaluate the improvement in discovery and fine-mapping of eQTLs through trans-ethnic meta-analysis and 1000 Genomes imputation. After imputation, the lead SNP at 1,629 probes achieved genome-wide significance (log10 BF > 6), compared with 1,591 for HapMap genotypes alone. An improvement in the strength of the association at the lead SNPs was observed after imputation for 1,239 probes. Probe ILMN_10409_6860670 (chr10:105156223; HGNC: USMG5; up-regulated during skeletal muscle growth 5) has the greatest increase in association signal after imputation. The lead typed SNP is rs7831 (log10 BF = 132), whilst the lead imputed SNP is 10-105161778 (log10 BF = 250). To determine which ancestry groups are most informative for fine-mapping, we defined 99% credible sets of imputed and HapMap SNPs that are most likely to contain the causal variant. As expected, the African group has the lowest median credible set size (44 SNPs, 154 Kb), followed by the European, South Asian and Hispanic group (82 SNPs, 227 Kb) and East Asian group (105 SNPS, 318 Kb). Combining all ancestral groups leads to further improvements in resolution (median 99% credible set size of 8 SNPs over 19 Kb). This study has demonstrated that trans-ethnic meta-analysis and imputation provides a powerful means for discovery and fine-mapping of eQTLs.

1907W

Statistical methods for the analysis of gene expression in single-family studies for genetic and complex disease.
S. Merella[1], P. Brambilla[3], F. Martinelli Bonesch[3], E. Stupka[1], P. Provero[1,2]. 1) Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute,Milan,Italy; 2) Dept of Molecular Biotechnology and Health Sciences, University of Turin, Turin, Italy; 3) The Laboratory of Genetic of Neurological Complex Disorders, INSPE & Division of neuroscience, San Raffaele Scientific Institute,Milan,Italy.

Multiplex families are a powerful tool to investigate heritability in complex diseases to investigate genetic components of the disease. Heritability can be investigated not only through genetic approaches (SNP genotyping or re-sequencing) but also using gene expression analysis. There are specific challenges in working with gene expression data. Each family can present different genetic makeup producing heterogenous gene expression across families. Previous approaches developed for multiplex families were focused on identifying significant genes across multiple families. Moreover in each family there will be a strong genetic component of gene expression from which the disease effect needs to be distinguished and identified. We developed a statistical method for the analysis of gene expression data on single-family studies to minimize the problem in interpreting these data, and in particular to disentangle of genetic effects from other factors affecting gene expression, such as pathological status and environmental exposure. We combined the information obtained from the Ornstein-Uhlenbeck (OU) and Phylogenetically Independent Contrasts (PIC) models, that have been previously applied to describe evolutionary process. In this work the two models have been transferred to a context in which the quantitative trait has different values in diseased and healthy individuals. The models are defined on trees derived from family pedigrees to explicitly model the inheritance of gene expression and thus to identify those genes that are affected by the disease status. In order to validate the method and verify the sensitivity and specificity of the test, simulated datasets were generated starting from a family pedigree in order to take into accounts the genetic component and the disease status of every gene for every individual. We also used a real dataset from a family pedigree of a multiplex family with 4 multiple sclerosis affected individuals. The results obtained on simulated datasets reports that the two models provide increased sensitivity with respect to traditional gene expression analysis approaches (t-test within LIMMA). The analysis of the MS affected family also identified genes and pathways associated with the disease.

1908T

Probabilistic Phase Concordance For Enhanced Detection Of Somatic Allelic Imbalance.
C.M. Hahn[1,3], S. Vattathil[2,3], P. Scheet[1,2,3]. 1) Biomathematics and Biostatistics Program, The University of Texas Graduate School of Biomedical Sciences, Houston, TX 77030, USA; 2) Human & Molecular Genetics Program, The University of Texas Graduate School of Biomedical Sciences, Houston, TX 77030, USA; 3) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

DNA microarrays have been successful for interrogating somatically acquired chromosomal aberrations, such as deletions, duplications, and copy-neutral LOH, often using paired tumor and normal samples. However, studies of tumor genomes in settings of high levels of normal DNA contamination, or for complex mixtures of multiple clones, require computational methods sensitive enough to detect subtle allelic imbalance (AI). For this purpose, Vattathil & Scheet (2013, Gen Res 23:152) introduced hapLOH, a method that incorporated haplotype information by calculation of a phase concordance between the B allele frequencies (BAFs) and the estimated haplotypes. This method improved power by taking into account the dependence in the direction of the BAF shifts. However, it ignored two pieces of information that should aid identification of AI, namely the magnitude of the BAF shifts (dispersion about an expected value) and knowledge about haplotype accuracy (i.e., marker pairs with poorly estimated haplotypes should be downweighted). We now introduce an extension to hapLOH that accommodates these features by combining the original output with BAF values and possible haplotype reconstructions to define a probabilistic phase concordance and further refine AI detection. We accommodate the magnitudes and direction of BAF shifts by taking the density of a pair of observed BAFs as a mixture of 4 bivariate normal distributions. The probabilistic concordance can then be used to test specific regions of interest and designate emission probabilities in a hidden Markov model. We identify regions of increased dispersion in the BAFs of SNP data from various simulated and real mixtures of tumor and normal DNA, as regions of allelic imbalance and their characteristics are indicative of tumor aberration types found within such mixtures. We anticipate this extension will harness the power obtained from considering the complete distribution of BAFs while retaining the speed of hapLOH. Further applications could aid time-course monitoring of subtle changes in tumor profiles or response to therapy.

1909F

Conducting Large-Scale Imputation Studies on the Cloud.
S. Buyske[1,4], K. Vahi[2], E. Deelman[2], U. Peters[3], T.C. Matise[4]. 1) Statistics & Biostatistics, Rutgers Univ, Piscataway, NJ; 2) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 3) Public Health Sciences, Fred Hutchinson Cancer Research Center (FHCRC), Seattle, WA; 4) Department of Genetics, Rutgers University, Piscataway, NJ.

The PAGE (Population Architecture using Genomics and Epidemiology) study has genotyped 70K individuals on the gene-centric 196K SNP Metabochip. Imputation on a PAGE sub-study, using Mach-Admix with the 1000 Genomes Project full panel, has 700+ computational tasks and requires 35 days/1K subjects on a moderately fast workstation. Since managing these computations by hand is impractical and much of the computation is parallelizable, we constructed a computational workflow.

The fastest imputation tasks complete in seconds while the slowest take 12+ hours and require 7+ GB of RAM. The computational needs of such a workflow can strain typical resources, yet campus clusters often lack high-memory machines. Commercial computational clouds, however, offer a range of compute instances optimized for different use cases. We structured the imputation steps as a computational workflow and provisioned resources from the Amazon Web Services to support the application execution. Pegasus Workflow Management System (Pegasus WMS) is used to manage the workflow on Amazon EC2. Pegasus WMS represents the workflow of an application in an abstract form, agnostic of the resources available and the location of data and executables. It compiles workflows into executable form by querying catalogs. The executable workflow can be executed on remote and distributed resources such as remote clusters and clouds. When errors occur, Pegasus tries to recover by retrying tasks, by retrying the entire workflow, by providing workflow-level checkpointing, by re-mapping portions of the workflow, by trying alternative data sources for staging data, and, when all else fails, by providing a rescue workflow detailing only the work that still remains. We defined an imputation workflow in Pegasus and have applied it to two different sub-studies of PAGE. We will use this approach on 18 additional sub-studies. For the current runs, we provisioned 20 m2.xlarge EC2 instances with 2 cores each and 17 GB of memory. We configured a Condor pool on these resources and used Pegasus to map the workflow onto these instances and run the workflows. Imputation from the Metabochip on our first 2 substudies showed 32 to 36 hours/1K subjects of EC2 run time with a serial time equivalent of 34+ days/1K subjects. There were 5 failures caused by transient S3 access issues from within the EC2 cloud. Pegasus automatically recovered by retrying the failed jobs.

1910W

SKAT Admix: extending the SKAT method for rare variant association to admixed populations. *A.E. Byrnes[1], M.C. Wu[1], M. Li[2], Y. Li[1,3,4].* 1) Biostatistics, University North Carolina, Chapel HIll, NC; 2) Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia Pennsylvania; 3) Genetics, University of North Carolina, Chapel Hill, NC; 4) Computer Science, University of North Carolina, Chapel Hill, NC.

The Sequence Kernel Association Test (SKAT), first proposed in 2011, was designed to measure association between a set of genetic variants (including rare variants) and complex genetic traits quickly and efficiently. Here, we expand the SKAT methodology for use in admixed populations by first predicting the ancestry for each individual, at each locus, by utilizing additional sequencing data from each of the parent populations. We then apply the SKAT methodology to minor alleles likely to be inherited from each parent population separately, resulting in a SKAT statistic for each parental population. For example, in the case of African Americans, we conduct one SKAT test for those markers of European ancestry and one for those of African ancestry. To evaluate this method, we present a series simulations intended to mimic large sequencing studies of a quantitative trait in African Americans. In our simulations, the chosen variants are only causal if they originate from one of the parental populations, either African or European. We also examine different numbers of causal variants and directions of association. We find that SKAT Admix has increased power to find association compared to SKAT alone in most cases. In addition, the SKAT method has been proven to preserve type I error without requiring permutation. In our simulations, SKAT Admix also preserves Type I error for variants of both African and European ancestry. Data from the 1000 Genomes Project suggest that many more rare variants are population specific, as compared to common variants. It has also been suggested that rare variants show stronger patterns of stratification than common variants. For these reasons, we suggest a test that attempts to separate genomic data based on the predicted population of origin for admixed data.

1911T

The visualization of probabilistic results from consumer genetic testing for ethnicity at AncestryDNA. *R.E. Curtis, K.H. Freestone, M.J. Barber, J.M. Callaway, K. Noto, Y. Wang, C.A. Ball, K.G. Chahine.* AncestryDNA, Provo, US.

An important, but often overlooked, challenge in consumer genetics is the design of engaging and informative data visualization strategies that help consumers understand and fully appreciate the results of their genetic tests. Successful data visualizations that allow accurate and meaningful interpretation of consumer genetic test results must (1) help consumers overcome erroneous preconceptions or assumptions, (2) bridge gaps in the genetics education of consumers and (3) communicate the uncertainty associated with probabilistic predictions to consumers who may not have a strong understanding of statistics. Conveying uncertainty associated with predictions is crucial because many aspects of genetic science rely on probabilistic theory, including identity by descent (i.e., cousin matching), admixture predictions (i.e., ethnicity), and relative disease risk. If the stochastic nature of genetic algorithms is not properly conveyed in the visualization, naïve users often will either whole-heartedly accept the results as ground truth or dismiss them altogether. At AncestryDNA, we have considered this problem in the context of admixture prediction based on autosomal SNP testing. Over the past year, we have delivered over 125,000 admixture predictions to consumers across the US and conducted multiple surveys to assess our customer's satisfaction with and comprehension of those predictions. User surveys revealed several areas for improvement: some consumers had incorrect assumptions about the test; many consumers were confused about genetic concepts and most consumers were unable to understand the probabilistic nature of the predictions. In response to our consumer reactions, we have developed new visualization tools that attempt to address the types of shortcoming described above and to create consumer-friendly approaches to describe statistical concepts such as confidence intervals, sensitivity, positive predictive value, and standard deviations. As part of the natural life cycle of iterative consumer-facing improvements, we survey and evaluate our user interfaces using systematic questionnaires and focus groups to judge which tools most effectively convey the concept of genetics and probabilistic theory to a largely naïve consumer population. We will present our findings based on the conviction that the principles we have applied in our iterative development, testing and refinement of user experiences can also extend to other aspects of consumer genetics.

1912F

Next Generation of Genotype Imputation Methods. *S. Das, G.R. Abecasis.* Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Genotype imputation is a key step in the analysis of human genetic studies as it facilitates in increasing power of gene mapping, enables combination of results across different studies and accelerates fine-mapping efforts. Imputation works by finding the shared haplotype segments between the sample individuals typed on a commercial array and a reference panel of more densely typed individuals (e.g. The International HapMap Project, The 1000 Genomes Project etc.). Advances in high-throughput sequencing technologies have resulted in rapid increase in the size of these publicly available data-sets. Using these as reference panels would soon result in prohibitive computational costs. We introduce a strategy called 'state space reduction' which reduces the description of short genomic regions to the number of distinct haplotypes rather than the total number of haplotypes. The existing formulas have been refined to work with only these distinct haplotypes in a series of short genomic regions which covers the whole genome. The proposed algorithm maintains the accuracy of the current methods while reducing the computational costs. We also formulate a recursive algorithm to find the optimal structure/design of such genomic regions.

1913W

SECA: SNP effect concordance analysis using genome-wide association summary results. *D. Nyholt.* Neurogenetics Lab, QIMR, Brisbane, QLD, Australia.

Epidemiological and clinical studies indicate many human complex disorders co-occur within an individual, while family and twin studies suggest correlations in familial and genetic liabilities. To date, over 1350 genome-wide association studies (GWAS) have been performed to identify common single nucleotide polymorphisms (SNPs), associated with 830 phenotypes (disease endpoints or quantitative traits). A surprising finding of GWAS is that many loci show pleiotropic effects by being associated with more than one distinct phenotype. Indeed, a study of 1380 genes and 1687 SNPs listed in the NHGRI Catalog of Published GWAS (February 4, 2011) found 16.9% of genes and 4.6% of SNPs show pleiotropic effects (Sivakumaran et al. 2011). Identifying and taking advantage of polygenic overlap across phenotypes can improve detection of genetic risk factors, because when risk is correlated across phenotypes, pooled analyses will be better powered than individual-disorder analyses.

To facilitate the discovery of pleiotropic effects and examine polygenic risk shared across two phenotypes, I have developed a user-friendly web-based application called SECA to perform SNP effect concordance analysis. SECA takes two input files of GWAS summary results, each containing: i) SNP rsID, ii) effect allele (EA), iii) non-effect allele (NEA), iv) p-value from association test, and v) regression coefficient (beta), odds ratio, or z-score for the EA relative to the NEA. SECA first aligns the SNP effects across the two GWAS summary results (dataset1 and dataset2), and extracts a subset of independent SNPs (randomly or prioritized by p-values in dataset1) via linkage disequilibrium clumping. Next, SECA tests for enrichment of overlapping (pleiotropic) SNPs and whether the effect directions are concordant across the datasets via binomial, Fisher's, and false discovery rate statistical tests of dataset2 SNPs, after conditioning on association results in dataset1. SECA also prepares Q-Q, true discovery rate and bar plots to visualize enrichment and polygenic overlap.

Using publicly available GWAS summary data, SECA corroborates recent results from the *Cross-Disorder Group of the Psychiatric Genomics Consortium* (Lancet 2013 381:1371-9), finding significant polygenic overlap between bipolar disorder, major depressive disorder, schizophrenia and autism spectrum disorder, but not with attention deficit-hyperactivity disorder.

1914T

Testing Hardy-Weinberg Equilibrium of Multi-allelic Markers: Computational R code for Implementing MCMC algorithm via a Markov Base. *M. Rao, S. Venkatesan.* Environmental Hlth, Univ Cincinnati, Cincinnati, OH.

The goal of the presentation is to test Hardy-Weinberg Equilibrium for Multi-allelic markers a la Fisher. The test involves looking at all possible genotype data sets with the same allele frequencies. This collection of data sets could be humongous. If this is the case, Fisher's exact test cannot be implemented. We have developed a Markov base to negotiate the collection. We can then implement a Markov Chain Monte Carlo Simulation Algorithm exploiting the Markov Base to carry out the Fisher's exact test. An R code will be presented to complete the computations. The code will be run on some examples.

1915F

Relative performance and application of gene- and pathway-level methods for genome-wide association studies. *G.L. Wojcik, W.H. Kao, P. Duggal.* Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD.

Background: Genome-wide association studies (GWAS) have identified more than 10,000 associated SNPs with 840 traits. Despite this success, there still remains the problem of 'missing heritability' for most traits. One contributing factor may be the result of examining one marker at a time as opposed to a group of markers that are biologically meaningful in aggregate. To address this problem, a variety of gene- and pathway-level methods were developed to identify putative biologically relevant associations. Markers are put into either gene-level units or pathways as designated in databases such as Gene Ontology (GO). A simulation was conducted to systematically assess the performance of gene- and pathway-level methods. **Methods**: Using genetic data from the Wellcome Trust Case Control Consortium (WTCCC), case-control status was simulated based on an additive polygenic model where cases have more risk alleles. A total of 20 pathways and 226 genes were selected from GO biological processes. We evaluated 20 methods (e.g. VEGAS, MAGENTA, GATES, HYST, SRT, Fishers, etc.) based on their sensitivity, specificity, type I and type II error. The influences of gene and pathway size, number of causal single nucleotide polymorphisms (SNPs) in each gene/pathway, and effect size were assessed. Simulations using two sample sizes were examined: traditionally underpowered (n=250 cases, 250 controls) and larger (n=2250 cases, 2250 controls). **Results**: Despite low overall sensitivity (20-60%), specificity was high (89-100%) with low type I error (0.1-6%). Classical methods, not developed to handle linkage disequilibrium, have higher sensitivity, but higher type I error. Newer methods that directly estimate correlation structures were underpowered to detect genes with smaller effect sizes, but type I error was low. All methods were able to detect genes that would have been ignored in a traditional GWAS. Pathway-level methods' performance was dependent upon database annotation. Gene-level methods can rely on annotation from standardized databases but pathway-level methods use a heterogeneous pool. **Conclusions**: The interpretation of gene- and pathway-level methods is dependent upon their specific, inherent assumptions and the resources they rely on. Low type I error and high specificity increase confidence in the identified genes and pathways, despite low sensitivity. Gene- and pathway-level methods may provide valuable insight into the 'missing heritability' of traits.

1916W

Using Gentrain and Z-Call to Identify Problematic SNPs in Rare Variant Genotype Calling. *J. Romm, I. McMullen, M. Jewell, J. Zhang, E. Pugh, KF. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality Next-Gen Sequencing (NGS), Genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. Many commercially available genotyping arrays (HumanExome, HumanOmni2.5, HumanOmni5 and all Illumina 'plus exome' arrays) contain SNP content that enables users to assay low minor allele frequency (MAF) variants (<1%). In order to accurately call these low MAF SNPs, CIDR first clusters a project's data using GenomeStudio version 2011.1, Genotyping Module version 1.9.4, GenTrain Version 1.0. We then use a series of filters based on GenomeStudio/Gentrain metrics (call rate, cluster separation, etc.) and flag SNPs that have likely clustering or calling errors for manual review and adjustment of clusters. In addition to these filters, we use Z-Call (Goldstein, et al, Bioinformatics, 2012) to detect SNPs where Gentrain has called a SNP completely homozygous and Z-Call detects a heterozygote cluster. SNPs flagged with Z-Call are manually reviewed and clusters are modified or the SNP is dropped. This method works well for a breadth of project sizes. Project 1, 95 total samples, had 317 SNPs flagged and 31 were dropped after manual review. Project 2, 284 total samples, had 389 SNPs flagged and 110 were dropped after manual review. Project 3, 4,079 total samples had 223 SNPs flagged and 88 were dropped after manual review.

1917T

Accurate and rapid genetic analysis of genomic data under mixed model with multiple variance components. *H.M. Kang, S. Yang.* Biostatistics Dept, Univ Michigan, Ann Arbor, Ann Arbor, MI.

Ultra-high dimensional genetic analysis of genomic data, such as expression quantitative loci (eQTL) mapping, becomes increasing important to unravel the regulatory mechanisms of genetic variation causing molecular phenotypic changes. Analysis of high-throughput sequencing data dramatically increased the hypothesis space and the computational burden. At the same time, the inherent confounding factors such as batch effects or population structure requires us to apply sophisticated statistical models such as mixed models, which even more increase the computational costs compared to simple linear models. Here we demonstrate an accurate and computationally efficient procedure for genetic association analysis with high dimensional expression data under linear mixed models with multiple variance components, at a computational cost comparable to the simple linear model. Our methods have three important advantages over existing methods. First, the computational complexity of our method is linear to the number of samples, markers, and genes, similar to the GRAMMA-Gamma method (Svishcheva et. al. 2012). However, our method does not rely on the simplifying assumption of constant inflation factor across the genome. Instead, it accurately incorporates marker-specific inflation factor (Kang et. al. 2010) without sacrificing the computational complexity. Second, our method robustly deconvolutes the variance component attributed by the technical batch effects from the variance component attributed by the hidden sample structure using an importance sampling approach. As a result, our method robustly controls Type I error for each phenotype while avoiding overcorrection due to misspecified correlation structure between samples. Third, our method achieves linear time estimation of variance component parameters for multiple variance components by utilizing multiple sets of precomputed singular vectors. As a result, the computational cost of our method across n samples, m markers, and g genes becomes $O(ngm)$ when n is smaller than g and m, which is the same to the time complexity of simple linear model. Our method is being implemented into the widely used EPACTS (Efficient and Parallel Association Container Toolbox for Sequence data) software package, with many utilities and visualization supports for eQTL analysis. We expect that the EPACTS-eQTL will be an extremely useful software package to unravel the causal regulatory mechanism from DNA and mRNA sequencing data.

1918F

Detecting genetic heterogeneity in complex diseases with a weighted U statistic. *Q. Lu[1], C. Wei[1], R.C. Elston[2].* 1) Dept of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio.

For most complex diseases, a large proportion of the genetic variants remain undiscovered. While current research interests have shifted toward uncovering rare variants, gene-gene/gene-environment interactions, and structural variations, the impact of genetic heterogeneity in human diseases has been largely overlooked. Converging evidence suggests that diseases with the same or similar clinical manifestations could have different underlying genetic etiologies. Most of the existing analytical approaches assume the disease under investigation has a homogeneous genetic cause and could, therefore, have low power if the disease undergoes heterogeneous biological pathways. In this paper, we propose a statistical approach, a heterogeneity weighted U (HWU) approach, for high-dimensional association analysis taking genetic heterogeneity into account. HWU can be applied to various types of traits (e.g., binary and continuous), and is designed for detecting heterogeneous genetic effects. Through simulations, we compared HWU with a non-heterogeneity weighted U (NHWU) and the conventional generalized linear model (GLM). The results showed that HWU has substantial gain in power compared to NHWU and GLM in the presence of genetic heterogeneity, while retaining a performance similar to that of NHWU and GLM when the effects are homogeneous. Using HWU, we conducted a genome-wide analysis to study genetic heterogeneity in nicotine dependence. The genome-wide analysis of nearly one million SNPs from the Study of Addiction: Genetics and Environments (SAGE) took 5.8 hours, identifying heterogeneous effects of two new genes (i.e., CYP39A1 and VDAC3) on nicotine dependence.

1919W

Efficient and Accurate Population-Scale KIR Typing from SNP Chip Data. *S. Leslie[1, 2], D. Vukcevic[1], J. Traherne[3, 4], S. Naess[5], M. Lathrop[6, 7], T.H. Karlsen[5], M. Moffatt[8], W. Cookson[8], J. Trowsdale[3, 4], G. McVean[9], S. Sawcer[10].* 1) Statistical Genetics, Murdoch Childrens Research Institute, Parkville, Victoria, Australia; 2) Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, Australia; 3) Cambridge Institute for Medical Research, University of Cambridge, UK; 4) Department of Pathology, University of Cambridge, UK; 5) Research Institute of Internal Medicine, Department of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, Oslo, Norway; 6) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 7) Fondation Jean Dausset - CEPH, Paris, France; 8) National Heart and Lung Institute, Imperial College London, UK; 9) Wellcome Trust Centre for Human Genetics, Oxford, UK; 10) Department of Clinical Neurosciences, University of Cambridge, UK.

KIR (killer immunoglobulin-like receptor) genes are of great interest in regard to resistance to viruses, autoimmune disease, reproductive conditions and cancer. Like HLA, they are highly polymorphic and some KIR molecules interact with HLA class I. Better understanding genetic associations of KIR with disease requires accurate typing of both HLA and KIR variants. KIR exhibit copy number variation (CNV) and haplotypes may comprise from 4 to 15 or more genes. This kind of diversity makes KIR typing expensive and time consuming. Thus to date KIR has been understudied. SNP imputation for other SNPs is now routine, and performs extremely well, and SNP-based imputation for HLA is very good and becoming more widely used, but currently there are no imputation methods for KIR. To address this we have developed a high-throughput, accurate statistical imputation methodology for KIR typing using SNP variation data. We use a reference data set of approximately 600 haplotypes of European descent typed at 305 SNP loci covering 400kb in the vicinity of the KIR genes. These SNPs were typed on Immunochip. We fit a statistical model to these SNPs that enables us to impute KIR alleles from SNP variation alone. Validation experiments, using cross-validation and in a separate group of individuals, show accuracy for typing KIR gene copy number is at least 95% for the overwhelming majority of KIR genes, with better that 90% accuracy for the rest. Despite the relatively small reference database our results provide strong proof of principle that accurate imputation of KIR from SNPs is feasible. Experience with other loci (e.g. HLA) indicates that accuracy is likely to improve dramatically with increased reference panel size, and that imputation is likely to perform well for non-European populations provided good reference data is available. Our results show that accurate and efficient typing of KIR genes from SNP-chip data is possible. This is a novel approach that will make KIR data available to large cohort studies using SNP chips, meaning that disease association with KIR can be directly studied for the first time. This will facilitate significant insights into the role of KIR in human disease.

1920T

PSEUDOMARKER 2.0: efficient computation of likelihoods using NOMAD. *A. Schaffer[1], E.M. Gertz[1], T. Hiekkalinna[2], S. Le Digabel[3], C. Audet[3], J.D. Terwilliger[2,4,5].* 1) National Center for Biotechnology Information, NIH, Bethesda, MD; 2) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, FINLAND; 3) GERAD and Départment de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, Montréal, CANADA; 4) Department of Psychiatry, Department of Genetics and Development, and Columbia Genome Center, Columbia University, New York, NY, USA; 5) Division of Medical Genetics, New York State Psychiatric Institute, New York, NY USA.

PSEUDOMARKER [Hiekkalinna et al. Hum Hered 2011; 71:256-266] is a software package that tests for significant relationships between a putative trait-predisposing locus and genotypes at a series of markers, such as microsatellites or single nucleotide polymorphisms (SNPs). Typically, such packages either test for cosegregation of markers with a putative trait-predisposing locus within families (genetic linkage), or test for differences in marker genotype frequencies between unrelated cases and controls (linkage disequilibrium, LD). With PSEUDOMARKER, one can jointly analyze linkage and LD, or LD conditional on linkage, using pedigree relationships among individuals and singleton individuals. PSEUDOMARKER version 1 maximizes several likelihood functions using a generalized pattern search (GPS) algorithm [Dennis and Torczon, SIAM J Optim 1991; 1:448-474] implemented in a custom version of the ILINK [Cottingham et al. Am J Hum Genet 1993; 53:252-263] program. The running time of PSEUDOMARKER depends on the number of times the optimization algorithm evaluates a likelihood function. We modified ILINK substantially to use the software package NOMAD [Le Digabel ACM Trans Math Softw 2011; 37:1-15] to maximize likelihoods instead of GPS, resulting in generally comparable or better optima with many fewer evaluations of the likelihood functions. Out of 288 tests, NOMAD reported a log likelihood more than 0.005 better than that found by GPS 68 times, while GPS was at least 0.005 better only 7 times. On 12 datasets with multiple tests per set, the number of likelihood evaluations in PSEUDOMARKER version 2.0 is reduced by a multiplicative factor of at least 2 on each dataset and reduced by a factor of more than 7 on the best test set, as compared to version 1. PSEUDOMARKER is available from http://www.helsinki.fi/~tsjuntun/pseudomarker.

1921F

MMAP: a comprehensive mixed model program for analysis of pedigree and population data. *J. O'Connell.* Div Endo/Diabetes/Nutrition, Univ Maryland, Baltimore, MD.

The application of mixed models for genetic analysis has seen a sharp increase the last several years. The power and flexibility of the mixed model has been applied to both population and pedigree data, to both single SNP and multi-SNP data as random and fixed effects, to gene expression and general omics data, to both estimation and prediction of genetic risk, to mapping both common and rare variants, and to understanding the genetic architecture of complex traits. Kernel machine methods, kriging, and non-parametric regression can be cast into the mixed model framework to take advantage of available machinery. The basic ingredients in the mixed model are covariance matrices that measure similarity between subjects based on genetic data.

MMAP is an optimized and flexible mixed model analysis platform that incorporates a wide range of covariance structures such as additive, dominance, epistasis, maternal and imprinting using pedigree and/or genotype data and also allows users to define their own covariance structures. Likelihood calculations use multi-threaded optimized matrix libraries to handle multiple random effects. MMAP can import data from a variety of imputation programs to avoid data manipulation and IBS/IBD programs to build covariance structures.

MMAP uses a fast low-memory method to calculate additive and dominant genetic covariance structures using SNP data, which can be quite challenging for large data sets. For polygenic SNP analysis MMAP can store SNP-covariance products to reduce the complexity subsequent analyses with the same subjects to linear regression, independent of phenotype or covariates.

We present results in both animal and human data. We show that the genetic architecture of complex traits can include significant non-additive variance such as dominance in milk traits estimated in 35,000 Holstein cows and in blood pressure in 3500 Amish subjects. We show that including sources of non-additive variance in estimating SNP effects can improve predictive ability. We also present a detailed comparison between pedigree and genomic estimates of heritability in the Amish to measure missing heritability.

1922W

SALAD: A software suite for admixture linkage analysis and discovery.
R. Johnson[1, 2], *G. Nelson*[1], *C. Winkler*[3]. 1) BSP CCR Genetics Core, Frederick National Laboratory, SAIC Frederick, Inc, Frederick, MD; 2) Chaire Bioinformatique, Conservatoire National des Arts et Mètiers, Paris, France; 3) Basic Research Laboratory, Frederick National Laboratory, SAIC Frederick, Inc, Frederick, MD.

The importance of accounting for population structure in genetic studies and the power of admixed populations has long been recognized. Technological advances over the last decade have made it possible to sample and analyze these populations for regions of admixture linkage disequilibrium associated with clinical outcomes. Admixture linkage analysis remains a powerful, important tool in genetic studies, and many good software packages exist to aid the researcher. Multiple software packages requiring disparate data formats are required to fully analyze and prepare all but the simplest admixture linkage models. The Suite for Admixture Linkage Analysis and Discovery (SALAD) was developed to bring together tools for admixture linkage studies from study design to manuscript preparation. The backbone of SALAD, local and global ancestry inference, is done using a Hidden Markov Model similar to other admixture linkage software. These ancestry estimates can be further analyzed in a flexible modeling framework, allowing the user to model clinical outcomes as a function of global and local ancestry in addition to relevant covariates, using the appropriate generalized linear model. The case-only locus-genome statistic implemented in ANCESTRYMAP (Patterson et. al. 2004) is also included, as well as basic locus-genome statistics for other data types (e.g. continuous or survival). This can be generalized to compare any two user defined models based on Bayes factors. Also included in SALAD are tools for ancestry informative marker panel discovery, reading and writing data in common formats, GWAS data analysis and various publication quality graphics functions. Power and accuracy of SALAD is tested using simulated data, and relevant comparisons with similar software packages are made.

1923T

Local Ancestry Inference for Whole Genome Sequence Data.
B. Maples[1,2], *S. Gravel*[3], *E. Kenny*[4,5,6,7], *C. Bustamante*[1]. 1) Genetics, Stanford University, Stanford, CA; 2) Biomedical Informatics, Stanford University, Stanford, CA; 3) Human Genetics, McGill University, Montreal H3A 1B1, Canada; 4) Genetics and Genome Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Statistical Genetics, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sanai, New York, NY.

Local ancestry inference is an important step in the genetic analysis of fully sequenced human genomes. Current methods can only detect continental-level ancestry (i.e., European vs. African vs. Asian) accurately even when using millions of markers. Here, we present RFMix, a powerful discriminative modeling approach that is faster (~30X) and more accurate than existing methods. We accomplish this by using a conditional random field (CRF) parameterized by random forests trained on reference panels. RFMix is capable of learning from the admixed samples themselves to boost performance and autocorrect phasing errors. RFMix shows high sensitivity and specificity in simulated Hispanic/Latinos, African Americans, and admixed Europeans, Africans, and Asians. Finally, we demonstrate that African Americans in HapMap contain modest (but non-zero) levels of Native American ancestry (~0.4%).

1924F

Estimating Sample Size and Power for Nested Cohort Studies.
R. Hoffmann. QHS, Pediatrics, Medical College of Wisconsin, Milwaukee, WI.

Objective: Because of the rarity of certain diseases, the high cost of gene sequencing and the high cost of obtaining gene expression data the sample size and power of a nested cohort study needs to be obtained when <u>all</u> the cases, but only <u>some</u> of the non-diseased are sampled.

Methods: A nested cohort study may sample all the cases in each cohort, but only some of the controls from each cohort. Obtaining estimates of Relative Risk (RR) or the expression ratio involves choosing the correct sample from the non-diseased from both cohorts. The statistical power was determined separately for matched, stratified (using a stratified random sample) and unmatched (using a simple random sample) non-diseased. The power was determined using simulated data with both rare variant and common variants making up a complex disease structure. Rare variants are defined as alleles that have <0.5 percent, < 1 percent or less than 5 percent prevalence. The disease structure is patterned on either (1) a disease structure that has a few well-defined candidate genes, as well as a disease structure that potentially involves as many as 5000 potential locations. Stratification ranged from no stratification to stratification that is related to the disease to stratification that is unrelated to the disease.

Results: A sample of matched non-diseased to the diseased is found to have the least power and the highest probability of a false discovery. A stratified random sample had high power and a substantially lower probability of false discovery. An unmatched sample was found to have substantially higher power, but a higher probability of false discoveries if substantial stratification is present and not related to the disease system. Continuous expression data had higher power, but otherwise showed the same pattern with respect to sample selection of the non-diseased.

1925W

Impact of measurement error on testing genetic association with quantitative traits.
J. Liao[1,2], *X. Li*[2,3], *T.Y. Wong*[1,2,4], *J.J. Wang*[5], *C.C. Khor*[1,6], *E.S. Tai*[4,7,8], *T. Aung*[1,2], *Y.Y. Teo*[3,4], *C.Y. Cheng*[1,2,4,8]. 1) Department of Ophthalmology, National University of Singapore and National University Health System, Singapore, Singapore; 2) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 3) Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore; 4) Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, Singapore; 5) Centre for Vision Research, University of Sydney, Sydney, Australia; 6) Division of Human Genetics, Genome Institute of Singapore, Singapore, Singapore; 7) Department of Medicine, National University of Singapore and National University Health System, Singapore, Singapore; 8) Duke-NUS Graduate Medical School, Singapore, Singapore.

Background: Ignoring measurement error of a phenotypic trait will lead to underestimation of the standard error and hence reduce the power to detect genetic associations. We examined the impact of sample size, allele frequency and effect size in the presence of measurement error for quantitative traits.

Methods: The statistical power to detect genetic association by comparing phenotype mean and variability was investigated analytically. The non-centrality parameter for a non-central F distribution was derived and verified using simulation study. We obtained equivalent formulas for the cost of phenotype measurement error while maintaining equal power. Effects of differences in measurements were examined in a genome-wide association study (GWAS) of two grading scales for cataract and a replication study of genetic variants influencing blood pressure with multiple measurements.

Results: The mean absolute difference between the analytic power and simulation power for comparison of phenotypic means and variances were less than 0.005, and the absolute difference did not exceed 0.02. To maintain the same power, a one standard deviation (SD) in measurement error of a standard normal distributed trait required a one-fold increase in sample size for comparison of means, and a three-fold increase in sample size for comparison of variances. GWAS results revealed almost no overlap in the significant SNPs ($p < 10^{-5}$) for the two cataract grading scales. Genetic variant rs1458038 associated with blood pressure was replicated in averaged blood pressure measurements but not in single blood pressure measurement due to larger measurement error.

Conclusions: We have developed a framework for researchers to quantify power or to calculate sample size in the presence of measurement error, which will be applicable to studies of phenotypes in which the measurement is highly variable.

1926T

Slicing the Genome: A New Approach to Association in Complex, Longitudinal Diseases. *A. Musolf[1], D. Londono[1], A.Q. Nato, Jr.[2], P. Vuistiner[3], C.A. Wise[4,5,6], L. Yu[1,7], S.J. Finch[8], P. Bovet[9], M. Bochud[3], T.C. Matise[1], D. Gordon[1].* 1) Department of Genetics, Rutgers University, Piscataway, NJ, USA; 2) Division of Medical Genetics, University of Washington, Seattle, WA, USA; 3) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 4) Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX, USA; 5) Department of Orthopedic Surgery, Texas Scottish Rite Hospital for Children, TX, USA; 6) Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX, USA; 7) Center of Alcohol Studies, Rutgers University Piscataway, NJ, USA; 8) Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA; 9) Unit for Prevention and Control of Cardiovascular Disease, Section of Non Communicable Diseases, Ministry of Health and Social Services, Seychelles.

We previously published a method that tests for association between a longitudinal phenotype and genetic variants. The method uses growth mixture models (GMM) to determine longitudinal trajectory curves. The Bayesian posterior probability (BPP) of belonging to a specific curve, an outcome variable from the GMM, is used as a quantitative phenotype in association analyses. Though the method proves to be powerful for a single causal variant under multiple inheritance scenarios, power significantly decreases when more than one causal variant is considered. Here, we present a new method designed to detect multiple causal SNPs associated with longitudinal phenotypes in both family and population studies. The method also allows for the incorporation of covariates. This novel method retains several ideas from our first method, however instead of performing individual association tests with each SNP, we slice the genome into non-overlapping blocks of 50 SNPs (which we term a "mega-locus") and obtain a significance value on each mega-locus. This is accomplished via the SumStat method, developed by Jurg Ott and colleagues. As SumStat works for population studies only, we use a modified procedure (TDT-HET) to test for family-based association. We consider various scenarios in our simulations, including four causal variants located within a single mega-locus and eight causal variants spread between two mega-loci on different chromosomes. We also introduce environmental covariates. Our data set is highly stratified to ensure robustness in the presence of population stratification. P-values for each mega-locus on each data set are computed. To adjust for multiple testing, the final p-values are combined via Fisher's method (per mega-locus) and by the false discovery rate (FDR). We report that our simulations: 1) appear to maintain the proper type I error and 2) have greater than empirical 75% power for most simulations. These results suggest that our method can detect multiple causal SNPs located in multiple regions across the genome. We believe that this method will be useful to researchers who are studying complex diseases with longitudinal phenotypes. It allows for potentially high power for association of causal loci with disease progression phenotypes for both population and family studies, even in the presence of confounding elements such as population stratification and environmental variables.

1927F

Use of Electronic Medical Records to Measure Phenotypic Heritability. *S. Hebbring[1], Z. Ye[2], M. He[1,2], J. Mayer[2], S. Schrodi[1], M. Brilliant[1].* 1) Center for Human Genetics, Marshfield Clinic, Marshfield, WI; 2) Bioinformatics Research Center, Marshfield Clinic, Marshfield, WI.

A priority of the NIH is to understand the genetic etiology of common and rare diseases so that genetics may be used in 'personalized medicine.' All diseases are the result of a combination of environmental and/or genetic factors. Conducting genetic/genomic studies is difficult without evidence of a strong genetic component by heritability measurements. Heritability can be measured in a variety of family structures including twins and extended families. A particular challenge in conducing heritability estimates is identifying appropriate families with available phenotypic information. To address this challenge, and to demonstrate proof-of-principle that electronic medical records (EMRs) can be used to efficiently characterize disease heritability, we estimated the heritability of muscular dystrophy (MD) in a Marshfield Clinic patient cohort (Personalized Medicine Research Project, PMRP) with available family structure data (4,475 families) all linked to patient EMRs. MD was chosen as a positive control because of its well characterized genetic etiology and has a well-defined ICD9 code (359.1) easily extractable from the EMR. In this population, 6 out of 18 MD cases were familial in nature. The broad sense heritability estimated for MD using the EMR was 0.63 (SE 0.16), and closely matched expected estimates. To further follow-up on the use of EMRs to characterize the potential genetic contribution of human disease, we established a Marshfield Clinic twin cohort (15,802 twin pairs) and again ascertained MD diagnosis using the EMR. In the twin cohort, there was an enrichment for disease concordance in twins (4 affected pairs out of 17 individuals, p=1.59E-8). These results, both from extended and twin families, demonstrate proof-of-principle that the EMR may be a powerful and efficient tool when doing studies of heritability even in the absence of existing genetic/genomic data.

1928W

Optimal Strategies for Identifying Disease Associated Singletons. *S. Rashkin, G. Jun, G. Abecasis.* Center Statistical Genetics, Univ Michigan, Ann Arbor, MI.

Rare functional variants are hypothesized to explain much of the heritability of common, complex diseases. These variants are generally missing from genotyping array based studies and understanding their role in disease requires sequencing. Singletons, the rarest of these variants, can only be detected with high power by deep sequencing, which remains expensive. At lower depth, costs can be greatly reduced, enabling sampling sizes to increase, but many singleton variants will be missed. Here, we examine the balance between power to detect rare variants at different sequencing depths and sample sizes, so as to maximize power and minimize cost for studies that explore the role of singleton variants in human disease.

We used a simulated multisample caller to calculate power of singleton discovery based on different read depths (2-50x), sample sizes (20-5000), and error rates (0.001-0.01). While all three factors influence power, read depth appears to have the largest effect. At 10x coverage, power to detect a singleton variant is greater than 0.75 for all combinations of sample sizes and error rates examined. At 20x coverage, power exceeds 0.95, regardless of sample size and error rate. We validated our simulation by down-sampling deeply sequenced exome samples and assessing our ability to detect previously called singletons.

While we are interested in our ability to detect singletons, it is also important to study how this affects association results. We approximated the power of an association test using analytical computations that considered sample size, population frequency of singletons, relative risk for carriers, prevalence of disease, and read depth. Analysis of association study power allows for determination of the ideal sample size and read depth to use in studying singletons under a variety of conditions. Changes in sample size, relative risk, or frequency of singletons have a large effect on power: increasing frequency of singletons or relative risk allows for smaller sample sizes to attain high power. For a fixed sample size, at lower depth (<10X), increasing coverage increases power; as depth increases (typically beyond >15-20X), power remains relatively constant. Thus, our results suggest that studies of the contribution of rare variants to human disease will be optimally powered at coverage of ~10X and that coverage should only be increased when such an increase does not require a decrease in sample size.

1929T

Response Dependent Sampling Designs And Analysis In Studies With Rare Variants. *L. Sun[1,2], A. Derkach[2], J.F. Lawless[3,1].* 1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 2) Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada; 3) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON, Canada.

Rare variants play an important role in studies of complex human diseases and traits, and next generation sequencing technology provides rich data for analysis. This recent focus on rare variants has produced a significant number of sequencing studies as well as a large number of association tests. However, theoretical and empirical results have shown that in order to achieve high power large sample sizes are needed. Compared to random sampling, it has been shown that response-selective sampling designs can greatly improve the power of association studies with common variants and recently, rare variants. We extend the range of tests for multiple rare variants and show how adjustments to existing tests can increase power. We also investigate sampling strategies where random samples are taken from the tails of the response distribution and there may also be stratification on covariates. Through theoretical calculations and extensive simulations, we investigate what designs and testing approaches are efficient and robust across various genetic models.

1930F

Quantitative-trait-dependent sampling designs for genetic association analysis of a rare variant score. *Y.E. Yilmaz[1, 2, 4], J.F. Lawless[2, 3], S.B. Bull[1, 2]*. 1) Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada; 4) Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

For rare variant analysis related to a quantitative trait (QT), selection of individuals for sequencing according to their QT value can improve cost-efficiency. We consider QT-dependent two-phase sampling designs for regression analysis of a QT on a rare variant score obtained by aggregating rare genetic variants in a specified functional unit. In the first phase, we obtain trait values for all individuals in a cohort; in the second phase, we obtain genetic sequence data for a subset of individuals selected according to their trait values. Under stratified sampling, individuals in strata defined by high or low trait values can be selected with a higher fraction than those with an intermediate trait value. Under extreme-trait sampling, only individuals with high or low traits values are selected. In such trait-dependent sampling designs, inference based on standard linear regression methods that treat the QT as the dependent variable, ignoring the selection, may be misleading. We review the key features of two well-known methods for analysis under two-phase designs: likelihood-based inference for response-biased samples and inverse probability weighting (IPW) based on estimating equations, and describe analytic and simulation-based approaches to compare various sampling designs according to the number and size of strata, and the allocation of the phase 2 sample to the specified strata, as well as the distribution of the rare variant score and associated effect size. The relative efficiency of alternative designs depends on the method of analysis. Under a stratified sampling design, and assuming a linear relationship between the trait and rare variant score with normally distributed error term, likelihood inference favours extreme-trait-selection whereas IPW analysis requires non-zero sampling fractions in all strata and can be sensitive to sparsely sampled strata. Based on evaluation of type I error and power of a likelihood ratio test of rare variant association by simulation, we recommend likelihood-based design and analysis for both extreme-trait and more general stratified sampling, with the choice between them based on considerations of robustness as well as relative efficiency.

1931W

Caveats of extreme sampling strategies for resequencing studies. *H. Qin[1], W. Ouywang[1], S. Cao[1,2], T. Yang[1], Y.-P. Wang[1,2], H.-W. Deng[1]*. 1) Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112, USA; 2) Department of Biomedical Engineering, Tulane University, 534 Lindy Boggs Building, New Orleans, LA 70118, USA.

Selective sampling strategies are crucial to enrich causal rare genetic variants for cost-effective next-generation resequencing studies. One widely employed selective sampling strategy is extreme phenotype sampling - sampling individuals with extreme phenotypes. It seems to be a common sense that extreme phenotype sampling can enrich the presence of causal rare genetic variants compared to random sampling. An intuitive alternative of extreme phenotype sampling is extreme residual sampling - sampling individuals with extreme residuals after adjusting for covariates (e.g., identified genetic variants, population structure surrogates, and environmental factors) from phenotypes. In this paper, we mathematically modeled the causal architecture among phenotype, target genetic variants, and covariates. Under the causal architecture, we analytically investigated the probabilities of the three sampling strategies to detect target genetic variants. We proved the following results: Extreme phenotype sampling is inferior to random sampling when covariates are negatively correlated with target genetic variant. Extreme residual sampling is superior to random sampling; but it is inferior to extreme phenotype sampling if covariates are positively correlated to target genetic variant and target genetic effect is mild. Typical examples are reported for illustrations. In conclusion, no sampling strategy can uniformly dominate all the others under all circumstances. For a resequencing study, the 'optimal' sampling strategy depends on the underlying causal architecture among phenotype, target genetic variants, and covariates.

1932T

Identifying potential cancer vaccine targets with high-throughput sequencing. *E. Aronesty[1,2], K. Robasky[2], W.D. Jones[2,3]*. 1) Bioinformatics, Johns Hopkins University, Washington, DC; 2) Expression Analysis a Quintiles Company, Durham, NC; 3) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC.

Since the HPV vaccine was introduced to prevent cervical cancer, other vaccine targets remain elusive. Here, we use RNA-Seq to detect viral homologs in tumor sample expression data obtained from TCGA. We compare with normal tissue expression obtained from GTEX data sets, and to 'paired-normal' from blood samples where available. We have inspected RNA sequence of for breast, liver, cervical, ovarian, prostate, colon, lung, and brain cancers, and found that the overall viral presence is higher in cancer tissue compared to uninfected patient tissue. We present a method for producing a virome, removing low-complexity reads, counting viral presence by genus, as well as two methods for comparing counts between patient tissue and GTEX reference tissue. Finally, we assemble contigs for statistically significant viruses, and annotate results in a manner sufficient for vaccine target identification.

1933F

A Dynamic Model for Classification of Gene Regulation with RNA-seq Data. *L. Li, M. Xiong*. University of Texas School of Public Health.

Characterizing gene regulation and capturing its feature will provide valuable information for understanding biological process. Variation in gene expression underlies many biological processes and holds a key to unraveling mechanism of gene regulation. However, the gene expression measured by microarray provides limited information to reveal the features of gene regulation. The rapidly developed NGS technologies have been becoming the platform of choice for gene expression profiling. RNA-seq for expression profiling offers comprehensive picture of transcriptome and has made a number of significant qualitative and quantitative improvements on gene expression analysis and provides multiple layers of resolutions and transcriptome complexity:the expression at exon, SNP, and positional level; splicing; post-transcriptional RNA editing across the entire gene; isoform and allele-specific expression. To unravel the features of gene transcription, we propose to use differential equation to model the observed number of reads across the gene. We view the number of reads or expression level at each position as a function of the genomic position and view the transcription process as a dynamic process of transcription along the genome. Instead of taking derivative of expression level with respective the time, we calculate derivative of the expression level with respect to genomic position. We use a second order differential equation to model the dynamics of transcription process along the genome. Iterative principal differential analysis is used to estimate the coefficients in ODE by specifying the ODE as the data-driven penalty in the B-spline smoothing. We iterate between curve smoothing and ODE estimation until convergence occurs. The proposed methods are applied to ovarian cancer RNA-seq data with 412 tumor samples from TCGA dataset. We study the stability and transient response of transcriptional process for each gene using its fitted differential equation. These dynamic features of the gene transcription can reveal various alternative splicing, alternative start and end of transcription, and isoforms. Based on the dynamic features of gene transcriptional process we use the coefficients of ODE to classify the gene transcription into four categories. Our results will provide valuable information for understanding the mechanism of gene regulation and unraveling disease process. These results may also open a new way to find drug target and disease treatments.

1934W

Telomere Length in Circulating Leukocytes Is Associated with Lung Function and Disease. *E. Albrecht[1], E. Sillanpää[2], S. Karrasch[3], A. Couto Alves[4], V. Codd[5,6], I. Hovatta[7,8], J.L. Buxton[9], S. Hägg[10,11], M. Mangino[12], G. Willemsen[13], K.H. Pietiläinen[14,15], C.P. Nelson[5,6], L. Broer[16,17], M.A.R. Ferreira[18], I. Surakka[15,19], C. Gieger[1], N.G. Martin[18], N.L. Pedersen[10], D.I. Broomsma[13], T.D. Spector[12], C.M. van Duijn[16,17], J. Kaprio[7,15,20], N.J. Samani[5,6], M.R. Jarvelin[4,21,22,23,24], H. Schulz[25].* 1) Institute of Genetic Epidemiology, Helmholtz Zentrum Munchen, Neuherberg, Germany; 2) Gerontology Research Center and Department of Health Sciences, University of Jyväskylä, Finland; 3) Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany; 4) Dept of Epidemiology and Biostatistics, MRC-HPA (Health Protection Agency) Centre for Environment and Health, School of Public Health, Faculty of Medicine, Imperial College London, UK; 5) Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; 6) Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, UK; 7) Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland; 8) Department of Biosciences, Viikki Biocentre, University of Helsinki, Finland; 9) Department of Medicine, Imperial College London, UK; 10) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 11) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 12) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 13) Dept of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands; 14) Obesity Research Unit, Department of Medicine, Division of Endocrinology, Helsinki University Central Hospital and University of Helsinki, Finland; 15) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland; 16) Netherlands Consortium for Healthy Aging, Leiden University Medical Center, Leiden, the Netherlands; 17) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 18) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 19) Public Health Genomics Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland; 20) University of Helsinki, Hjelt Institute , Dept of Public Health, Helsinki, Finland; 21) Institute of Health Sciences, University of Oulu, Finland; 22) Unit of General Practice, University Hospital Oulu, Finland; 23) Biocenter Oulu, University of Oulu, Finland; 24) Department of Lifecourse and Services, National Institute for Health and Welfare, Oulu, Finland; 25) Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

Clinical and experimental studies suggest the involvement of premature aging in COPD. Using an epidemiological approach we studied whether accelerated aging indicated by reduced telomere length, as a marker of biological age, is associated with COPD and asthma, and whether intrinsic age-related processes contribute to the inter-individual variability of lung function. Our meta-analysis of 14 studies included >1,000 COPD cases with >15,000 controls, >2,500 asthma cases with >28,000 controls, and the spirometric indices forced expiratory volume in one second (FEV1), forced vital capacity (FVC), and their ratio FEV1/FVC in >13,000 individuals. Associations were tested by linear regression, adjusting for age, sex, and smoking status. We observed negative associations between telomere length and COPD as well as asthma, with stronger effects in women compared to men. The analysis of spirometric indices showed positive associations between telomere length and all investigated spirometric measures. The associations were weaker in apparently healthy subjects compared to COPD or asthma patients. Our results indicate that lung function may reflect biological aging primarily due to intrinsic processes which are likely to be aggravated in lung diseases. Shortened telomeres in COPD and asthma suggest that induced aging is involved in the pathogenesis of these diseases.

1935T

How much does family history information add to other risk factors when predicting a patient's risk of colorectal cancer? *L. Jonah[1], B.K. Potter[1], J. Little[1,2], J.C. Carroll[3], Q. Hasanaj[1], J. Allanson[4], D. Castle[5], B.J. Wilson[1], CIHR Emerging Team in Genomics in Screening.* 1) Department of Epidemiology & Community Medicine, University of Ottawa, ON, Canada; 2) Canada Research Chair in Human Genome Epidemiology; 3) Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 4) Department of Pediatrics, University of Ottawa, ON, Canada; 5) ESRC Innogen Centre, University of Edinburgh, Edinburgh, UK.

Background: Positive family history (FH) is associated with risk of many common complex diseases. Most clinical guidelines include FH in assessing disease risk. However, the evidence base to inform the routine use of FH in primary care is largely lacking. The 2009 NIH State-of-the-Science Panel on Family History and Improving Health concluded that, for FH to be established as an evidence-based tool, there is a need to evaluate its predictive ability and prognostic value in combination with traditional risk factors. Published research suggests that FH might offer statistically significant improvements to disease risk classification, but clinical utility still needs to be evaluated. We report the findings of a study of colorectal cancer (CRC) risk prediction, in which the impact on clinical decisions of adding FH to other risk factor information is provisionally evaluated. **Objectives**: To assess the incremental improvement in individual CRC risk prediction which is gained by adding FH to other clinical information recommended in risk models, and its potential impact on patient classification and management. **Methods**: This is a secondary cross-sectional analysis of data from the Assessment of Risk of Colon Tumors in Canada Study, a case control study. The dataset contains information on a range of risk factors as well as FH. We used a risk assessment model developed by Freedman et al (2009) to select variables (diet, NSAID use, physical activity, etc) for univariate and logistic regression analyses. We developed three models: Model 1 included non-FH covariates; Model 2 added FH to Model 1; and Model 3 incorporated a broader range of risk factors available in the dataset, including FH. For each model, the predictive/discriminatory ability was evaluated using ROC curves. **Results**: We present the three models, their accompanying ROC curves, and the Δarea under the curve achieved by adding FH and other variables to the base risk model. We also present the results of reclassification analyses which clarify how individuals shift between risk strata under the three models, including alterations in false positives and/or false negative rates. We comment on the likely clinical implications of risk reclassification, and offer recommendations for future research to clarify the thresholds for cost-utility analyses. These analyses offer a framework for the practical assessment of genomic information in clinical and patient decision-making.

1936F

Population based study of permanent teeth agenesis in Japanese. *J. Machida[1], T. Nishiyama[2], S. Yamaguchi[5], M. Kimura[3], A. Shibata[3], T. Tatematsu[3], Y. Abe[1], S. Makino[1], H. Miyachi[3], K. Shimozato[3], Y. Tokita[4].* 1) Oral and maxillofacial surgery, Toyota Memorial Hospital, Toyota, Aichi, Japan; 2) Public health, Aichi Medical School, Nagakute, Aichi, Japan; 3) Maxillofacial Surgery, School of dentistry, Aichi-Gakuin University, Nagoya, Japan; 4) Perinatology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan; 5) Oral and Maxillofacial Surgery, Aichi Children's Health and Medical Center.

Tooth agenesis is one of the most common congenital anomalies in humans and it is characterized by developmental absence of teeth. Clinical and epidemiologic studies of defined geographic populations can serve as a means of establishing data for genetic counseling and as a first step in identification of causes. In this study, we carried out clinical and epidemiologic studies at Aichi Prefecture, Japan. Four thousand eighty eight medical records of 15 year old school children from west Mikawa area were reviewed and prevalence rates calculated. In addition, medical and dental records of 71 unrelated patients with tooth agenesis of Aichi-Gakuin University were reviewed to obtain sibling recurrence risk. The prevalence of hypodontia with absence of one to five missing permanent teeth, excluding third molars and oligodontia with absence of more than six permanent teeth were 6.84%(95% CI: 6.06%; 7.68%) and 0.13%(95% CI: 0.04%; 0.30%), respectively. Sibling recurrence ratio in absence of one, two, three to five, and six or more teeth were 0.245% (95% CI: 0.138%; 0.383%), 0.250% (95% CI: 0.073%; 0.524%), and 0.222% (95% CI: 0.064%; 0.476%), and 0.438% (95% CI: 0.264%; 0.623%), respectively. This suggests severe phenotype, oligodontia fitted with an autosomal dominant inheritance, whereas mild phenotype, hypodontia is most compatible with an autosomal dominant trait with incomplete penetrance and/or polygenic inheritance. Our findings on the intra oral distribution of agenesis of permanent teeth in children may help us better understand the etiology of agenesis.

1937W

Population Architecture using Genomics and Epidemiology (PAGE): The association of trans-ethnic genetic variation with glucose and insulin levels in PAGE. *S.A. Rosse[1], C.S. Carlson[1], D. Crawford[4], J. Haessler[1], C.A. Haiman[6], T. Matise[5], K.E. North[2,3], J. Pankow[7], N. Pankratz[8], U. Peters[1], A. Young[1], C. Kooperberg[1].* 1) Public Health Genetics, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC; 3) Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC; 4) Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN; 5) Department of Genetics, Rutgers University, Piscataway, NJ; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 7) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 8) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.

Type 2 Diabetes Mellitus (T2DM) has an increasing burden of morbidity and mortality worldwide, particularly in populations of non-European descent. Although Genome-Wide Association Studies (GWAS) have established numerous loci associated with diabetes-related glycemic traits, the subsequent identification of causal variants remains a considerable challenge. Leveraging differences in local linkage disequilibrium (LD) structure between genetically diverse populations is a powerful method to narrow candidate region. To explore the generalization of susceptibility loci for fasting glucose (FG) and fasting insulin (FI) we fine-mapped 17 previously identified loci by investigating a total of 12,227 variants genotyped on the MetaboChip in individuals without diabetes from the PAGE study. A previous analysis demonstrated that exploring European GWAS loci in African Americans is a powerful approach to identify likely functional candidate(s) and secondary signals for FG and FI within PAGE. In this analysis we expand upon previous work to investigate the generalization of European GWAS loci and search for novel SNP effects for FG and ln(FI) across multiple ethnicities in a total of 21,002 Hispanic, 8140 Asian, 3500 Hawaiian and 11,795 African American individuals. Association between SNPs and trait was evaluated using linear regression for FG and logistic regression for ln(FI) under an additive genetic model (adjusted for age, sex, BMI, diabetes status, and principal components). Each ethnicity and study was first analyzed separately. Meta-analysis was conducted on results across studies within each ethnicity and with all ethnicities combined. In our preliminary analysis we attempted to fine-map 17 European GWAS loci by genotyping an average of 720 SNPs per region in 3,215 Hispanic women from Women's Health Initiative. Using a Bonferroni adjusted threshold, we detected one significant SNP (rs116132956, β=1.29 mmol/L, p=6.12×10^{-5}) in the *DGKB* region previously identified in AA and EA. Although the direction of effect was the same across populations, the most extreme p-values were at three statistically independent SNPs for AA, HA, and EA suggesting fine mapping of functional variants at the *DGKB* locus. A deeper understanding of the genetic influences across populations on glucose and insulin levels will provide insight into the underlying biology of T2DM that could have important implications for the development of therapeutic targets and prevention strategies.

1938T

Estimating genetic correlations between traits from summary statistics. *H.K. Finucane[1], S. Lindstrom[2], A.L. Price[2].* 1) MIT, Cambridge, MA; 2) Harvard School of Public Health, Boston, MA.

Estimating genetic correlation, a measure of the shared genetic origin of two traits, is a problem of high interest (Lee et al. 2012 Nat Genet). Here, we present a method to estimate genome-wide genetic correlations from GWAS summary statistics, which are much more widely available than raw genotype data. We focus on the genetic correlation of genotyped SNPs, as in Lee et al. 2012. Our method is applicable to either quantitative or case-control traits, and includes corrections for cryptic relatedness between samples, and for linkage disequilibrium (LD) between SNPs. For a fixed set of SNPs, let b and b' be vectors of normalized SNP effect sizes for two phenotypes. The genetic correlation between these two phenotypes is defined as r = E[bb']/(Var(b)Var(b'))^(1/2). We estimate E[bb'] by using the estimated effect sizes (i.e. summary statistics) and applying a correction for LD. We estimate Var(b) and Var(b') by using the estimated effect sizes and applying corrections for sampling noise, LD, and cryptic relatedness. Our derivations quantify the impact of both cryptic relatedness and LD on average chi-square statistics, extending the derivation of Yang et al. 2011 Eur J Hum Genet for strictly unrelated samples with LD. Theory and simulations show that the estimates of E[bb'], Var(b) and Var(b') are unbiased, and the resulting estimate of genetic correlation is consistent. Under the infinitesimal model in which true effect sizes are normally distributed, this estimator of genetic correlation is the maximum likelihood estimator. The corrections for LD and cryptic relatedness are based on a reference panel of raw genotypes from the same population, but we evaluate the method's robustness to mismatched reference panels. We validate the method using data from the Wellcome Trust Case Control Consortium, and apply the method to estimate the genetic correlation between mammographic density and breast cancer using summary statistics from over 50,000 samples.

1939F

Derivation of a genome-wide significance threshold for African populations. *M.D. Fortune[1,2], I. Tachmazidou[1], E. Zeggini[1].* 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Department of Applied Mathematics and Theoretical Physics, Cambridge, UK.

Genome-wide association studies examine common variation across the genome for association with complex traits of interest. Significance is declared at the widely-accepted threshold of p<5.0E-08. This has been derived from the total number of effective common variant (minor allele frequency (MAF) >0.05) tests in European populations and has been based on HapMap data. As the GWAS field is shifting to the study of more structured and heterogeneous populations, for example of African descent, a new statistical significance level has to be defined. Lower levels of linkage disequilibrium between common variants may necessitate a more stringent threshold. In addition, the availability of sequence data further empowers the assessment of the effective number of independent tests, as common variation has been comprehensively assayed. Many methods exist which exploit the correlation structure, either haplotypic or genotypic, between the variants. We have implemented several of these on two African datasets, Luhya in Webuye, Kenya (LWK) and Yoruba in Ibadan, Nigeria (YRI) , from the 1000 Genomes Project (sequence data, phase I integrated public data release), in order to estimate the effective number of tests for common genetic variation (MAF over 1 or 5%) in African populations. For comparison we also used the Utah residents (CEPH) with Northern and Western European ancestry CEU dataset from the same source. Using the haplotypic correlation coefficients, as proposed by Moskvina and Schmidt resulted in an estimate of 3.0E-8 for the European dataset, and 1.15E-8 for the African datasets at MAF over 5%. The same algorithm at MAF over 1% gave an estimate of 1.5E-8 for the European dataset, 7.0E-9 for LWK and 4.4E-9 for YRI. This reflects the greater genetic variation in present day sub-Saharan African populations.

1940W

Mediating genetic effects using twin data. *A. Ulgen[1], W. Li[2], J. Hjelmborg[3].* 1) Faculty of Medicine, Eastern Mediterranean University, Famagusta, TRNC, Mersin-10-Turkey; 2) Robert S Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore LIJ Health System, Manhasset, New York, United States of America; 3) Department of Biostatistics, Institute of Public Health, University of Southern Denmark, Odense, Denmark.

We apply a genetic modeling via simulation to twin data for obesity and cancer related measurements. For obesity, we use both BMI and other quantities derived from BMI to measure the weight growth. More specifically, the log(BMI) at baseline is regressed over other factors: ln(BMI_ij)= beta_i + alpha_i *j +gamma*age_i+epsilon_ij, for individual I and timepoint j (time since the baseline, in years). The alpha_i is then the log-weight growth rate (Hjelmborg et al. Obesity, 16(4), 2008). For cancer data, we use phenotypes available in the registry. We follow the genetic modeling of twin data proposed by (Dite et al. and Stone et al; Cancer Epidemiol Biomarkers Prev; 17(12), 2008 and 17(12), 2012, respectively.) In this modeling, the phenotype of a twin in a twin pair is regressed over both twins' co-variates. If the two twins in a twin pair are labeled as 1 and 2, Y denotes phenotype and X covariate, then E(Y)= alpha + beta_1 X_1 +beta_2 X_2. It was shown that by varying a covariate experimentally, the expected value of the phenotype measure would change.In our analysis, we assume a bivariate normal distribution for both (Y1,Y2) and (X1,X2). We treat phenotype measurements such as BMI, growth rate, at the baseline time as X, and that at the later time as Y. This approach would incorporate measurements at two points along a time course, thus enhance the power to detect the genetic component. We also introduce a random effects model for the stratification effects.

1941T

Detecting recent coevolution through ancestry association on different chromosomes in African-Americans. *H. Wang[1], Y. Choi[2], X. Wang[3], B. Tayo[4], U. Brockel[5], C. Hanis[6], S. Kardia[7], S. Redline[8], R. Cooper[4], N. Risch[9], H. Tang[2], X. Zhu[1].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Biostatistics, Harvard, Boston, MA; 4) Department of Public Health Science, Loyola University Medical Center, Maywood, IL; 5) Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI; 6) Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston, Houston, TX; 7) Department of Epidemiology, University of Michigan, Ann Arbor, MI; 8) Department of Medicine, Harvard Medical School, Boston, MA, USA; 9) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA.

Genetic coevolution could be maintained through compensatory mutations over evolutionary history. Recent studies have found evidence of coevolution between physically unlinked genes and loci. We hypothesized that recent allelic coevolution could be detected by examining simultaneous or complementary transmission of ancestry alleles in different genomic regions. This hypothesis can be tested in recently admixed populations such as African Americans. We analyzed the correlations of local ancestry across the genome in 20,097 African Americans sampled from three different study cohorts: CARe, WHI, and FBPP. The ancestral proportions were estimated using HAPMIX and SABER+. Inverse-variance weighted meta-analysis approach was used to combine results from the three study cohorts. Our analysis revealed 79 pairs of genetic regions on different chromosomes which are significantly correlated (p-value < 10-9 in meta-analysis), suggesting possible evidence of coevolution. Several of these regions contain genes previously demonstrate to be under selection such as SLC30A9 and ASPM.

1942F

Investigating bias due to population stratification in pharmacogenetic studies. *X. Chen, C. Molony, C. Zhang, H. Zhou.* Informatics and Analysis, Merck , Boston, MA.

It is well known that bias will be introduced in population-based association studies in the presence of population stratification (PS). Failing to control for PS might lead to identification of spurious association signals from ancestry-informative markers, rather than the causal disease loci. However, the impact and magnitude of this confounding by ethnicity remain unclear in the context of pharmacogenetic studies (PGx), where genome-wide associations studies (GWAS) are applied retrospectively on a subset of population from Phase II/III randomized clinical trials to discover genetic biomarkers associated with drug responses. Through using a simulation study design, we systemically investigated the type I error rates and bias in the estimates caused by PS in PGx. Various population substructures and treatment response models were generated to evaluate the bias in main genetic effects and the interactions between gene and treatment effects. Simulation results indicated that it is crucial to perform association analyses within each genetically homogeneous population even in well conducted randomized clinical trials. Otherwise false positive results due to differences in drug response rates across ethnicities were likely to be observed. With respect to the Phase II clinical trials conducted in a mixed population (e.g. African American), the bias due to PS appeared limited when differences of drug response rates among ethnicities are small to moderate. We also demonstrated that applying standard statistical approaches of adjusting ancestral histories can effectively provide unbiased estimations of main genetic effects and genetic by treatment interaction effects in PGx for identifying predictive biomarkers of drug response , while maintaining valid false-positive rates.

1943W

Validation of an ancestry estimation analysis method using a comparison of FRAPPE and STRUCTURE. *J.J. Bryan[1], K. Tang[1], R. Kittles[2], C.L. Mouritsen[1].* 1) Sorenson Genomics, 2495 South West Temple Salt Lake City, UT. 84115; 2) University of Chicago 5841 South Maryland Avenue AMB W601 MC6091 Chicago, IL 60637.

An analytical test method for estimating genetic ancestry has been developed wherein 190 Ancestral Informative markers (AIMS) are genotyped and compared to reference populations. This study was performed to ascertain the test's ability to accurately estimate admixture. Ancestry estimation was performed for 44 unrelated, self-declared 'African American' (ASW) samples from the HAPMAP3 dataset using a genome-wide dense marker set (~860k SNPs) to establish the 'true' genetic affinities to 5 reference populations (European, East Asian, India Subcontinent, Indigenous America, and Sub-Saharan Africa). The same set of 44 samples was analyzed again, using the 190 AIMs in our test. For the 190 AIMs, two different statistical programs were independently employed, FRAPPE and STRUCTURE, both recognized for their validity in this type of comparative analysis. Each sample analysis was performed 50x, with a permutation procedure to account for the variability in the reference sample panels. The genetic affinities were thus estimated with well defined confidence intervals. The ancestry estimation results with the 190 AIMs were then compared to that of the high density ~860k set and then to each other. The results showed that our ancestry tests with either method corresponded very well to the data from the ~860k estimations; and strong consistency was also observed between the FRAPPE and STRUCTURE programs. More than 90% of the samples having a comparative r-squared value greater than 0.92 between the two programs. The overall difference between results obtained using FRAPPE and STRUCTURE were minimal, with an average difference between the corresponding ancestry percentages less than 0.2%. These data were further validated using an additional 144 self-declared 'African American' samples, with near equivalent results.

1944T

Geographic Population Structure (GPS) of worldwide human populations infers biogeographical origin down to home village. *E. Elhaik[1], T. Tatarinova[2,3], D. Chebotarev[3], I.S. Piras[4], C.M. Calò[4], A.D. Montis[5], M. Atzori[5], M. Marini[5], S. Tofanelli[6], P. Francalacci[7], L. Pagani[8], C. Tyler-Smith[8], Y. Xue[8], G. Cucca[4], T.G Schurr[9], J.B. Gaieski[9], C. Melendez[9], M.G Vilar[9], R. Gomez[10], R. Fujita[11], F.R. Santos[12], D. Comas[13], O. Balanovsky[14,15], P. Zalloua[16], H. Soodyall[17], R. Pitchappan[18], A. GaneshPrasad[18], M. Hammer[19], L. Matisoo-Smith[20], S.R. Wells[21].* 1) Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205; 2) Glamorgan Computational Biology Research Group, University of Glamorgan, Wales, CF371HR, United Kingdom; 3) Laboratory of Applied Pharmacokinetics and Genomics, Children's Hospital Los Angeles, University of Southern California, 4650 Sunset Blvd, Los Angeles, CA 90027; 4) Department of Sciences of Life and Environment , University of Cagliari, Monserrato, SS 554, 09042, Italy; 5) Research Laboratories, bcs Biotech S.r.l., Viale Monastir 112, 09122 Cagliari, Italy; 6) Department of Biology, University of Pisa, Via Ghini 13, 56126 Pisa, Italy; 7) Department of Science of Nature and Territory, University of Sassari, Località Piandanna, Sassari, Italy; 8) The Wellcome Trust Sanger Institute, CB10 1SA, Hinxton, UK; 9) University of Pennsylvania, Philadelphia, PA; 10) CINVESTAV, Mexico City, Mexico; 11) University of San Martin de Porres, Lima, Peru; 12) Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil; 13) Institut de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain; 14) Vavilov Institute for General Genetics, Moscow, Russia; 15) Research Centre for Medical Genetics, Moscow, Russia; 16) The Lebanese American University, Chouran, Beirut, Lebanon; 17) University of the Witwatersrand, Johannesburg, South Africa; 18) Chettinad Academy of Research and Education, Chennai, India; 19) University of Arizona, Tucson, AZ; 20) University of Otago, Dunedin , New Zealand; 21) National Geographic Society, Washington DC, USA.

The search for a method that utilizes biological information to predict human's place of origin has occupied scientists for millennia. Modern biogeography methods are accurate to 700 km in Europe but are highly inaccurate elsewhere, particularly in Southeast Asia and Oceania. The accuracy of these methods is bound by the choice of genotyping arrays, the size and quality of the reference dataset, and principal component (PC)-based algorithms. To overcome the first two obstacles, we designed GenoChip, a dedicated genotyping array for genetic anthropology with an unprecedented number of ~12,000 Y-chromosomal and ~3,300 mtDNA SNPs and over 130,000 autosomal and X-chromosomal SNPs carefully chosen to study ancestry without any known health, medical, or phenotypic relevance. We also 615 individuals from 54 worldwide populations collected as part of the Genographic Project and the 1000 Genomes Project. To overcome the last impediment, we developed an admixture-based Geographic Population Structure (GPS) method that infers the biogeography of worldwide individuals down to their village of origin. GPS's accuracy was demonstrated on three data sets: worldwide populations, Southeast Asians and Oceanians, and Sardinians (Italy) using 40,000-130,000 GenoChip markers. GPS correctly placed 80%; of worldwide individuals within their country of origin with an accuracy of 87%; for Asians and Oceanians. Applied to over 200 Sardinians villagers of both sexes, GPS placed a quarter of them within their villages and most of the remaining within 50 km of their villages, allowing us to identify the demographic processes that shaped the Sardinian society. These findings are significantly more accurate than PCA-based approaches. We further demonstrate two GPS applications in tracing the poorly understood biogeographical origin of the Druze and North American (CEU) populations. Our findings demonstrate the potential of the GenoChip array for genetic anthropology. Moreover, the accuracy and power of GPS underscore the promise of admixture-based methods to biogeography and has important ramifications for genetic ancestry testing, forensic and medical sciences, and genetic privacy.

1945F

AdmixKJump: Identifying population structure in recently diverged groups. *T. O'Connor*[1,2,3]. 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland; 2) Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland; 3) Department of Genome Sciences, University of Washington, Seattle, Washington.

Correctly identifying population structure is important both to understand population history and to mitigate the potential for confounding in association analyses. Statistically, recent population divisions can be difficult to recognize, as there has not been substantial time for the groups to differentiate. Objective methods to identify recent population divisions are needed. ADMIXTURE has developed a cross-validation approach to select the correct number of K (i.e. clusters or putative populations), but how this statistic performs on recent population divisions with realistic simulations has yet to be evaluated. Also, alternative approaches from the informatics and statistical literature may be better suited to recent demographic events. I have implemented a new metric for admixture analysis, AdmixKJump, and compared it with the performance of the cross-validation statistic. I use a coalescent simulation framework based on parameters estimated from the Exome Sequencing Project to generate whole genome sequences with multiple populations. I vary the split time between populations to evaluate how accurate each method is in identifying the correct number of clusters over 50 replicates. With a sample size (n) of 50 for each of two populations, the admixture parameters themselves have almost no error at ~6KYA (ie ~1K years prior to the expansion). The cross-validation metric gains 100% power at about 14KYA, whereas AdmixKJump reaches 100% accuracy at 10KYA. I also find that the new measure has more power with smaller sample sizes (testing n=10, 20, ..., 50), for instance n=30 is 100% at 12KYA for AdmixKJump. I then apply the new metric to the results of the Great Apes Project and 1000 Genomes Project. I find comparable results to the cross-validation approach, emphasizing the narrow improvement window of AdmixKJump and the need for future improvements to fully capture the population structure in these data sets. In conclusion, I have developed a new objective approach to identifying population structure, which has more power than previous methods, especially with recent demographic events.

1946W

Distribution of genetic ancestry and candidate disease allele frequencies in Puerto Rico. *Y. Afanador*[1], *I. Rivera*[1], *W. Guiblet*[1], *LGDS. Consortium*[1], *M. Yeager*[2], *V. Washington*[3], *J.C. Martinez-Cruzado*[1], *T. Oleksyk*[1]. 1) Biology, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR; 2) Frederick National Laboratory for Cancer Research, Frederick, MD; 3) Biology, University of Puerto Rico, Rio Piedras Campus, San Juan, PR.

In modern human populations, sensitivity to complex diseases is often determined by past demographic events, selection, and admixture. Puerto Ricans are an excellent model to understand this relationship as they acquired characteristics from three ancestral origins: African, European, and Native American. We are using a geographically distributed random sample of reference from each of the 78 municipalities of Puerto Rico collected by undergraduate students from a NSF funded educational curriculum, the Local Genome Diversity Studies (LGDS). This study provides estimates for individual ancestry proportions across the island. Different admixture contributions carry different disease alleles that can account for prevalence of diseases that are on the top of the list of death causes in the Island: diabetes, kidney disease, heart disease, and cancer. Candidate genes for these diseases are being studied for their association with ancestry as well as the disease occurrence in the general population compared to the patients. For example, preliminary analysis of candidate gene polymorphisms for the end stage kidney disease (ESKD) patients from the island showed significant differences in frequencies of some disease-relevant variants, specifically in the NPHS2, ATF6, and ENPP1 genes when compared with the reference population, and the geographical distribution of the risk alleles shows a pattern consistent with the history of admixture across the island. These results demonstrate the importance of history and geography of admixture to the pubic health and personalized medicine decisions. Funding for this study came in part from NSF grant: DUE 1044714 TUES: Integration of Research and Undergraduate Education: Local Genome Diversity Studies in Puerto Rico.

1947T

Allelic Frequency Determination of Asthma-Related Single Nucleotide Polymorphisms and the Relation of Genetic Admixture in Asthma disease Prevalence among Puerto Ricans. *I. Rivera*[1], *Y. Afanador*[1], *C. Garcia*[1], *W. Guiblet*[1], *E. Suárez*[2], *J.C. Martinez-Cruzado*[1], *T.K. Oleksyk*[1], *Local Genome Diversity Studies.* 1) Biology Dept, University of Puerto Rico, Mayaguez Campus, Mayaguez, PR; 2) Biology Dept, University of Puerto Rico, Ponce Campus, Ponce PR.

Asthma is one of the most recognized complex human disorders characterized by an airway obstruction due to an exacerbated immune response. Currently, asthma is estimated to affect more than 235 million people worldwide with an annual cost of around $115 billion. In Puerto Rico asthma prevalence has been estimated to be 15% from 2000-2007, reaching an alarming 20% in 2003, making Puerto Ricans one of the populations with the highest asthma prevalence worldwide. Since asthma is more frequently seen in populations from African descent (around 12-14% Afro-Americans) and Puerto Ricans share on average 21% African ancestry, it has been suggested that an admixture factor might be involved. In order to evaluate this hypothesis, two sets of samples were collected: 1) a set of asthma-diagnosed patients; and 2) a systematically collected random sample set of individuals representing all 78 municipalities of the island: the Local Genome Diversity Studies cohort (LGDS). We assayed the allelic frequency of candidate asthma-related single nucleotide polymorphisms (SNPs) in patients and compared them to the general population reference samples from the LGDS cohort. In order to evaluate an additional admixture component that might contribute to asthma occurrence, we further analyzed the admixture proportions of African, European and Native American ancestry using a panel of ancestry informative markers (AIMs). Our results help validate previous studies in genetically diverse Hispanic populations, and indicate the potential for novel ancestry-related candidate gene polymorphisms that correlate with asthma occurrence in Puerto Rican population. Funding for this study came in part from NSF grant: **DUE 1044714** *TUES*: *Integration of Research and Undergraduate Education: Local Genome Diversity Studies in Puerto Rico* .

1948F

Evidence for interaction of population-specific EFHC1 alleles with genetic ancestry in juvenile myoclonic epilepsy. *R.L. Subaran*[1], *J.M. Conte*[2], *W.C.L. Sterart*[1], *D.A.G. Greenberg*[1]. 1) Research Institute, Columbus, OH. Nationwide Children's Hospital; 2) Columbia University, New York, NY. Mailman School of Public Health.

Missense mutations in the EFHC1 gene have been reported to cause juvenile myoclonic epilepsy (JME) with high penetrance in patients reporting Hispanic ancestry. However, observations from other studies call into question the magnitude and scope of this effect. To help probe the relationship of EFHC1 to JME, we examine the frequency of these mutations in a group of Hispanics identified in New York City and in participants from non-Hispanic populations. We screened 117 healthy controls (60 Hispanic, 57 non-Hispanic of various ethnicities) for EFHC1 mutations purported to cause JME in Hispanics. To search for novel pathogenic mutations, we sequenced the exons of EFHC1 in a newly ascertained group of Hispanic JME patients. We compared our findings to the frequencies found in 1KGP. The EFHC1 coding mutations we found in our patients were also found in our controls except for a novel splice-donor mutation carried by a single JME patient. Of the five coding mutations previously reported to cause JME with high-penetrance, we found three at appreciable frequencies in our non-Hispanic controls and in the non-Hispanic participants of the 1KGP. In general, EFHC1 coding mutations are not a major cause of JME in Hispanics. However, ascertaining specific Hispanic subpopulations on the basis of JME causes enrichment of EFHC1 alleles seen otherwise only in non-Hispanic populations, suggesting a hypersensitivity to disruption of EFHC1 in these Hispanic subpopulations. Importantly, our findings help shed light on the nature of this effect on a common disorder in a historically understudied population.

1949W

Genomic scans for haplotypes of Denisova and Neanderthal ancestry in modern human populations. *F.L. Mendez, M.F. Hammer.* University of Arizona, Tucson, AZ., USA.

Evidence of archaic introgression into modern humans has accumulated in recent years. While most efforts to characterize the introgression process have relied on genome averages, only a small number of introgressive haplotypes have been shown to have an archaic origin after rejection of the alternative hypothesis of incomplete lineage sorting. Accurate identification of introgressive haplotypes is crucial both to characterize potentially functional consequences of archaic admixture and to quantify more precisely the genomic impact of archaic introgression. We perform two independent genomic scans for haplotypes of Denisova and of Neanderthal origin in a geographically diverse sample of complete genome sequences. These scans are based on the local sharing of polymorphisms and linkage disequilibrium, respectively. The analysis of concordance between the methods is then used to estimate the power and to compare demographic inference when performed using either all the data or just the genomic regions with no evidence of introgression. Moreover, we evaluate the extent to which Denisova haplotypes are observed in non-Melanesian populations, and investigate whether the presence of such haplotypes is better explained by their persistence in the population since introgression or by more recent gene flow from Melanesians.

1950T

Admixture Estimation in a Founder Population. *Y. Banda[1], M. Kvale[1], T. Hoffmann[1], S. Hesselson[1], H. Tang[3], D. Ranatunga[2], L. Walter[2], C. Schaefer[2], P. Kwok[1], N. Risch[1].* 1) Institute Human Genetics, University California San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California, Division of Research, Oakland, CA; 3) Department of Genetics, Stanford University, Stanford, CA.

Admixture between previously diverged populations yields patterns of genetic variation that can aid in understanding migrations and natural selection. An understanding of individual admixture (IA) is also important when conducting association studies in admixed populations. However, genetic drift, in combination with shallow allele frequency differences between ancestral populations, can make admixture estimation by the usual methods challenging. We have, therefore, developed a simple but robust method for ancestry estimation using a linear model to estimate allele frequencies in the admixed individual or sample as a function of ancestral allele frequencies. The model works well because it allows for random fluctuation in the observed allele frequencies from the expected frequencies based on the admixture estimation. We present results involving 3,366 Ashkenazi Jews (AJ) who are part of the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort and genotyped at 674,000 SNPs, and compare them to the results of identical analyses for 2,768 GERA African Americans (AA). For the analysis of the AJ, we included surrogate Middle Eastern, Italian, French, Russian, and Caucasus subgroups to represent the ancestral populations. For the African Americans, we used surrogate Africans and Northern Europeans as ancestors. For the AJ, we estimated mean ancestral proportions of 0.380, 0.305, 0.113, 0.041 and 0.148 for Middle Eastern, Italian, French, Russian and Caucasus ancestry, respectively. For the African Americans, we obtained estimated means of 0.745 and 0.248 for African and European ancestry, respectively. We also noted considerably less variation in the individual admixture proportions for the AJ (s.d. = .02 to .05) compared to the AA (s.d.= .15), consistent with an older age of admixture for the former. From the linear model regression analysis on the entire population, we also obtain estimates of goodness of fit by r2. For the analysis of AJ, the r2 was 0.977; for the analysis of the AA, the r2 was 0.994, suggesting that genetic drift has played a more prominent role in determining the AJ allele frequencies. This was confirmed by examination of the distribution of differences for the observed versus predicted allele frequencies. As compared to the African Americans, the AJ differences were significantly larger, and presented some outliers which may have been the target of selection (e.g. in the HLA region on chromosome 6p).

1951F

Using a haplotype-based model to infer Native American colonization history. *C. Lewis[1], D. Balding[1], S. Myers[2], G. Busby[2], C. Capelli[2], D. Falush[3], A. Ruiz-Linares[1], G. Hellenthal[1].* 1) University College London, London, United Kingdom; 2) University of Oxford, Oxford, United Kingdom; 3) Max Planck Institute for Evolutionary Anthropolgy, Leipzig, Germany.

We apply a powerful haplotype-based model (described in Lawson *et al*. 2012) to infer the population history of 410 individuals from ~50 Native American groups, using data interrogated at >470,000 genome-wide autosomal Single-Nucleotide-Polymorphisms (SNPs). The model matches haplotype patterns among individuals' chromosomes to infer which individuals share recent common ancestry at each location of the genome, an approach that has previously been demonstrated to increase power substantially over widely-used alternative approaches that consider SNPs independently. We apply this methodology to 1861 samples described in Reich *et al*. (2012), incorporating 263 additional samples from 32 relevant world-wide regions collated from other publicly available resources and currently unavailable data. We utilize these methodology and data in two ways. First, we infer intermixing (i.e. "admixture") events among different Native American groups by identifying the groups that share the most haplotype segments. Using additional unpublished techniques, we determine the dates of these intermixing events, the proportions of DNA contributed, and the precise genetic make-up of the groups involved. These unique characteristics set this methodology apart from all presently available software, allowing us to place these mixing events into a clear historical context and thus identify the factors (e.g. the rise or fall of various Native American empires) that have contributed most to the genetic architecture of present-day Native American groups. Second, we match DNA patterns from each Native American group to a set of over 30 populations from Siberia and East Asia, describing each Native American group as a mixture of DNA from these regions. This enables us to shed light on the widely debated number of distinct migrations into the Americas during the initial colonization across the Bering Strait, comparing our results to previous inference from the literature. Our application demonstrates the power gained by using rich haplotype information relative to approaches that ignore this information.

1952W

Diversity of the Mexican Mestizo population using 18 X-STR. *E. Ortiz[1], G. Noris[2], C. Santana[2], M.A. Meraz[3], R. Gómez[1].* 1) Dep. de Toxicología, Cinvestav-IPN, Mexico City, Mexico; 2) Laboratorio BIMODI, Querétaro, Qro., Mexico; 3) Dep. de Biomedicina Molecular, Cinvestav-IPN, Mexico City, Mexico.

The use of the X chromosome STRs (X-STR) in population genetics, forensic studies and kinship testing have been amply used in recent years. The X-STRs shows intrinsic characteristics as lower mutation and recombination rates, and faster genetic drift due to smaller effective population size. In addition, population history and demographic factors have a deep impact on the population stratification. Previous studies have indeed highlighted the particularly complex genetic composition of the Mexican-mestizo population using Y chromosome STRs and mitochondrial DNA, however the population genetic studies using X-STR are still very limited. In order to assess the population stratification of the Mexican mestizo population we genotyped 200 unrelated women from central states of Mexico using 18 X-STR. Our results showed DXS10011 (k=32, PIC=0.94), DXS10134 (k=17, PIC=0.83), DXS10146 (k=22, PIC=0.84), and DXS10101 (k=21, PIC=0.90) were the most informative markers. Hardy-Weinberg departure was found in six out of eighteen loci studied (DXS7132: FIS=0.17061, P=0.001; DXS6809: FIS = -0.25683, P=0.001; DXS10134: FIS = -0.01425, P=0.0036; DXS10146: FIS = 0.14370, P=0.009; DXS10101: FIS =0.08344, P=0.001; DXS10079: FIS = 0.1204, P=0.0048). Gametic association was tested for all pairs of loci, however only three pairs of loci show P values below the significant level of 0.0027 (obtained after Bonferroni correction): DXS6809-DXS10134 (P= 0.00146), DXS6809-DXS8378 (P=0.00183), DXS10134-DXS10101 (P= 0.00245). The degree of population subdivision assessed using a hierarchical analysis of molecular variance (AMOVA) showed 91.59% variation within populations. In addition principal component analysis (PCA) suggest high population stratification. Our results established the first database for X-STR markers diversity for Mexican-mestizos. The relatively high number of allele encountered proved that these markers could be considered as highly informative and an important tool for human identification purposes and general anthropological research.

1953T

Native American, European and African ancestry from genotype by sequencing in Argentinean populations. *M. Muzzio[1,2,3], J.M.B. Motti[2,3], T. Cooke[1], L.S. Jurado-Medina[2], M.C. Yee[4], A. Adams[1], J. Beltramo[2], R. Santos[2,3], V. Ramallo[5], M. Schawb[2], O. Cornejo[1], G. Bailliet[2], E.E. Kenny[6,7,8,9], C.M. Bravi[2,3], C.D. Bustamante[1].* 1) Stanford University School of Medicine, Stanford, CA; 2) IMBICE CCT-La Plata CONICET- CICPBA, Argentina; 3) Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Argentina; 4) Carnegie Institution for Science, Stanford, CA; 5) Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 6) Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY; 7) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, NY; 8) The Center for Statistical Genetics, Icahn School of Medicine at Mount Sinai, NY; 9) The Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NY.

In this work we aim to describe the demographic history and population structure of Argentina, by analyzing a collection of 2904 DNA samples from 15 dispersed regions in Argentina, from the Andean populations in the Northwest to the river lands of the Northeast. Argentina has history of broad cultural diversity, which included hunter-gatherers and agro-pastoralists to the expansion of the Inca Empire, the Spanish arrival, slave trade from Africa and the European migration between the mid XIXth and XXth centuries. Thus far, we have 390 genotypes defined from the Illumina Exome Array 250K and 89 genome-wide sequences from a genotype-by-sequencing strategy we developed, which covers 1.5% of the genome. We will continue to sequence more samples from our collection of 2904 participants from 15 Argentinean populations. From our preliminary data, we have identified proportions and patterns of European, African and Indigenous American ancestry in the Argentinean cohort via local ancestry inference and found a correlation between latitude and proportion of Native American ancestry, where the highest proportion corresponds to the northernmost populations, (where it averages 75.06%) descending towards the center of Argentina (averaging 34.02%). We will focus on our sequencing data and compare it to full genomes of Native Americans, studying the genomic tracts of Native American ancestry for fine-scale examination of sub-continental structure.

1954F

The Genetics of Craniofacial Morphometry in Latin America. *K. Adhikari, A. Ruiz-Linares.* Dept. of Genetics, Evolution and Environment, University College London, London, United Kingdom.

Craniofacial morphometry has been one of the classical components of human physical anthropology. Recently there have been several studies on constructing 3-d models of the face and performing genetic associations. In this project, we bridge the two aspects of facial morphology and obtain interesting patterns with genetic ancestry.

The CANDELA (Consortium for the Analysis of the Diversity and Evolution of Latin America) project is exploring the effect of European-Native American admixture on several thousand subjects from five different countries (Brazil, Chile, Colombia, Mexico and Peru). From photographs, we construct a 3-d model of their faces with 36 landmarks. We also construct symmetrized distances, left-right deviations, and principal components of shape from these landmarks, each covering a different aspect of facial morphometry. We observe these to be strongly correlated with ancestry. In general, European ancestry contributes to a larger facial shape, just as it contributes to taller height in our subjects. The noses and the lips show the largest amount of variation according to ancestry.

Following the style of classical anthropology, we independently phenotype our subjects on various traits using ordinal categories, such as the presence of epicanthic fold or mono-brow. We match these phenotypes with the quantitative distance measurements, and they show similar associations with European ancestry, with nose and lip traits again being the strongest.

Finally, we genotype some of our Colombian samples on an Illumina 730K chip and perform a genome-wide association study on the traits. The GWAS implicates several genes such as NPAS2 and PSD3. As we genotype more samples from the other countries, we will be able to replicate the findings and probably find other associated genes as the sample size increases.

1955W

The Centenarians of Nicoya, Costa Rica: a genomic evolutionary approach. *J. Azofeifa[1], E.A. Ruiz-Narvaez[2], A. Leal[1], L. Rosero-Bixby[3, 4].* 1) Escuela de Biologia, Universidad de Costa Rica, San Pedro, San Jose, Costa Rica; 2) Slone Epidemiology Center and Department of Epidemiology, Boston University; 3) Central American Population Center, University of Costa Rica; 4) Department of Demography, University of California, Berkeley.

The disclosure of the genetic contribution to longevity through polymorphisms that delay the rates of progression of senescence has proven to be elusive partially due to the fact that most of current approaches lack an evolutionary perspective. We propose an approach based on both the antagonist pleiotropy theory, and the mutation accumulation theory to explain senescence, and consequently, longevity. Both theories allow formulating predictions that can be explored by genomic screenings. We present results of a population genetics study derived from a demographic program on healthy aging in Costa Rica (CRELES). We concentrated on the population of Nicoya, in the Costa Rican province of Guanacaste, which is characterized by an extreme longevity (according to the demographic results of CRELES, the Nicoyan nonagenarian males could have the highest life expectancy in the world). We performed a pilot genome-wide association study (GWAS) using a sample of 20 of the oldest old (99-105 years old, 11 males and 9 females), from the Nicoya region, and their respective controls (60-65 years old, 11 males and 9 females) using the Illumina Human Omni Express array (~700,000 markers). No SNP reached genome-wide significance. We then proceeded to estimate individual genetic ancestry based on 464 ancestral informative markers included in the Illumina array to assess its relationship with longevity. Amerindian ancestry was higher in centenarians compared to non-centenarians (43.3% vs. 36.0, $p = 0.027$). We estimated that by each 5% increment of Amerindian ancestry was associated with a 49% higher probability of being a centenarian, OR (95% CI) = 1.49 (1.01-2.23), $p = 0.04$. These results suggest the existence of Amerindian-specific alleles that are associated with extreme longevity, and offer evidence in favor of the mutation accumulation theory, which has lacked support because it is difficult to test empirically. An explanation could be that negative effects of recessive late-acting alleles, accumulated in different populations through long periods of genetic isolation, are masked by heterozygosity. The importance of ancestry and the historical and geographical origin of parental populations have, to our knowledge, never been considered in the context of the origins of senescence and longevity. More research is needed in larger sample sizes to advance the promising results of this pilot study.

1956T

Population Structure and Genetic Diversity in a Population of 15,000 Patients from East Harlem, NY. *G. Belbin[1], D. Ruderfer[2,3,4], E.A. Stahl[2,3,5,6], J. Jeff[5], Y. Lu[5], R.J.F. Loos[5,7], O. Gottesman[5], S. Purcell[2,3,4,5], E. Bottinger[5], E.E. Kenny[1,4,5,6].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Broad Institute, Cambridge, MA; 3) Division of Psychiatric Genomics, Icahn School of Medicine at Mt Sinai, New York, NY; 4) Center for Statistical Genetics, Icahn School of Medicine at Mt Sinai, New York, NY; 5) Institute for Personalized Medicine, Icahn School of Medicine at Mt Sinai, New York NY; 6) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mt Sinai, New York, NY; 7) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mt Sinai, New York, NY.

New York City has historically been a significant point of entry for immigration into the United States and as a consequence is today peopled by a highly structured and ethnically diverse population. Census ethnic labeling reveals some of this diversity, but does not fully capture the variety of cultural groups, with complex and diverse demographic origins, foods and traditions, living in New York. Using genome-wide data, it is possible to detect such population structure which can both inform population history inference and result in better outcomes for medical genetics studies. We present a diversity of approaches for the analysis of fine-scale population structure in a population of 29,093 patients enrolled in the Icahn School of Medicine BioMe Biobank Cohort (BioMe), of which ~13,500 have available Illumnia Omni Express and Exome Chip data (~900K SNPs). BioMe comprises 34%, 47% and 19% participants with self-reported African-American (AA), Hispanic-Latino (HL) and European-American (EA), respectively, and is representative of the population of Northern Manhattan. We combined these data with both data generated from the 1000 genomes project and an additional unique database of genomic variation in over 4,000 individuals from diverse European, Middle Eastern, East Asian, African and Native American populations. Population genetic analysis using standard Principle Component Analysis (PCA) and ADMIXTURE, and a novel ancestry-specific PCA method using Native American, European and African local ancestry haplotypes from AA and HL genomes, reveals diverse sub-continental structure in the BioMe cohort. In particular, we detect a large proportion of Ashkenazi Jewish and Eastern European ancestry in the BioMe EAs. We also performed identity-by-descent (IBD) analysis and detect elevated cryptic relatedness in the AAs and HLs, which results in increased genetic tract sharing compared to EAs. For example, analysis of IBD haplotype sharing between any two less-than-fourth-degree relatives in our cohort indicates a larger percent of their genome is shared (~0.4% and 3.72%, for AA and HL, respectively), compared with the same analysis in EA (~0.12%).

1957F

Ultra fast and sample-aware local ancestry inference using population specific variants. *R.P. Brown[1], B. Pasaniuc[1,2].* 1) Bioinformatics, UCLA, Los Angeles, CA; 2) Pathology and Laboratory Medicine, Geffen School of Medicine, Los Angeles, CA.

Inferring the ancestry at each genomic locus of recently admixed individuals (e.g. Latino Americans) plays a key role in medical and population genetic inferences from finding disease risk loci to inferring recombination rates. Current local ancestry methods are designed for genotyping arrays without utilizing the full spectrum of data available from sequencing and are very computationally expensive. In addition, existing methods do not make use of all the admixed samples when calling ancestry for a given individual and only rely on external reference panels used as proxies for the true ancestral populations. We present a fast and accurate method using population specific variants (PSVs) (i.e. variants identified by sequencing to be present in only one continental population) to infer local ancestry in sequenced admixed genomes. We use the real 1000 Genomes data to find an abundance of such PSVs (e.g. an average of ~13.6 informative PSVs per Mb in Europeans, 109 in Africans) and model them within standard Hidden Markov Models for local ancestry inference to achieve an ultra-fast and accurate approach. Our method includes an iterative framework that rebuilds its reference panels from the confidently called ancestry segments in the admixed individuals themselves to further boost accuracy while reducing bias introduced by the reference panels. Using simulations of Puerto Ricans (Mexicans) from the 1000 Genomes, we show that our method achieves an accuracy of 0.93 (0.91) (quantified by the average r2 between inferred and simulated ancestry) as compared to 0.91 (0.90) achieved by existing methods. Most importantly, our approach is orders of magnitude faster than existing methods for full sequencing data (e.g. our method can infer local ancestry in 10,000 individuals in ~18 CPU days as compared to ~8 CPU years for existing methods). We also explore whether similar results can be attained in real data and show that our approach yields comparable local ancestry to the provided calls in the real Puerto Rican and Mexican individuals from 1000 Genomes. We extend our approach to low coverage sequencing and show that accurate local ancestry inference is attained at low coverage (e.g. r2 of 0.73 in our Puerto Ricans simulations at 4x coverage). Finally, we extend our method to GWAS-array genotyped individuals using a PSV tagging procedure that achieves accuracy similar to when full sequence data is available (e.g. 0.94 in African American simulations).

1958W

Y chromosomes in surname samples: insights into surname frequency and origin. *F. Calafell, N. Solé-Morata, J. Bertranpetit, D. Comas.* Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Barcelona, Spain.

In most societies, surnames are inherited through the paternal line, exactly as Y chromosomes are, with the exceptions of adoption, false paternity, and the inheritance of the maternal surname (as in the case of single mothers). The well-established phylogeography of the Y chromosome and the existence of fast-evolving STR markers implies that genotyping the Y chromosomes in samples of men bearing the same surname allows to i) count the number of founders of a surname and the frequency of their descendant in the current population and ii) pinpoint the remote origins of the founder of a surname. We have collected >2,500 samples from volunteers bearing one of 50 different Catalan surnames (http:// http://cognoms.upf.edu/), in which we are genotyping 17 Y-chromosome STRs and a custom-designed set of 64 SNPs (typed in a single reaction with the OpenArray™ Real-Time PCRplatform), with the following objectives: i)Discover and quantify the processes that drive surname frequency. Surname polyphiletism (which can be measured from the number of founders detected from the Y chromosome diversity) can drive surname frequency, as was found in a sample of English surnames, but drift and natural selection (associated with high-status surnames) may also have a role. 2) Were the founders of surnames that are linguistically Arab or Hebrew North Africans or Jews themselves? 3) Were the founders of Germanic patronymic surnames of a different genetic origin form the rest of the population? In Catalonia, as in France, a frequent source of surnames are former first names of Germanic origin (*Albert*, *Robert*, *Grau*,...). We will compare some of those to patronymic surnames of Latin origin (*Oriol*, *Pons*,...). Haplogroup frequency differences between Germany and Catalonia provide sufficient power for this comparison. 4) Does an ethnonymic surname indicate a foreign origin? Some Catalan surnames (*Alemany*, *Danés*, *Guasch*) denote geographic origin (they mean *German*, *Dane*, *Gascon*, respectively). We have completed this objective and we have found that haplogroups that are more frequent in Germany or Gascoigne than in Catalonia were not overrepresented in the founders of the *Alemany* and *Guasch* surnames. Thus, these surnames may have originated as bynames not necessarily linked with the origins of the founders.

1959T

The Population Genetics of Sub-Saharan African Populations. *M. Capredon[1], J. Hussin[1,2], J. Quinlan[1,3], Y. Idaghdour[1], T. de Malliard[1], JC. Grenier[1], V. Bruat[1], E. Gbeha[1], L. Barreiro[1], P. Awadalla[1].* 1) Pediatrics, CHU Sainte Justine, Fac Med, Univ Montreal, Montreal, Quebec, Canada; 2) Biochemistry, Faculty of medicine, , University of Montreal, Montreal, Quebec, Canada; 3) School of public health, Faculty of medicine, University of Montreal, Montreal, Quebec, Canada.

Characterizing genetic diversity in African populations is critical for population genetics and personalized medical-genetic studies. We describe 14 populations, including four novel African datasets, two Western Africa (Cotonou and Zinvie) populations that are geographically very close, and two from Eastern Africa (Batwa pygmies and Kiga from Uganda). We compared these populations to ten published datasets from Africa Hapmap phase 3 (YRI, LWK, MKK) and HGDP (Bantu from Kenya, Bantu from South Africa, Biaka, Mandenka, Mbuti, San and Yoruba) across Africa. We also used a European dataset (Hapmap phase 3 CEU) as controls. Population structure analysis (PCA and admixture), revealed that the Kiga population is genetically very close to the LWK and to the Bantus from HGDP, while populations from Cotonou and Zinvie are clustering closely with the Yoruban and the Mandenka. Although the Zinvie population is geographically very close to Cotonou, clear differentiation was highlighted. Finally, the Batwa pygmies are substantially differentiated genetically from other populations. We observed that variability exists in recombination rates among African populations; sometimes this variability was associated with a geographic cline in Africa, but also was observed among more closely related populations (Yorubans from Nigeria versus Yorubans from Benin).

1960F

A scalable and effective local ancestry deconvolution algorithm for Latinos. *G. Genovese[1,2], A.L. Williams[2], S.A. McCarroll[1,2].* 1) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 2) Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

Genomes of admixed individuals are a mosaic of genetic segments inherited from different ancestral populations. Local ancestry deconvolution is the process of statistically inferring the ancestral source population of all sites in the genome of admixed individuals. Most local ancestry deconvolution algorithms are designed to work with SNP array genotype data and require reference panel data that are good representatives of each ancestral population. Obtaining reference panels may be especially challenging for Native American groups, and small or inaccurate panels have the potential to bias current methods. We developed a new algorithm (LATOOLS) for Latino local ancestry deconvolution which utilizes whole genome sequence data and does not require reference panels. The algorithm works in two steps. The first step identifies what we termed signature alleles: alleles that are present to an appreciable frequency in at least one ancestral population, but not in all ancestral populations for which the local ancestry deconvolution is computed. The second step uses these signature alleles to infer local ancestry in a given genome. Thus only sites that are fully informative for ancestry are incorporated into the model. While this method does not utilize all available information in its inference, our experimental results demonstrate that this approach is effective for Latinos. Furthermore, because of its relative simplicity, the algorithm is fast and scalable: its run time is linear in signature alleles count and cohort size. We show that in Latino samples from the 1000 Genomes project we are able to achieve, once several potential sources of sequencing biases are carefully removed, over 90% concordance with ancestry calls obtained using more sophisticated algorithms. Finally, we briefly show an application of the obtained local ancestry deconvolutions for the problem of admixture mapping the missing pieces of the human genome reference.

1961W

A two-sex model for the admixture history of a hybrid population: the X-chromosome. *A. Goldberg[1], P. Verdu[2], N.A. Rosenberg[1].* 1) Biology, Stanford University, Stanford, CA; 2) CNRS-MNHN, UMR7206 Eco-Anthropology and Ethno-Biology, Paris France.

Complex social practices such as residence and descent rules, polygyny, and dominance relationships can produce sex-biased demographic histories in admixed human populations. Genetic signatures of sex-biased admixture have been empirically detected in samples from a variety of human populations throughout the world, usually without mechanistically modeling the full complexity of the sex-specific history. Expanding on the model of Verdu & Rosenberg (2011), we developed a model that mechanistically considers sex-specific admixture histories for autosomal DNA. Here, we extend our model to the sex chromosomes, allowing multiple source populations to contribute to the admixed population, potentially with varying contributions from male and female lineages across generations. Under the model, the X-chromosome is particularly informative about sex-biased admixture, as, unlike the autosomes, it can be used to identify which sex has a greater contribution to the admixed population. With no sex bias, the ratio of the expected X-chromosomal ancestry fraction to the expected autosomal ancestry fraction is one, but this ratio deviates from one with increasingly sex-biased admixture. The variance of the ancestry fraction is larger for a random X-chromosome sampled from a male from the admixed population rather than from a random female X-chromosome, in line with the decreased X-chromosomal population size for males compared to females. For admixture processes that are constant in time, we perform approximate Bayesian computation to infer sex-specific contributions from each source population. Considering autosomal DNA and the X-chromosome together, our approach can contribute to methods for inference of complex sex-biased admixture histories.

1962T

Is Genetic Ancestry Associated with Sleep? *I. Halder[1], M. Hall[2], D. Buysse[2], S.E. Reis[1], K. Matthews[2].* 1) Dept. of Medicine, Univ Pittsburgh, Pittsburgh, PA; 2) Dept. of Psychiatry, University of Pittsburgh, Pittsburgh, PA.

Background: Sleep characteristics vary by race. Compared to Whites, African Americans (AA) take longer to fall asleep, have lower sleep efficiency, lesser sleep duration and spend less time in the restorative slow wave sleep stage. The causes underlying these differences are not fully known but substantial heritability of sleep suggests some genetic underpinnings. We hypothesized that an overall genetic effect on sleep might be observed in AAs who have varying proportions of African genetic ancestry. We tested whether sleep phenotypes show an association with objectively measured individual genetic ancestry. **Methods:** The sleep of 101 Whites and 71 AA adults (mean age 58 ± 6 years; 57% female) was examined as part of the University of Pittsburgh SleepSCORE project. Polysomnography was used to assess average sleep duration, efficiency, Apnea-hypopnea index (AHI) and slow wave sleep (delta band of spectral EEG analyses). The Pittsburgh Sleep Quality Index (PSQI) was used to measure sleep quality. Biometric, psychosocial factors, health behaviors, and environmental factors were measured in all subjects. 1600 ancestry informative genetic markers were genotyped in all AAs which were used to infer overall individual African ancestry (IAA) by a maximum likelihood method. Multivariate tests were used to compare sleep phenotypes between Whites and AAs. Next, in AAs only, hierarchical linear regression was used to test whether IAA predicts sleep duration, efficiency, quality, architecture and AHI. Age, gender and BMI, education, income, use of sleep medication and antidepressants were used as covariates. All covarites were entered in step one, AHI (when not tested as the outcome) was entered in step two and IAA was entered in step three of the model. **Results:** Sleep duration, efficiency and architecture differed significantly between Races. In AAs, IAA ranged from 10%-88% with a mean of 67%. Higher African genetic ancestry was associated with lower slow wave sleep in AAs (β (SE) = -4.6 (1.5); P = 0.002). This association persisted after all covariate adjustment. After adjusting for all relevant covariates individual African genetic ancestry explains 11% of the variation in slow wave sleep in AAs. Sleep duration, efficiency, quality or AHI was not associated with African genetic ancestry. **Conclusion:** Racial differences in slow wave sleep (but not other sleep phenotypes) appear to be explained partly by genetic variation related to continental ancestry in AAs.

1963F

Evidence of social marginalisation leading to strong genetic differentiation among the Ari of Ethiopia. *G. Hellenthal[1], L. van Dorp[1], S. Myers[2], L. Pagani[3], C. Tyler-Smith[3], E. Bekele[4], A. Tarekegn[4], M. Thomas[1], N. Bradman[5], D. Balding[1].* 1) Genetics, Evolution, and Environment, University College London, London, United Kingdom; 2) Department of Statistics, University of Oxford, Oxford, United Kingdom; 3) The Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 4) Addis Ababa University, Addis Ababa, Ethiopia; 5) Henry Stewart Talks Ltd, London, United Kingdom.

A major debate among population geneticists is whether the large amount of observed genetic diversity among human groups (e.g. from Africa) is primarily attributable to ancient substructure or to more recent inter-mixing or drift events. A specific example involves the caste-like occupational groups within the Ari of Ethiopia, which include the Cultivators and socially marginalized Blacksmiths. There are two competing theories about the origins of the Ari Blacksmiths as (i) remnants of an ancient hunter-gatherer population that inhabited Ethiopia prior to the arrival of agriculturists (e.g. Cultivators) during the Neolithic period, versus (ii) relatively recently related to the Cultivators but presently marginalized in the community due to their craft skills. A recent paper by Pagani et al (2012) collected and analysed genome-wide DNA from samples of Ari Blacksmiths and Cultivators, observing genetic differentiation between these two occupational groups (F_{ST} = 0.04) at a similar level to that observed across multiple ethnic groups sampled throughout Ethiopia (F_{ST} range 0.02 - 0.06). Furthermore, an analysis using the clustering algorithm ADMIXTURE assigned the Ari Blacksmiths almost entirely to a single genetically homogeneous cluster, with other Ethiopian groups -- including the Cultivators -- displaying varying levels of genetic similarity to this unique cluster. The authors noted these patterns were consistent with model (i), with subsequent assimilation of the indigenous hunter-gatherers into the expanding agriculturalist community. We analysed the same samples using a novel haplotype-based approach to test this hypothesis, comparing strings of DNA patterns among different Ethiopian groups and to various groups outside of Ethiopia. Importantly, this new model is able to distinguish genetic structure attributable to allelic drift within a population from that attributable to shared ancestry with outside groups. Using this technique, we provide compelling evidence that the genetic differences among Ari Blacksmiths and Cultivators are entirely attributable to recent social marginalisation, i.e. hypothesis (ii). This finding serves as both a cautionary tale about interpreting results from the widely popular program ADMIXTURE, and a clear demonstration of how social constructions can contribute directly to genetic differentiation among previously similar groups.

1964W

Pinpointing the Indian origin and revealing the Caucasus chapter in the genetic ancestry of the European Roma. *M. Karmin[1], M. Baldovič[2], N. Jeran[3], M. Reidla[1], S. Cvjetan[4], S. Rootsi[1], T. Šaric[3], J. Šarac[3], M. Cenanovic[5], T. Haller[6], A. Raidvee[7], R. Mägi[6], A. Leskovac[8], L. Kovacevic[5], D. Marjanovic[5], H.D. Auguštin[3], N. Novokmet[3], A. Ficek[2], G. Chaubey[1], P. Rudan[3], V. Ferak[2], E. Metspalu[1], M.D.M. Behar[1,9], M. Metspalu[1], R. Villems[1].* 1) Intitute of Molecular and Cellular Biology, University of Tartu and Estonian Biocenter, Tartu, Estonia; 2) Department of Molecular Biology, Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia; 3) Institute for Anthropological Research, Zagreb, Croatia; 4) Mediterranean Institute for Life Sciences, Split, Croatia; 5) Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina; 6) Estonian Genome Center, University of Tartu, Tartu, Estonia; 7) University of Tartu, Faculty of Social Sciences and Education, Department of Psychology, Tartu, Estonia; 8) Vinca Institute of Nuclear Sciences , Belgrade, Serbia; 9) Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa, Israel.

According to linguistic evidence, the Indian exodus of the ancestors of the European Roma most probably took place around the end of the first millennium. By the 13th - 15th centuries, different groups of the Roma had spread throughout Europe. A virtual lack of written records prior to their arrival to Europe has left us with scarce knowledge about their historical migratory routes. Therefore, valuable insight comes from genetic studies. The origin of the maternal (mitochondrial DNA - mtDNA) and paternal lineages (Y chromosome - NRY) of the European Roma can be broadly classified as South Asian and West Eurasian. Out of the nearly 600 mtDNAs and 340 NRY lineages studied here, about 30% of the gene pool of both maternal and paternal lineages of the European Roma were classified as South Asian. To analyze the genetic ancestry of the Roma on the genome-wide level we genotyped ca 650 000 SNPs in 30 Roma individuals from six European countries with Illumina genotyping arrays. STRUCTURE-like analyses and principal component analyses (PCA) were performed in the context of Eurasian populations (N=1219). These analyses showed that in general the European Roma have a three-part genetic ancestry, with components from South Asia, the Caucasus area and Europe. Notably, on average, 38% of each Roma genome comprises the genetic component currently most abundant among populations in the Caucasus region, revealing a genetic legacy that has not received much attention until now. To further pinpoint the Indian origin of the Roma we used two methods based on haplotype data and a pre-defined reference population. 1) With the local ancestry deconvolution method we assigned the parts of genome with South Asian ancestry, and further used information only from those parts for PCA. 2) We used a novel in-house method assigning best-fit ancestry proportions for individuals as amalgamates of ancestral populations. The Roma formed a cluster bordering and overlapping with individuals from Uttar Pradesh and several South Indian states on PCA results, and the best-fit method revealed that the Indian ancestry of the Roma comes primarily from the Southern Indian populations. Thus, our results reveal a substantial amount of ancestry shared with populations from the Caucasus and several lines of evidence suggest a more southern origin of the Roma within India than previously thought.

1965T

Reconstructing Austronesian population history. *M. Lipson[1], P.-R. Loh[1], N. Patterson[2], P. Moorjani[2,3], Y.-C. Ko[4], M. Stoneking[5], B. Berger[1,2], D. Reich[2,3].* 1) Department of Mathematics and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Graduate Institute of Clinical Medical Science, China Medical University, Taichung, Taiwan; 5) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Present-day populations that speak Austronesian languages are spread across half the globe, from Easter Island in the Pacific Ocean to Madagascar in the Indian Ocean. Evidence from linguistics and archaeology suggests that the "Austronesian expansion," a vast cultural and linguistic dispersal that began 4--5 thousand years ago, had its origin in Taiwan. However, genetic studies of Austronesian ancestry have been inconclusive, with some finding affinities with aboriginal Taiwanese, others advancing an autochthonous origin within Island Southeast Asia, and others proposing a model involving multiple waves of migration from Asia. Here, we analyze genome-wide data from a diverse set of 31 Austronesian-speaking and 25 other groups typed at 18,412 overlapping single nucleotide polymorphisms (SNPs) to trace the genetic origins of Austronesians. We use a recently developed computational tool for building phylogenetic models of population relationships incorporating the possibility of admixture, which allows us to infer ancestry proportions and sources of genetic material for 26 admixed Austronesian-speaking populations. Our analysis provides strong confirmation of widespread ancestry of Taiwanese origin: at least a quarter of the genetic material in all Austronesian-speaking populations that we studied---including all of the Asian ancestry in populations from eastern Indonesia and Oceania---is more closely related to aboriginal Taiwanese than to any populations we sampled from the mainland. Surprisingly, we also show that western Austronesian-speaking populations have inherited substantial proportions of their Asian ancestry from a source that falls within the variation of present-day Austro-Asiatic populations in Southeast Asia. No Austro-Asiatic languages are spoken in Island Southeast Asia today, although there are some linguistic and archaeological suggestions of an early connection between mainland and island populations. The most plausible explanation for these findings, in light of the historical evidence, is that western Island Southeast Asia was settled by Austronesian groups who had previously mixed with Austro-Asiatic speakers on the mainland.

1966F

Patterns of genetic variation in populations of African ancestry observed in whole genome sequencing of 691 individuals from CAAPA.
R. Mathias[1], L. Huang[1], T.D. O'Connor[2,3], C. Vergara[1], M. Taub[1], A. Deshpande[4], C.R. Gignoux[5], N. Rafaels[1], S. Shringarpure[6], R. Torres[5], J. Galanter[5], R. Hernandez[5], E.E. Kenny[7], D. Locke[4], W. Grus[4], K. Gietzen[8], I. Ruczinski[1], K.C. Barnes[1], CAAPA Consortium. 1) Johns Hopkins University, Baltimore, MD; 2) University of Washington, Seattle, WA; 3) University of Maryland, Baltimore, MD; 4) Knome, Inc., Cambridge, MA; 5) University of California at San Francisco, San Francisco, CA; 6) Stanford University School of Medicine, Palo Alto, CA; 7) Icahn School of Medicine at Mount Sinai, New York, NY; 8) Illumina, Inc., San Diego, CA.

The **C**onsortium on **A**sthma among **A**frican-ancestry **P**opulations in the **A**mericas (CAAPA) includes high coverage whole genome sequence (WGS) data (~30x depth) on 1,005 subjects of African ancestry and extends the patterns of variation catalogued in the Thousand Genomes Project (TGP) and Exome Sequencing Project (ESP) to a spectrum of populations representing a wide range of African ancestry. An interim data freeze (N=691) of CAAPA includes: 329 African Americans from 8 sites in the United States (Chicago, Atlanta, Baltimore/Washington, Nashville, New York City, Detroit, San Francisco, and Winston-Salem); 125 African Caribbeans from three sites (Barbados, Jamaica and Honduras); 212 African ancestry samples with a notable Latino component sampled from 5 Central and South American sites (Cartagena, Conde and Salvador in Brazil, Dominican Republic, and Puerto Rico); and 25 samples from Nigeria. The average Yoruba (YRI) component in CAAPA is 58%, ranging from <40% in the Puerto Rican samples to >80% in the Jamaican samples. Principal components and FST analysis reveals minimal differences between the 8 African American sites. Of the 47.4 million (M) variants observed in CAAPA, a little more than half (~24M) are unique to CAAPA and not observed in TGP sequence data. 68.7% (~33M) have an MAF≤1%, 13.8% (~6.6M) have an MAF 1-5% and only 17.5% (~8.4M) are common (MAF>5%). A dramatic 37% (~18M) sites represent private variation. Per genome, we observe ~4.1M variant sites and observe strong correlation between variation per genome with level of African ancestry (r=0.97 with YRI content). Exome variation in CAAPA is similar to observations in the ESP. Relying on transcription factor binding motifs, DNase-seq, FAIRE-seq, and CHIP-seq data from ENCODE, we annotated each CAAPA genome and observe ~31K variants mapping to CTCF enriched elements, ~23K predicted enhancer site variants, ~3K sites in predicted promoter flanking regions, ~407K sites in predicted transcribed regions, ~13K sites in predicted promoter regions including TSS, and ~12K sites in predicted weak enhancer or open chromatin cis regulatory elements per genome. WGS data from CAAPA will provide an expansive understanding of genetic variation in populations of African ancestry and will be available as a public resource to the scientific community to improve our understanding of the role of genetic variation in complex disease.

1967W

Correlation of Native American Ancestry with Body Mass Index in an Admixed Community.
T.M. Norden-Krichmar[1], I.R. Gizer[2], O. Libiger[1], K.C. Wilhelmsen[3], C.L. Ehlers[4], N.J. Schork[1]. 1) Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA; 2) Department of Psychological Sciences, University of Missouri, Columbia, MO; 3) Department of Genetics and Neurology, University of North Carolina, Chapel Hill, NC; 4) Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, La Jolla, CA.

Obesity is highly prevalent in Native Americans (NAs). Because of this, the analysis of ancestral background is essential in association studies of obesity phenotypes in genetic studies of admixed populations with NA ancestry because population stratification can confound the results. In addition, there might be a set of polygenes present in the NA ancestral population that contributes to gradations in obesity phenotypes among admixed individuals with different levels of admixture. We explored the relationship between genetic admixture, cultural and socioeconomic factors, and body mass index (BMI) in a community sample of Native Americans with a high prevalence of obesity (n=822). We estimated genetic admixture by self-report as well as genetic marker-based supervised and unsupervised algorithms determined by: 1) Illumina low-coverage (3X to 20X) whole genome sequencing (WGS) and 2) genotype data from an Affymetrix Exome1A chip. The ancestry estimates from the different genotyping methods had a very high Pearson's correlation of 0.9. Linear regression was performed between the morphological phenotypes of BMI, height, and weight and the estimated percent Native American ancestry. BMI and weight showed a significant positive correlation with Native American ancestry, whereas height showed a negative correlation. Age was also strongly associated with BMI, while gender and socioeconomic measures, such as education and income, were not. Two different cultural identity measures were also used to test for correlation with BMI: the Orthogonal Cultural Identification Scale (OCIS) and the Indian Culture Scale (ICS). The Indian Culture Scale (ICS) showed a positive correlation with BMI , but the Orthogonal Cultural Identification Scale (OCIS) showed no significant correlation. Heritability (h2) estimates for BMI were also calculated from linkage analysis and from the genotyping data to be between 0.38 - 0.49, suggesting a high heritability of this phenotype. These results suggest that genetic and cultural environmental factors influence BMI in this population. Additionally, this study suggests that admixture-mapping may be a successful approach in uncovering genetic variants associated with obesity and obesity-related illnesses in this population.

1968T

Deep coverage Bedouin genomes reveal Bedouin haplotypes shared among worldwide populations in the 1000 Genomes Project.
J.L. Rodriguez-Flores[1,2], K. Fakhro[3], F. Agosto-Perez[2], A. Robay[3], R.G. Crystal[1], J.G. Mezey[1,2]. 1) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY; 2) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Department of Genetic Medicine Weill Cornell Medical College - Qatar, Doha, Qatar.

The 1000 Genomes Project (1000G) has sampled and sequenced over 2500 genomes that are representative of the genetic diversity in populations worldwide. The Arabian Peninsula has not been previously included in 1000G, hence the connections between genetic variation in the indigenous Bedouin people and worldwide populations is unknown. We have sampled genomes from Bedouin individuals in the nation of Qatar as a window into the genetic variation in this understudied region. Our goal was to use this sample to assess the hypothesis that there is detectable shared ancestry between Bedouin and Southern European populations resulting from the history of empires that spanned both the Mediterranean and Arabian regions and the hypothesis that there is shared ancestry between Bedouin and contemporary Latin American populations, since the majority of European settlers in Latin America from the past half millennia are primarily from Southern European countries. We selected 60 Qataris with over 95% Bedouin ancestry and at least 3 generations of ancestry in Qatar for deep coverage genome sequencing. Genomes were sequenced by the Illumina Genome Network using TruSeq DNA PCR-free sample preparation, generating over 120 gigabases of paired-end 100 base pair reads per genome on a HiSeq 2500, yielding over 30x depth and genotypes for >96% of the genome using both the ELAND/CASAVA and BWA/GATK pipelines. Using these genotypes, we inferred haplotypes using SHAPEIT for Bedouin Qataris and for 1000G populations on a set of sites polymorphic in both 1000G and Bedouins. We used admixture analysis to assess shared ancestry between our Bedouin sample and 1000G populations using the ancestry deconvolution method SUPPORTMIX. Given the lack of appropriate ancestral populations, we conducted a leave-one-out approach, where for each population (1000G + Bedouin = n), we removed the population and used the remaining n-1 populations as an ancestral reference panel. Using this approach, we observed up to 15% Bedouin ancestry in European, South Asian, and American populations. Likewise, we observed ancestry from Europe, South Asia, and America in the Bedouin. For individuals from the Americas, the analysis identified a considerable number of segments shared with Bedouins previously classified as European ancestry.

1969F

Fossil free sequencing of archaic hominin metagenomes. *B. Vernot, JM. Akey.* Department of Genome Sciences, University of Washington, Seattle, WA.

To date, the genetic analysis of extinct archaic hominins has required the isolation and sequencing of ancient DNA obtained from fossilized remains, which is technologically challenging and limited by the number of available specimens. Here, we describe an alternative, fossil free paradigm, to sequence archaic genomes. The rationale of this approach is that ~3% of non-African genomes are estimated to have been inherited from Neanderthal ancestors. However, the precise introgressed sequences will vary among individuals. Thus, if enough individuals are analyzed, the various haplotypes of archaic sequence can be identified and stitched together to generate what we refer to as an 'archaic metagenome', as it consists of lineages from multiple archaic ancestors. To enable fossil free sequencing of archaic metagenomes, we developed a novel and computationally efficient statistic to identify introgressed DNA sequences, which is agnostic to the availability of an archaic reference sequence and can therefore facilitate the discovery of previously unknown archaic hominins, if such groups exchanged genes with modern humans. We rigorously evaluated the power and false discovery rate of our method through extensive coalescent simulations under a wide variety of demographic models and admixture scenarios. We applied our method to 379 European whole genome sequences, and identified a total of 5.2 Gb of putatively introgressed sequence, covering 1.1 Gb of the genome. These sequences are significantly enriched for matches to the high-coverage Neanderthal genome, supporting the hypothesis that they were inherited from admixture events with Neanderthal ancestors. Moreover, they are significantly depleted among coding regions, suggesting a fitness cost to hybridization. The recovered sequences also enable a variety of population genetics inferences to be made, such as effective population sizes, time of introgression, and number of archaic ancestors. In summary, we find that approximately 1/3 of the Neanderthal genome survives in modern humans and we anticipate that fossil free sequencing of archaic genomes will be a significant advance for the burgeoning field of paleogenomics, allowing genetic analyses that have heretofore not been possible.

1970W

No indication of Khazar genetic ancestry among Ashkenazi Jews. *M. Metspalu[1,13,14], D.M. Behar[2,1,14], Y. Baran[3], S. Rosset[4], N. Kopelman[5], B. Yunusbayev[1,6], A. Gladstein[7], M.F. Hammer[7], S. Tzur[2], E. Halperin[3,8,9], K. Skorecki[2,10], R. Villems[1,11], N.A. Rosenberg[12].* 1) Evolutionary Biology, Estonian Biocentre & Tartu Univ, Tartu, Estonia; 2) Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa 31096, Israel; 3) The Blavatnik School of Computer Science, Tel Aviv University, Tel-Aviv 69978, Israel; 4) Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; 5) Porter School of Environmental Studies, Department of Zoology, Tel Aviv University, Tel Aviv 69978, Israel; 6) Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa 450054, Russia; 7) ARL Division of Biotechnology, University of Arizona, Tucson, Arizona 85721, USA; 8) Department of Molecular Microbiology and Biotechnology, George Wise Faculty of Life Science, Tel-Aviv University, Tel-Aviv 69978, Israel; 9) International Computer Science Institute, Berkeley, California 94704, USA; 10) Ruth and Bruce Rappaport Faculty of Medicine and Research Institute, Technion-Israel Institute of Technology, Haifa 31096, Israel; 11) Estonian Academy of Sciences, Tallinn 10130, Estonia; 12) Department of Biology, Stanford University, Stanford, California 94305, USA; 13) Department of Integrative Biology, University of California Berkeley, 94720, USA; 14) these authors contributed equally.

The origin and history of the Ashkenazi Jewish population have long been of great interest. Most studies have concluded that the population derives its genetic ancestry from both Europe and the Middle East, and that it retains high genetic similarity to other Jewish groups such as the Sephardi Jews in Europe and Jewish communities in Northern Africa. It has recently been claimed, however, that a large part of the ancestry of the Ashkenazi population originates with the Khazars, a conglomerate of multi-ethnic, mostly Turkic-speaking tribes who consolidated into a powerful state just north of the Caucasus mountains between ca. 1,400 to 1,000 years ago. It has been difficult to explicitly test for Khazar contributions into Ashkenazi Jews, because it is not clear which extant populations can be used to represent modern descendants of the Khazars, and because the proximity of the southern Caucasus region to the Middle East makes it difficult to attribute any potential signal of Caucasus ancestry to Khazars rather than Middle Eastern populations. Here, we assemble the largest sample set available to date for assessment of Ashkenazi Jewish genetic origins, containing genome-wide single-nucleotide polymorphism data in 1,774 samples from 107 Jewish and non-Jewish populations that span the possible regions of potential Ashkenazi Jewish ancestry: Europe, the Middle East, and 15 populations from the region historically associated with the Khazar kingdom at its peak. Employing a variety of standard techniques for the analysis of population structure, we find that Ashkenazi Jewish samples share the greatest genetic ancestry with other Jewish populations, and among non-Jewish populations, with groups from Mediterranean Europe and the Middle East, and that they have no particular signal of genetic sharing with populations from the Caucasus. Thus, analysis of the most comprehensive set of Jewish and other Middle Eastern and European populations together with a large sample from the region of the Khazar kingdom does not support the hypothesis of a significant contribution of the elusive Khazars into the gene pool of the Ashkenazi Jews.

1971T

The Brazilian EPIGEN Initiative: admixture, history and epidemiology at high resolution. *A.C. Pereira[1], M.L. Barreto[2], B.L. Horta[3], M.F. Lima-Costa[4], A. Horimoto[1], N. Esteban[1], F.S.G. Kehdy[5], W.C.S. Magalhães[5], M.R. Rodrigues[5], M. Gouveia[5], M. Machado[5], R. Moreira[5], J.M. Sanches[1], H. Santos[1], F. Soares[1], G.B. Soares-Souza[5], T. Muniz[5], H. SantAnna[5], E. Tarazona-Santos[5], The Brazilian EPIGEN Consortium.* 1) INCOR, Universidade de São Paulo, Brazil; 2) Universidade Federal da Bahia, Brazil; 3) Universidade de Pelotas, Brazil; 4) Fundacão Oswaldo Cruz, Centro de Pesquisa René Rachou, Brazil; 5) Universidade Federal de Minas Gerais, Brazil.

As part of the largest Latin American initiative in population genomics and genetic epidemiology, we studied three Brazilian longitudinal populational cohorts: Salvador (n=1309), Bambui (n=1442) and Pelotas (n=3736) from Northeast, Southeast and Southern Brazil respectively. We genotyped the Omni 2.5 M Illumina array for the 6487 individuals, the Omni 5.0 M Illumina array for 265 individuals and sequenced 30 complete genomes (average coverage: 42X). EPIGEN individuals show a very large extent of individual variability in ancestry proportions. While Native American ancestry was low (5-7% at population level, with no individual with > 30% of this ancestry), the three populations showed individuals with all possible combinations of African and European ancestry. At population level, African ancestry ranged from 14-15% in Pelotas and Bambui to 51% in Bahia. Our unprecedented high resolution analysis of population structure of Brazilians in the context of worldwide variation shows that African and European ancestral subcomponents of Brazilians differ from African-Americans, reflecting the prevalent contribution from Mediterranean countries and African immigration from different geographic areas such as Mozambique and Angola. Our large and highly admixed dataset gives us high power to make inferences based on the distribution of local chromosome ancestry. We are currently inferring the dynamics of the demographic admixture process in different parts of Brazil, as well as the time and mode of arrival to Brazil of clinically relevant mutations. We are also identifying regions with significant excess/deficit of European, African or Native American ancestry, to identify candidate regions to be affected by Post-Columbian natural selection. Moreover, our high-quality sequencing data allowed us to identify between 3.6 M and 4.4 M of autosomal SNPs per each whole-genome sequenced individual, and the high levels of diversity and admixture of the Brazilian population allowed us to identify around 2.3 M of new autosomal SNPs, most of them rare. We are separating the African, European and Native American constituents of the 30 complete genomes to analyze the distributions of different class of variants in function of their ancestry. The EPIGEN Initiative is also performing several GWAS and admixture mapping studies on different complex traits, including longitudinal data. Funding: Brazilian Ministry of Health/FINEP.

1972F

The Genetic Structure and Admixture Analysis of Brazilian Populations: The Brazilian EPIGEN Initiative. *H.C. Santos[1], A. Horimoto[1], A.C. Pereira[1], E. Tarazona-Santos[2], M.L. Barreto[3], B.L. Horta[4], M.F. Lima-Costa[5], M. Gouveia[2], M. Machado[2], F. Soares[2], T.M. Silva[3], J.M. Sanches[1], N. Esteban[1], W.C.S. Magalhaes[2], M.R. Rodrigues[2], F.S.G. Kehdy[2], The Brazilian EPIGEN Consortium.* 1) Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of São Paulo; 2) General Biology Department, Federal University of Minas Gerais, Brazil; 3) Instituto de Saúde Coletiva, Federal University of Bahia, Brazil; 4) Universidade Federal de Pelotas, Brasil; 5) Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil.

Brazilians form one of the most heterogeneous populations in the world, the result of interethnic crosses between peoples from three continents: the European colonizers, African slaves and the autochthonous Native Americans. To analyze patterns of human genetic variation in Brazilian populations, we used data from 370,539 shared SNPs between EPIGEN-BRAZIL Project genotyped by Omni2.5 Illumina array (6,487 individuals from three Brazilian cohorts: Pelotas, Southern Brazil, 3,736 individuals, Minas Gerais, South East Brazil, 1,442 individuals and Bahia, north east Brazil, 1,309 individuals), 688 individuals from 6 populations from HapMap Project (CEU, YRI, ASW, TSI, MEX and LWK) and 95 individuals from 4 Native American populations from HGDP (Karitiana, Surui, Pima and Maya). Principal components and individual admixture analysis (by ADMIXTURE software) were performed to these populations to identify structure, ancestry percentages and a minimum set of SNPs needed to capture the admixture components of Brazilian populations. As expected, Brazilian samples fell between Africans, Europeans and Mexicans. Furthermore, the mean percentage of individual European, African and Native American ancestry, respectively were 0.773, 0.155 and 0.072 for Pelotas cohort, 0.798, 0.142 and 0.060 for Minas Gerais cohort and 0.435, 0.505 and 0.060 for Bahia cohort. From 370,539 analyzed SNPs, we found a minimum number of 256 SNPs able to capture 91%, 93% and 78% of respectively African, European and Native American ancestries components. Our results confirm the highly admixed character of Brazilian populations suggest this population as a good tool for admixture mapping studies. More studies about intracontinental components that contributed to the Brazilian population formation are needed and will be performed. Funding: Brazilian Ministry of Health/FINEP.

1973W

Population stratification detection and correction in rare variant collapsing methods using principal component analysis. *J.R. Wallace[1], C.B. Moore[1,2], A.T. Frase[1], M.D. Ritchie[1].* 1) Center for Systems Genomics, The Pennsylvania State University, University Park, PA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

Principal Component Analysis (PCA) has often been used in genome-wide association studies (GWAS) correct for population stratification and prevent increased type I errors. With the proliferation of Next Generation Sequencing (NGS) technology, rare variant collapsing methods have gained popularity. However, applying PCA to rare variant collapsing methods has not been well studied, and the effectiveness of adjusting for population stratification using principal components (PCs) on rare variants is largely unknown. To explore population stratification correction in rare variant data, we collapsed the low frequency (< 5% MAF) variants in the 1000 Genomes Project Phase 1 data based on Entrez gene boundaries using BioBin, a bioinformatics tool for automatically binning rare variants into biologically relevant bins (genes, pathways, regulatory regions, etc.) based on prior biological knowledge gleaned from various public data sources such as Entrez, Gene Ontology (GO), and Protein Families database (PFAM). These bins can then be evaluated with any number of popular rare variant collapsing statistical methods.

We analyzed the data for each pairwise combination of the 14 populations available in the 1000 Genomes data and found dramatic stratification and natural clustering of populations into continental groups. We then examined multiple approaches to construct PCs using different subsets of the genetic data. To compare approaches, we defined a normalized distance metric between sets of PCs as well as notions of correctness for a given stratification and predictive power without given stratification. We illustrate these concepts with respect to the 1000 Genomes data and show how similar concepts can be applied to a natural dataset in the course of analysis. We found that for populations close in ancestral history, rare variants should be excluded from PCA, and we developed a method for increasing the sensitivity of PCA, though this method may disguise other hidden stratification. Identifying and properly correcting for stratification remains an important issue; using the techniques described, we demonstrate solutions that identify and correct the ancestry stratification as well as stratification along sequencing technology.

1974T

Surveying European and West African Population Structure Using >2,300 Samples with Spatial Information. *Y. Wang, K. Noto, J.B. Byrnes, R.E. Curtis, N.M. Myres, M.J. Barber, J.M. Granka, C.A. Ball, K.G. Chahine.* AncestryDNA, San Francisco, CA.

Population structure arises as a consequence of the interplay between geography, genetic drift, and gene flow. Knowledge of human population structure is fundamental to understanding how demographic history has shaped human genetic diversity. In this work, we investigated the genetic structure of European and West African populations using high-density SNP data. First, we analyzed ~2000 European individuals genotyped at >700,000 SNPs using principal component analysis (PCA), spatial ancestry analysis (SPA), and admixture-model-based method (ADMIXTURE). Despite an overall high level of genetic similarity, we observed significant patterns of genetic variation among European populations, ranging from the regional level (e.g., Northern European vs. Southern European and Eastern European vs. Western European) to the local level (e.g., Iberian vs. Italian and English vs. Irish). We then conducted similar analyses on ~330 individuals sampled from nine West African countries (Senegal, Mali, Ivory Coast, Ghana, Benin, Togo, Nigeria, Cameroon and Congo). Our results revealed strikingly significant structure at the country level, despite geographical proximity. Comparing the three approaches, we found that SPA has better power in predicting the population of single origin individuals, given an accurate reference panel of reasonable size. We further applied SPA to identify SNPs showing large gradients in allele frequency, which can be used to tag candidate regions under natural selection. In addition, we selected two sets of ancestry informative markers (AIMs) that carry substructure information for European and West African populations, respectively. Lastly, we investigated the pattern of genetic variation in African Americans using a large cohort (>5,000) of African American individuals and the European and West African reference panels. Our results shed new light on the population and migration history of African Americans.

1975F

Using evolutionary profiles to better-inform model organism selection for human disease research. *A.D. Baxevanis, E.K. Maxwell, C.E. Schnitzler, A.D. Nguyen, R.T. Moreland.* Genome Technology Branch, Division of Intramural Research, National Human Genome Research Institute, NIH, Bethesda, MD.

While the standardization of methods for studying human diseases in traditional animal models has yielded many clinically actionable results, it has effectively narrowed the breadth of species in which we choose to look for insights. The recent expansion of whole-genome sequence data available from a diverse array of animal lineages provides an opportunity to investigate the feasibility of using non-traditional model organisms to advance human disease research. Cases in which traditional animal models have led to conclusions that are not applicable to humans are becoming more commonplace, and the concern that this may be a growing problem calls for a re-evaluation of how appropriate models are selected for different disease classes. To that end, we have used a comparative genomics approach that encompasses a wide range of animals across the metazoan tree to determine which organisms could serve as viable models for studying various classes of human diseases. We show that some emerging non-bilaterian model organisms have surprisingly high proportions of human disease gene homologs despite their great evolutionary distance from humans; these organisms may confer advantages as animal models in terms of their ease of use, short generation times and cost-effectiveness. Conversely, while it has been previously shown that the genes implicated in the causation of most human diseases are of ancient origin, our results indicate that some disease classes involve a significantly large proportion of genes that appear to have emerged relatively recently within the Metazoa. These disease classes, having a more recent evolutionary history, may be difficult to replicate phenotypically outside of our closest animal relatives. Taken together, these findings demonstrate why model organism selection should be done on a disease-by-disease basis, with evolutionary profiles in mind.

1976W

A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. *J.A. Capra[1], M.J. Hubisz[2], D. Kostka[3], K.S. Pollard[4], A. Siepel[2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Developmental Biology and Computational & Systems Biology, University of Pittsburgh, Pittsburgh, PA; 4) Institute for Human Genetics and Division of Biostatistics, University of California, San Francisco, CA.

GC-biased gene conversion (gBGC) is a recombination-associated process that favors the fixation of G/C alleles over A/T alleles. In mammals, gBGC is hypothesized to contribute to variation in GC content, rapidly evolving sequences, and the fixation of deleterious mutations, but its prevalence and general functional consequences remain poorly understood. gBGC is difficult to incorporate into models of molecular evolution and so far has primarily been studied using summary statistics from genomic comparisons. Here, we introduce a new probabilistic model that captures the joint effects of natural selection and gBGC on nucleotide substitution patterns, while allowing for correlations along the genome in these effects. We implemented our model in a computer program, called phastBias, that can accurately detect gBGC tracts ~1 kilobase or longer in simulated sequence alignments. When applied to real primate genome sequences, phastBias predicts gBGC tracts that cover roughly 0.3% of the human and chimpanzee genomes and account for 1.2% of human-chimpanzee nucleotide differences. These tracts fall in clusters, particularly in subtelomeric regions; they are enriched for recombination hotspots and fast-evolving sequences; and they display an ongoing fixation preference for G and C alleles. They are also significantly enriched for disease-associated polymorphisms, suggesting that they contribute to the fixation of deleterious alleles. The gBGC tracts provide a unique window into historical recombination processes along the human and chimpanzee lineages. They supply additional evidence of long-term conservation of megabase-scale recombination rates accompanied by rapid turnover of hotspots. Together, these findings shed new light on the evolutionary, functional, and disease implications of gBGC. The phastBias program and our predicted tracts are freely available.

1977T

Conserved combinatorial transcription factor binding identifies regulatory variants in human disease pathways. *M.D. Wilson[1,3,6], B. Ballester[2,4], A. Medina-Rivera[1], D. Schmidt[3], M. Gonzàlez-Porta[4], M. Carlucci[1], K.N. Chessman[1], A. Faure[4], A. Funnell[7], A. Goncalves[4], C. Kutter[3], M. Lukk[3], S. Menon[3], W.M. McLaren[4], K. Stefflova[1], S. Watt[3,5], M. Crossley[7], J.C. Marioni[4], D.T. Odom[3,5], P. Flicek[4,5].* 1) Genetics and Genome Biology, SickKids Research Institute, Toronto, ON, Canada; 2) INSERM U1090, TAGC, Aix-Marseille University, Marseille, France; 3) University of Cambridge, Cancer Research UK, Cambridge Institute, Robinson Way, Cambridge CB2 0RE, United Kingdom; 4) European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom; 6) Department of Molecular Genetics, University of Toronto, Canada 101 College St. East Tower, Toronto, ON, M5G 1L7, Canada; 7) School of Biotechnology and Biomolecular Sciences, University of New South Wales, NSW 2052, Australia.

In order to test whether evolutionary conservation of combinatorial protein-DNA interactions gives insight into human gene regulatory function and disease, we experimentally determined transcription factor (TF) binding locations of four master regulatory TFs (ONECUT1/HNF6, FOXA1/HNF3A, HNF4A, and CEBPA) in the livers of human, macaque, mouse, rat and dog. Approximately two thirds of the TF binding events for these four TFs resided in heterotypic clusters, which we designated as cis-regulatory modules (CRMs). Less than half of the identified human liver CRMs could be detected in orthologous regions from one or more of our study species. We found that CRMs conserved in human and at least one non-primate to be disproportionately found in regulatory regions implicated in rare liver-related genetic diseases including those affecting blood coagulation and lipid homeostasis. For example, several of the promoter regions in the blood coagulation cascade contain highly conserved CRMs that are recurrently mutated in bleeding disorders. In comparison, primate or rodent-specific CRMs indicated more lineage-specific pathways involving drug detoxification. Deeply conserved CRMs were also found in close proximity to liver and blood lipid-disease loci identified by genome-wide association studies (GWAS). Together, our data demonstrates how the rapid evolution of human combinatorial transcription factor binding can be used to identify and prioritize the functional study of pathologic regulatory mutations.

1978F

Functional characterization of Toll-like receptor signaling pathways in primates. *J.F. Brinkworth[1,2], J.N. Kohn[3,4], J. Boulais[1,5], J.C. Grenier[1], R.E. Lanford[6], Z.P. Johnson[3], L.B. Barreiro[1,2,5].* 1) Centre Hospitalier Universitaire Sainte-Justine Hospital Centre de Recherche; 2) Department of Pediatrics, University of Montreal, Quebec, Canada; 3) Yerkes National Primate Research Center, Atlanta, GA, United States; 4) Graduate Program in Neuroscience, Emory University, Atlanta, Georgia, United States; 5) Department of Biochemistry, University of Montreal, Quebec, Canada; 6) Department of Virology and Immunology, Texas Biomedical Research Institute, San Antonio, Texas, United States.

Despite close genetic relatedness, humans, apes and monkeys exhibit inter-species differences in susceptibility to certain pathogenic infections that are major causes of severe disease in humans. For example, humans are highly susceptible to Gram-negative bacterial sepsis, HIV/AIDS, and tuberculosis, while some monkey and ape species remain resistant or progress very rapidly. Such differences between humans and other primates are thought to result, at least in part, to inter-species differences in immune response to infection. However, due to the lack of comparative functional data across species, it remains unclear in what ways the immune systems of humans and other primates differ. Here we report a genome-wide comparative study of immune responses among primates, specifically focusing on Toll-Like receptor (TLR) signaling pathways. To examine if human and nonhuman primate TLR pathway activity differs, we stimulated leukocytes from humans, common chimpanzees and rhesus macaques with TLR2, 4 and 7 -detected bacterial and viral ligands (LPS from *Escherichia coli*, Lipomannan from *Mycobacterium smegmatis*, single-stranded RNA viral mimetic Gardiquimod). Blood was stimulated for 4 and 24 hours and innate immune response was assessed via global expression profiling of total blood leukocytes using RNA sequencing. Inter-species differences in gene expression were noted across many immune gene families, with humans manifesting many unique gene expression profiles in response to both bacterial and viral ligands. These observations suggest that innate immune responses of human and non-human primates have evolutionary diverged, lending support to the notion that human-specific immune responses might account for some of the known differences in susceptibility to infectious diseases between humans and our closest evolutionary relatives. This study was supported by National Science and Engineering Council of Canada (# grant number LBB) Réseau de Médecine Génétique Appliquée (RMGA), the Fonds de Recherche du Québec - Santé (FRQS) and the Canadian Institutes of Health Research (CIHR, Grant # TGF-96109) (JFB).

1979W

Accelerated evolution of primate-specific microRNAs in the human genome. *M. Lopez-Valenzuela[1], N. Petit-Marty[1], A. Navarro[1,2,3], Y. Espinosa-Parrilla[1].* 1) Institut de Biologia Evolutiva (IBE, Universitat Pompeu Fabra-CSIC), Barcelona, Catalonia, Spain; 2) Instituto Nacional de Bioinformática (INB), Barcelona, Catalonia, Spain; 3) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain.

MicroRNAs (miRNAs) are a class of small non-coding RNAs with a prominent role on gene regulatory networks. miRNA genes are continuously being added to metazoan genomes allowing the creation of lineage-specific miRNAs. The differential regulation provided by these lineage-specific miRNAs may contribute to shape different phenotypes in species with genomes of similar protein-coding content. The rise of primate lineage is marked by an outstanding emergence of new miRNA families, but how natural selection has acted on them is still not known. We studied the action of natural selection on human miRNAs applying a likelihood ratio test based in comparing substitution rates in the sequences of the miRNAs themselves with those in ancestral repeat (AR) sequences that are used as the reference for neutrality. The test can detect positive selection acting upon a small number of sites and distinguish it from relaxation of purifying selection. Primate PhyloP scores for each nucleotide in the studied miRNAs were also calculated as a complementary and alternative method. Comparison of 1,523 known human precursor miRNA sequences (miRBase release 18) with the reference genomes of chimpanzee and rhesus macaque showed nucleotide differences for 555 human miRNAs. The great majority of these miRNAs did not reach significance in the likelihood ratio test indicating that most sites in human miRNAs are either conserved or evolving neutraly. Nevertheless two miRNAs, hsa-mir-518a-1 and hsa-mir-3939, escaped markedly this tendency and presented the signature of accelerated evolution by positive selection. Interestingly, the primate-specific fraction of the analyzed miRNAs (79%) presented a higher proportion of significant nucleotides than the rest of miRNAs (p=0.00001) and showed lower average conservation index in the PhyloP analysis indicating that this group of primate-specific miRNAs could constitute a group of fast evolving miRNAs. We also found that, according to the TargetScan algorithm, these accelerated miRNAs tend to have less targets and that functions related to their target genes are similar to those of genes targeted by the set of more conserved miRNAs suggesting a redundant role for the recently emerged primate-specific miRNAs. Our results shed light on the evolution and function of human miRNAs and the way they might participate in the diversification of the primate lineage.

1980T

Addiction Drugs Cluster Functionally in the Genome with Concomitant Variation in Human Populations. *L. Jackson[1], Y. Liu[2], A. Tozeren[1].* 1) Biomedical Engineering, Science, and Health System, Drexel University, Philadelphia, PA., USA; 2) Department of Epidemiology and Biostatistics, University of Pennsylvania-Perelman School of Medicine, Philadelphia, PA, USA.

Understanding the common genetic mechanisms that underlie dopamine, opiate, and GABA addiction is of significant interest to human geneticists. We seek to use biologically relevant addiction genes as the starting pool to identify whether genes involved in addiction cluster together into loci of functional addiction control. We ask whether genes involved in addiction physically cluster in the genome, if there is a functional pattern to this clustering and whether such clusters show polymorphic variation in HapMap populations. Genes with biological relevance related to addiction (N=587) were gathered using NCBI Gene and then physically mapped onto the human genome using computational tools. Clusters of genes were identified using specific threshold criteria. These clusters were then surveyed to determine their drug addiction class membership. Polymorphisms in each of these regions were identified for each of the 11 HapMap populations. To test for deviations from neutral demographic expectations, we used neutrally evolving autosomal comparison regions and mapped within-cluster variation. Our analyses found nine addiction gene clusters with three clusters containing GABA specific addiction genes and six clusters containing addiction genes involved with all surveyed classes. Furthermore, analysis of the polymorphism underlying these clusters show patterns consistent with non-neutral GABA specific evolution in Chinese populations (sampled in either Beijing or Denver), and with non-neutral evolution at generalist addiction genes in African ancestry populations (Luhya, Masaai, African American, and Yoruba). Of particular interest are two polymorphisms (rs9467667 and rs9379817) which sit in the regulatory region of the HIST1H complex in a GABA specific addiction cluster, which show high derived variant frequencies in Chinese populations but which is nearly fixed for the ancestral variant in African and European ancestry populations. These variants may play a role in epigenetic histone regulation of alcohol sensitivity behavior.

1981F

Critical Assessment of Coalescent DNA Simulators in Modeling Recombination Hotspots. *T. Yang, H.W. Deng, T. Niu*.* Dept of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA.

Coalescent-based simulation plays a pivotal role in understanding population evolutionary models and demographic histories, as well as in developing new analytical methods for genetic association studies. A plethora of coalescent simulators are developed, but it often remains challenging to select the most appropriate program. We extensively compared performances of five most-widely used ones - Hudson's ms, msHOT, MaCS, Simcoal2, and fastsimcoal, to provide a practical guide considering three crucial factors: (i) speed, (ii) scalability and (iii) hotspots position accuracy. Hudson's ms, the most popular simulator has very robust performance under standard coalescent but lacks the ability to simulate sequences with recombination hotspots. We highly recommend ms in absence of hotspots. msHOT has compensated the deficiency of ms incorporating crossover and gene conversion hotspots at arbitrary location and intensity. Simcoal2, based on discrete-generation coalescent, could simulate more complex demographic scenarios, but is comparatively slow. MaCS and fastsimcoal, both adopting a fast sequential Markov coalescent model to approximate standard coalescent, are much more computationally efficient whilst keeping salient features of msHOT and Simcoal2, respectively. Our extensive simulations demonstrated that MaCS and fastsimcoal have significant advantages over other programs for a spectrum of demographic scenarios. We also evaluated accuracy of hotspot positions of the simulated data by LDhat 2.2 rhomap package, sequenceLDhot and haploview. We found that fastsimcoal has the best performance. In conclusion, while ms remains an excellent choice for general scenarios, MaCS and fastsimcoal are much more scalable and flexible in simulating many different demographic histories and diverse DNA sequence structures. *Corresponding Author.

1982W

The genome-wide distribution of gene conversion, cross-overs, and de novo mutations events in Western Chimpanzees. *O. Venn[1], I. Turner[2], I. Mathieson[1,2], N. de Groot[3], G. McVean[1,2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) University of Oxford, Department of Statistics, Oxford, OX1 3TG, United Kingdom; 3) Biomedical Primate Research Centre, Department of Comparative Genetics and Refinement, Rijswijk, 2288, Netherlands.

A wide spectrum of processes including cross-over, gene conversion, mutation and DNA repair impact genomes from the level of single nucleotide changes to multiple megabases. Inter-species comparisons of their genome-wide distributions may potentially be informative about the relative impact of each process and their evolution. The analysis of genetic variation measured through high-throughput population sequencing is a potentially powerful approach to study such processes. However, a major challenge in interpreting this information is the detection of erroneous, mis-assembled, or incorrectly genotyped genetic variants.

To characterize the spectrum of genome changes occurring in Western chimpanzees (*Pan troglodytes verus*) we sequenced an extended three-generation pedigree (3 founders, 3 F1, 3 F2) to ~30x across individuals. This family structure maximizes transmission information for the detection of rare events: 7/8 of the founder genomes is transmitted at least once, events arising in the F1 parent are naturally validated in the F2s, detected events can be assigned parental origin, and 99.2% of segregating sites can be phased unambiguously. At fine-scales, we describe the properties of de novo mutation events with respect to paternal age and the impact of local sequence. We estimate a sex-averaged de novo mutation rate of 1.38×10^{-8} bp per meiosis. To characterize meiotic recombination, we infer ~400 events for each cross-over and gene-conversion class, at median resolution of 835 bp, and compare these with historical LD-based cross-over hotspots. We find, human and chimpanzee share a male to female cross-over ratio of 0.6 and have an almost identical sex-averaged total genetic distance.

1983T

Moving backwards in time from present-day sequences: the paradox of sequential founder events. *M. Jeanpierre.* Department of genetics, Hôpital Cochin, Paris, France.

It is difficult to unravel the history of ancient populations, and events identified as single founder migrations may actually turn out to be a sequential series of founder events, with a small number of initial settlers gradually joined by an increasing number of related individuals of the same origin. The modeling of sequential founder events is an important issue, because each genetic variant is studied as a member of a sequence, and it is only possible to filter out these adjacent variations if we know the distribution of identical-by-descent segments. When modeling of the decay of haplotype sharing, simple geometric structures that can be described unambiguously in mathematical terms can provide an algebraic framework for analyses of the forces shaping the genealogy of a single allele. The challenge is to develop simplifications with a minimal loss of information, because the precision of the reconstruction depends on the dimensionality of the graph. The definition of blocks as physical entities, with clear borders, as for objects in the physical world, results in an apparent simplification, but is not really helpful because segments identical by descent are statistical entities with an ephemeral existence. Paradoxically, a hierarchical model, in which clusters are considered as border-less elements, gives a better representation of ancestral sequences, its precision steadily increasing with sample size. Large samples are not uncommon, with some containing more than 100,000 individuals, so this relationship between sample size and precision matters. The weakness of low-dimensional contingency tables may be one of several explanations for the poor reproducibility of association studies. This theoretical approach may be illustrated by the non random distribution of genetic variants for adult size. The clustering of variations defining ephemeral units of selection makes sense from a genetic point of view, because selection operates on visible physical characters, not on numerous, infinitesimally small genetic elements.

1984F

Effective population size - recombination, gene conversion and linkage disequilibrium. *D. Labuda, J.F. Lefebvre, C. Moreau.* Dept Ped, CHU Sainte-Justine, Univ Montreal, Montreal, PQ, Canada.

The genetic diversity of a population reflects the rate of mutation (μ) creating new alleles, and recombination events (c) that redistribute these alleles among homologous chromosomes to form a variety of haplotypes. Importantly, genetic diversity also reflects the historical effective size (Ne) of a population, since in larger populations more diversity may arise every generation and surviving variants persist for longer, thus resulting in greater population diversity. Quantitatively, the population parameters Θ=4Neμ and ρ=4Nec define the mutational and recombinational diversity, respectively. Combining different Θ estimates that reveal different aspects of population genealogies has led to statistical tests that are sensitive to selection events and/or demographic history. Here, by computer simulations, we explore different estimates of ρ and compare their sensitivity to detect recombination and gene conversion under a standard neutral population model at constant population size, as well as following population bottleneck and demographic growth. We compare ρ estimates obtained from counting past recombinations (R) with those evaluated from the extent of linkage disequilibrium (LD) along the genomic sequence. While ρLD is practically insensitive to gene conversions, thus monitoring only reciprocal crossover events, ρR monitors both recombination and gene conversion. Interestingly, following a population bottleneck these ρ estimates recover at different rates, as observed with different Θ such as Tajima's and Watterson's estimates. Both ρ estimates were also computed for the autosomes of all HapMap3 populations. An excellent correlation was observed when individual chromosomal estimates were compared, not only between our two methods, but also with LdHat as well as with recent pedigree estimates of the recombination rates. We expect that combining different recombinational diversity estimates and, especially, in combination with different mutational diversity estimates, should open up new ways to infer demographic as well as population genetic history, including events of natural selection.

1985W

Analysis of linkage disequilibrium associated with Southeast Asian Ovalocytosis (SAO). *M.K. Thompson, J.A. Wilder.* Northern Arizona University, Flagstaff, AZ.

Southeast Asian Ovalocytosis (SAO) is a type of hereditary elliptocytosis that confers protection from malaria-causing parasites, including both *Plasmodium falciparum* and *P. vivax*. This trait is caused by a 27-base pair deletion, or a 9 amino acid deletion, in the Band 3-encoding *SLC4A1* gene. While SAO provides protection from severe manifestations of malarial parasitism when heterozygous, the trait is lethal when homozygous. This pattern suggests it is maintained as a balanced polymorphism. Previous studies have found substantial short-range decay in linkage disequilibrium (LD) associated with the SAO-causal mutation, suggesting a relatively ancient age of the trait. Here we extend this analysis by targeted resequencing of loci that span 500 kilobases around the causal mutation in 52 SAO-carriers from Indonesia and Thailand. These data will allow us estimate the decay of LD over a relatively large chromosomal region, spanning a highly heterogeneous recombination environment. These results will refine our understanding of the evolutionary history of Southeast Asian Ovalocytosis.

1986T

Meiotic gene conversion in humans: rate, sex ratio and GC bias. *A.L. Williams[1], G. Genovese[1], T. Dyer[2], N. Patterson[1], J. Blangero[2], D. Reich[1,3], the T2D-GENES Consortium.* 1) Broad Institute, Cambridge, MA; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) Harvard Medical School, Boston, MA.

Gene conversions are short, 20-300 bp segments copied from one homologous chromosome to another during meiosis. While related to crossover, gene conversions are distinct in that they affect a much shorter region, are more frequent, and show bias towards G or C allele transmissions. To date, there has been no genome-wide study of *de novo* gene conversion and no direct inferences about their localization, sex differences, and the impact of GC bias on the genome.

We examined SNP array data from 190 individuals in 16 three-generation pedigrees that are informative for 42 meioses (21 paternal, 21 maternal) and 4.15×10^6 sites. Our study design uses three-generation pedigrees and requires that a gene conversion received by a second generation child is also transmitted to a third generation grandchild. This ensures that the identified gene conversions are either real or the result of at least two genotyping errors at the same site.

We identified 33 putative gene conversions and obtained a genome-wide *de novo* gene conversion rate estimate of 8.0×10^{-6} per bp per generation~a rate consistent with sperm typing and LD-based studies. We validated these gene conversions using whole genome sequence data for a subset of the genotyped samples, with genotype calls available for 19 of the 33 sites. Of these 19 sites, 18 had genotypes consistent with the SNP chip data. The single mismatching site is ambiguous as to the error source and appears to be a sequencing artifact. These results suggest a low false-positive rate and validate our study.

The gene conversions are significantly enriched in recombination hot spots, with 10/33 events in regions with recombination rate ≥10 cM/Mb (P= 1.1×10^{-8}). Male gene conversions predominantly localized to telomeres while females transmitted 1.54x more gene conversions than males. These results further validate our methodology and are consistent with the mechanism for forming gene conversions being similar to that of crossovers.

We observed extreme GC bias in allelic transmissions, with 23/31 sites heterozygous for AT and GC alleles transmitting a G or C allele (P=5.3×10^{-3}). This 3-fold enrichment of GC transmissions suggests that gene conversion substantially alters the allelic make up of the genome. We are examining a 2-fold larger dataset to further validate this and other findings.

1987F

Linkage disequilibrium and haplotype blocks determined by the analysis of 250K SNPs in three quilombo remnants communities.
E.S. Andrade[1], D.M. Salvanha[2], R.Z.N. Vêncio[2], L.M. Garrido[3], H. Krieger[3], A.L. Simões[1], C.T. Mendes-Junior[4]. 1) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil; 2) Departamento de Matemática e Computação, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil; 3) Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, 05508-000, São Paulo-SP, Brazil; 4) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil.

The non-random association between alleles of different loci characterizes what is called linkage disequilibrium (LD) between them. The LD extent in human populations can be influenced by many factors, such as recombination rate, demographic features and evolutionary events. The aim of this study was to describe the LD patterns of four Brazilian populations and correlate these patterns with their respective demographic histories. Samples from three quilombo remnants populations of the Piauí State, Gaucinha (GAU, n = 14), Mimbó (MIB, n = 15) and Sítio Velho (STV, n = 15) and the urban population of Teresina, Piauí (TES, n = 15), and seven population samples from the HapMap Project (CEU, CHB, JPT, ASW, LWK, MKK, YRI, all with n = 15) were analyzed. More than 250 thousand SNPs (Single Nucleotide Polymorphisms) were genotyped using the GeneChip ® Human Mapping 250K Nsp Array I - Affymetrix® in the samples of the four Brazilian populations. Raw data of the HapMap population samples for this array were obtained from the HapMap homepage. Genotypes for all samples were determined by CRLMM algorithm. LD analyzes and determination of haplotype blocks were performed using the Haploview software. Considering the number of haplotype blocks detected in each population, a consistent pattern was observed for all autosomes. The European population (CEU) and the two Asian populations (CHB and JPT) of the HapMap showed the highest numbers of blocks, while the lowest numbers were observed in the GAU and MIB quilombos and in the TES population. The African populations, LWK, MKK and YRI, and the African-American ASW exhibited intermediate values and the African-Brazilian population STV, presented a number of blocks just smaller than that observed for CEU, CHB and JPT. The great African contribution in the GAU and MIB quilombos may explain their lower LD. On the other hand, the lower LD in TES is probably due to its foundation that involved a larger number of individuals and was followed by a fast growth. A possible explanation for the higher LD observed in STV consists in its particular demographic history: admixture at the time of its foundation, slow growth and low differentiation. Thus, it was shown how the demographic events of each population influence their respective LD patterns. FINANCIAL SUPPORT: FAEPA-HCFMRP-USP/Brazil (Grant 1394/2011).

1988W

A recombination map of Latino populations inferred using local ancestry.
S. Shringarpure[1], S. Gravel[2], C. Gignoux[3], A. Moreno[1], C. Eng[3], S. Huntsman[3], D. Torgerson[3], E. Burchard[3], C. Bustamante[1]. 1) Genetics, Stanford University, Stanford, CA, 94305; 2) McGill University, Human Genetics, Montreal, H3A 1B1, Canada; 3) University of California, San Francisco, Bioengineering, San Francisco, CA, 94158.

Recombination maps have traditionally been inferred in populations using information about recombination events in pedigrees. Recent work (Hinch et al., Wegmann et al.) has demonstrated how knowledge of local ancestry can be used to infer relative recombination rates in admixed populations. We apply similar techniques to admixed Latino American populations to infer a high-resolution map of relative recombination rates. We present a method for inferring relative recombination rates in Latino populations by using signals from the admixture events in their evolutionary history. We used 2001 Latino individuals of Mexican and Puerto Rican ancestry from the GALA study, including 551 trios, for our analysis. Local ancestry inference was performed using software RFMix to infer tracts of African, European and Native American ancestry in the study individuals. The inferred ancestry switch points were then used to perform statistical inference and infer recombination rate variation along the genome. We used the statistical framework of Dirichlet Process Mixture Models to correct for the biases of local ancestry inference and estimate underlying rates of ancestry switches. We validated our method on data simulated from an admixture model of the evolutionary history of Latino populations, with admixture parameters from literature. Simulations show that our method can accurately resolve ancestry switch points to within a distance of 400 kb with 80% accuracy. Our analysis of the Latino dataset shows evidence of differentiation in recombination rates between populations. We also compare the recombination rates in the different admixed and nonadmixed populations and interpret the results in light of the PRDM9 variants observed in each. References [1] Wegmann, D., Kessner, D. E., Veeramah, K. R., Mathias, R. A., Nicolae, D. L., Yanek, L. R., Sun, Y. V., et al. (2011). Recombination rates in admixed individuals identified by ancestry-based inference. Nature genetics, 43(9), 847-53. doi:10.1038/ng.894 [2] Hinch, A. G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C. D., Chen, G. K., et al. (2011). The landscape of recombination in African Americans. Nature, 476(7359), 170-5. doi:10.1038/nature10336.

1989T

De novo genes in evolution: Apcdd1, a novel dual Wnt and Bmp inhibitor.
V. Luria[1,3], S.G. Oh[1], A.H. O'Donnell[2], A.M. Christiano[3]. 1) Systems Biology, Harvard Medical School, Boston, MA; 2) Children's Hospital Boston, Harvard Medical School, Boston, MA; 3) Genetics & Development, Columbia University Medical Center, New York, NY.

The dynamic instability of biological circuits enables them to decide between alternatives. In genetic circuits transducing signals to cells, circuit components can surprisingly be new genes, raising the question of how these arise evolutionarily and integrate into existing, ancient circuits. A new signal transduction component, Apcdd1, encodes a 514 amino-acid membrane-anchored extracellular protein that, intriguingly, has no significant similarity to any protein. Expressed in neurons, germline and ectodermal appendages, it can be induced by Wnts and, when mutated, causes a form of human hair loss. We showed APCDD1 inhibits Wnt signaling in chick, frog and humans (Shimomura*, Luria* et al., Nature 2010). We now show APCDD1 inhibits Bmp signaling using biochemistry, single-cell imaging and in vivo experiments in chick, frog and zebrafish. The position of APCDD1 at the intersection of Wnt/Bmp pathways may enable it to coordinate pathway activation dynamics. To determine whether Apcdd1 is a de novo gene, we asked when Apcdd1 appeared in evolution, and interrogated the likelihood of gene emergence by coopting random sequences. We found Apcdd1 appeared in Eumetazoa, has no homologues, and that most APCDD1 amino acid sequence is relatively disordered with no recognizable protein domain, suggesting Apcdd1 is a de novo gene. To determine the frequency and content of de novo genes, combining bioinformatics and mathematical modeling, we investigated how de novo genes may arise as a function of gene, genome and population parameters, and the protein structure content of de novo proteins. Surprisingly, we find de novo genes may arise more frequently than previously predicted, may be more disordered than ancient proteins, and that even random proteins have substantial secondary structure. Since de novo genes like Apcdd1 can acquire genetic partners and biochemical functions, the next challenge will be understanding what fraction are maintained in genomes that continuously generate and destroy new genes.

1990F

Detecting evolutionary strata on the human X chromosome in the absence of gametologous Y-linked sequences. *M. Wilson Sayres[1], R. Shanker Pandey[2], R. Azad[2,3].* 1) Integrative Biology, University of California, Berkeley, Berkeley, CA; 2) Biological Sciences, University of North Texas, Denton, TX; 3) Mathematics, University of North Texas, Denton, TX.

Mammalian sex chromosomes arose from a pair of homologous autosomes that differentiated into the X and Y chromosomes following a series of recombination suppressions that likely happened due to serial inversions on the Y. The stepwise recombination suppressions from the distal long arm to the distal short arm of the chromosomes are reflected as regions with distinct X-Y divergence, referred to as evolutionary strata on the X. Identifying evolutionary strata is central to understanding the history and dynamics of sex chromosome evolution. All current methods of stratum detection depend on X-Y comparisons but are severely limited by the paucity of X-Y gametologs. We have developed an integrative method that combines a top down, recursive segmentation algorithm with a bottom up, agglomerative clustering algorithm to decipher compositionally distinct regions on the X, which reflect regions of unique X-Y divergence. In application to human X chromosome, our method correctly classified a concatenated set of 35 previously assayed X-linked gene sequences by evolutionary strata. We then extended our analysis, applying this method to the entire sequence of the human X chromosome, in an effort to define specific stratum boundaries. The boundaries of more recently formed strata, 4 and 5, have been defined by previous studies, and are recapitulated with our method. The older strata, 1-3, have remained poorly resolved due to paucity of X-Y gametologs. By analyzing the entire X sequence, our method identified seven evolutionary strata in these ancient regions, where only three could previously be assayed. Through this study, we present a method for fine scale detection of evolutionary strata across the entire human X chromosome, independent of information about gametologous Y-linked sequences. Our study also provides information about the contribution of repetitive elements in shaping the compositional landscape of the human X chromosome following the suppression of X-Y recombination.

1991W

***NANOGP8*: Evolution of a Human-Specific Cancer-Promoting Retrogene.** *D.J. Fairbanks[1], T.H. Ogden[1], A.D. Fairbanks[1], G.J. Parker[1], P.J. Maughan[2].* 1) Department of Biology, Utah Valley Univ, Orem, UT; 2) Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT.

NANOGP8 is a human (*Homo sapiens*) retrogene, expressed predominantly in cancer cells where its protein product is tumorigenic. It arose through retrotransposition from its parent gene, *NANOG*, which is expressed predominantly in embryonic stem cells. Based on identification of fixed and polymorphic variants in a genetically diverse set of human *NANOG* and *NANOGP8* sequences, we estimated the evolutionary origin of *NANOGP8* at approximately 0.9 to 2.5 million years ago, more recent than previously estimated. We also discovered that *NANOGP8* arose from a derived variant allele of *NANOG* containing a 22-nucleotide pair deletion in the 3' UTR, which has remained polymorphic in modern humans. Evidence from our experiments indicates that *NANOGP8* is fixed in modern humans even though its parent allele is polymorphic. The presence of *NANOGP8*-specific sequences in Neanderthal reads provided definitive evidence that *NANOGP8* is also present in the Neanderthal genome. Some variants between the reference sequences of *NANOG* and *NANOGP8* utilized in cancer research to distinguish RT-PCR products are polymorphic within *NANOG* or *NANOGP8* and thus are not universally reliable as distinguishing features. *NANOGP8* was inserted in reverse orientation into the LTR region of an SVA retroelement that arose in a human-chimpanzee-gorilla common ancestor after divergence of the orangutan ancestral lineage. Transcription factor binding sites within and beyond this LTR may promote expression of *NANOGP8* in cancer cells, although current evidence is inferential. The fact that *NANOGP8* is a human-specific retro-oncogene may partially explain the higher genetic predisposition for cancer in humans compared with other primates.

1992T

Characterizing bias in population genetic inferences from uncertain genotype data. *E. Han[1], J. Sinsheimer[1], J. Novembre[1,2].* 1) Biostatistics, UCLA, Los Angeles, CA; 2) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA.

The site frequency spectrum (SFS) is of primary interest in population genetic studies, because the SFS compresses variation data into a simple summary from which many population genetic inferences can proceed. However, inferring the SFS from sequencing data is challenging because genotype calls from sequencing data are often inaccurate due to high error rates and if not accounted for, this genotype uncertainty can lead to serious bias in downstream analysis based on the inferred SFS. In our work, we compare two types of approaches to estimate the SFS from sequencing data by detailed simulations: one method that infers individual genotypes from aligned sequencing reads and then estimate the SFS based on the inferred genotypes (two-stage approach) and another method that directly estimates the SFS from aligned sequencing reads by maximizing a likelihood of the SFS (direct estimation approach). We find that the SFS by the direct estimation approach is unbiased even at low coverage, whereas those by the two-stage approach become biased as coverage decreases and most deviations come from the sites with rare variants. Interestingly, the bias by the two-stage approach is in opposite directions depending on the pipeline to infer genotypes: estimating genotypes by pooling individuals in a sample (multi-sample calling) results in underestimation of the number of rare variants, whereas estimating genotypes in each individual and merging them later (single-sample calling) leads to overestimation of rare variants. We characterize the impact of these biases on downstream analyses, such as demographic parameter estimation and rank-based genome-wide selection scans. We observe that at low coverage, the estimated growth rate based on the directly inferred SFS is almost unbiased, whereas using the SFS inferred by a two-stage approach is biased upwards using the single-sample calling and biased downwards using the multi-sample calling. In contrast, bias is less problematic for rank-based genome-wide selection scans because the rank ordering of genomic regions is more robust to genotype errors and biases in the inferred SFS. Our works highlight that depending on the pipeline to infer the SFS from sequencing data one might reach different conclusions in population genetic inferences with the same data set and care is vital for these analyses.

1993F

The evolutionary impact of GC-biased gene conversion on human populations. *J. Lachance, S. Tishkoff.* Dept Genetics, Univ Pennsylvania, Philadelphia, PA.

Gene conversion refers to the non-reciprocal transfer of genetic information between two recombining sequences, and there is evidence that this process is biased towards G and C alleles. Biased gene conversion influences both standing genetic variation and the probability of fixation of new mutants. Using high-coverage whole genome sequences of African hunter-gatherers (including 5 West African Pygmy, 5 Hadza, and 5 Sandawe genomes), other global human populations, and primate outgroups we quantify the effects of GC-biased gene conversion on population genetic data. Gene conversion results in modified allele frequency distributions, particularly when the ancestral allele is A or T and the derived allele is G or C. These shifted allele frequency distributions yield modified values of Tajima's D and Fay and Wu's H, potentially leading to false inferences of natural selection and population size changes. These effects vary across the recombination landscape of the human genome, and are more pronounced in high recombination regions. In addition, biased gene conversion can at least partly explain the two-fold difference in mutation rates estimated from human-chimp comparisons and pedigree data. Because molecular clock estimates depend on accurate estimates of mutation rates, incorporating gene conversion leads to more accurate inference of demographic history. Taken together, our findings reveal that molecular genetic phenomena like GC-biased gene conversion have important population genetic implications.

1994W

Characterization and evolution of LAVA elements in gibbons. *M.K. Konkel[1], J.A. Walker[1], B. Ullmer[2], T.J. Meyer[3], A. Damert[4], R. Hubley[5], A.F.A. Smit[5], L. Carbone[3], M.A. Batzer[1] for the Gibbon Genome Sequencing and Analysis Consortium.* 1) Dept. of Biological Sciences, Louisiana State University, Baton Rouge, LA; 2) School of Electrical Engineering and Computer Science, Center for Computation and Technology, Louisiana State University, Baton Rouge, LA; 3) Dept. of Behavioral Neuroscience, Oregon Health & Science University, Beaverton, OR; 4) Molecular Biology Center, Babes-Bolyai University, Cluj-Napoca, Cluj, Romania; 5) Institute for Systems Biology, Seattle, WA.

We have recently discovered a novel retrotransposon called 'LAVA' in the gibbon lineage. LAVA represents a composite element closely related to the hominoid-specific SVA element. Both share the VA part (VNTR (variable number of tandem repeat region) and *Alu*-like sequence). However, instead of the SVA-specific SINE-R region, LAVA elements contain unique sequence sections as well as ancient *Alu* and L1 sequence. In our investigations of this element, we first analyzed LAVA insertions to confirm that they harbor hallmarks of non-LTR (long terminal repeat) retrotransposons. We conclude that LAVAs most likely rely on the enzymatic machinery of L1 (long interspersed element 1) for their propagation. Next, we investigated the tempo and mode of LAVA expansion in the gibbon lineage leading to *Nomascus leucogenys* (NLE). Computational reconstruction followed by manual curation of the LAVA subfamily structure revealed 23 distinct full-length subfamilies in the NLE draft genome assembly [Nleu 1.1]. We identified >2500 LAVA elements since the origin of gibbons in the NLE genome. Based on the LAVA subfamily structure, activity of some subfamilies peaked prior to the radiation of gibbons, while other subfamilies show evidence for continued propagation until (at least) very recently. To better understand the expansion dynamics of LAVA elements and to investigate the still unresolved gibbon phylogeny, we selected 200 loci for our phylogenetic PCR analyses. We selected LAVA elements based on their divergence from their respective consensus sequence. PCR analyses were performed on a primate DNA panel containing 13 gibbon species from all four genera, with four great apes and green monkey as outgroups. Our wet-bench results confirmed our computational findings that several of the oldest appearing subfamilies ceased activity prior to the radiation of gibbons; while some subfamilies provide evidence for very recent retrotransposition. Even though mobile elements represent a marker system with advantages such as identity by descent and near-absence of homoplasy, we were not able to determine the radiation order of the four gibbon genera due to extended incomplete lineage sorting. This indicates rapid speciation and/or historically recurring hybridization events. Furthermore, our results show an expansion of LAVA elements in gibbons, provide evidence for ongoing LAVA propagation, and highlights that primate lineages evolve uniquely.

1995T

Genetic architecture of variants affecting splicing in human populations. *Y. Lee[1], E. Gamazon[1], H. Im[2], W. Hernandez[1], N. Cox[1].* 1) Department of Medicine, University of Chicago, Chicago, IL; 2) Department of Health Studies, University of Chicago, Chicago, IL.

Alternative splicing (AS) is a crucial mechanism producing unstable proteins in various human pathologies. Transcript isoforms due to differential splicing can occur with different frequencies in different human populations. We hypothesize that SNPs affecting splicing through Splicing Regulatory Elements (SRE) may show evidence for a signature of natural selection including potential evidence for population differentiation in allele frequency. We previously identified splicing-affecting SNPs within splicing regulatory element (SRE) sites which result in skipping of the adjacent exon. In this study, we have investigated SNPs in SREs considered to be silent genic variants (intronic and synonymous SNPs) to gain insight into their contribution to population differentiation by using population-pairwise *Fst* estimates and derived allele frequencies among four populations, resulting in six pairs of comparisons (AFR and AMR, AFR and ASN, AFR and EUR, AMR and ASN, AMER and EUR, and ASN and EUR). We found intronic SNPs in SREs have significantly higher *Fst* values than intronic SNPs not located in SREs; this was true across the allele frequency spectrum. Results are similar for synonymous SNPs in SREs when compared with synonymous SNPs not located in SREs. We also found that for both intronic and synonymous SNPs in SREs, the derived allele frequencies increase as the population differentiation (*Fst*) increases. These findings were observed across all study populations. Our study suggests that SNPs in SREs maybe be commonly subject to natural selection and that this class of functional variation may contribute to population differentiation in disease risk and phenotypes.

1996F

From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *B. Georgi, B.F. Voight, M. Bucan.* University of Pennsylvania, Philadelphia, PA.

Understanding the core set of genes that are necessary for basic developmental functions is one of the central goals in biology. Studies in model organisms identified a significant fraction of essential genes through the analysis of null-mutations that lead to lethality. Recent large-scale next-generation sequencing efforts have provided unprecedented data on genetic variation in human. However, evolutionary and genomic characteristics of human essential genes have never been directly studied on a genome-wide scale.

In this study we used detailed phenotypic resources available for the mouse and deep genomics sequencing data from human populations to characterize patterns of genetic variation and mutational burden in a set of 2,472 human orthologs of known mouse essential genes. Using the 1000 Genomes Phase 1 dataset we compared the genomic characteristics of these genes with all protein-coding genes, as well as a set of 3,811 genes with non-lethal mouse phenotypes. Consistent with the action of strong, purifying selection, the essential genes exhibit significantly reduced levels of sequence variation (Wilcoxon $P=1.66\times10^{-180}$), skew in allele frequency towards more rare (Wilcoxon $P=3.12\times10^{-35}$), and increased conservation across the primate and rodent lineages (Wilcoxon $P=1.28\times10^{-75}$). In individual genomes we observed ~12 rare mutations within essential genes predicted to be damaging. Consistent with the hypothesis that mutations in essential genes are risk factors for neurodevelopmental disease, we show that de novo variants in patients with Autism Spectrum Disorder (ASD) are significantly more likely to occur in this collection of genes ($P=2.7\times10^{-4}$ by gene permutation). Within a core set of 179 essential genes with de novo events exclusively in ASD cases, we show an enrichment of protein connectivity (P=0.0019 based on 1000 DAPPLE permutations).

In currently ongoing work, we are expanding on the role of essential genes in Schizophrenia and bipolar disorder based on previously reported de novo variants, as well as inherited variants identified in family-based association studies.While incomplete, our set of human orthologs shows characteristics fully consistent with essential function in human and thus provides a resource to inform and facilitate interpretation of sequence data in studies of human disease.

1997W

Exploring the Relationship Between Immune System Related Genetic Variants and Complex Traits and Disease Through a Phenome-Wide Association Study (PheWAS). *A. Verma[1], H. Kuivaniemi[2], G. Tromp[2], D.J. Carey[2], G.S. Gerhard[1,2], J.E. Crowe Jr.[3], M.D. Ritchie[1], S.A. Pendergrass[1].* 1) Center for Systems Genomics, The Pennsylvania State University, State College, PA; 2) Geisinger Health System, Danville, PA; 3) Vanderbilt University, Nashville, TN.

Exploring the relationship between immune-system related genetic loci and a wide array of traits and outcomes provides a way to elucidate interrelationships among the immune system and diagnoses, as well as identify pleiotropic loci. To explore these connections further and identify novel associations and pleiotropy, we selected 132,467 single nucleotide variants (SNPs) with allele frequency > 0.01 that were previously associated with autoimmunity and the immune system to perform a Phenome-Wide Association Study (PheWAS). We calculated associations between the SNPs and 480 clinical diagnoses. To define case-control status we used ICD9 diagnosis codes from 3,035 subjects using de-identified electronic medical records (EMRs) from the Geisinger Clinic MyCode biorepository. We required ≥10 case subjects for ICD9 code inclusion, and used logistic regression for all associations, adjusting models for age and sex. With an exploratory P-value cutoff < 0.001 we found a total of 7,910 SNP-diagnosis associations, and 420 SNPs associated with more than one diagnosis. The most significant novel association was the *GLRB* SNP rs17035787 and the diagnosis of 'hemorrhagic disorder due to circulating anticoagulants' (P = 4.78×10-9; 29 cases, 3,007 controls). Additional novel associations included *CAMTA1* SNP rs79204895 and 'chronic sinusitis' (P =3.5×10-8; 19 cases, 3,017 controls) and the SNP rs11842088 and 'syncope and collapse' (P = 4.29×10-8; 106 cases, 2,930 controls). We found replication of previously reported GWAS results, including the second most significant result of this study, the *F5* SNP rs6025 associated with the diagnosis of 'venous thrombosis' (P = 3.5×10-8; 55 cases, 2,981 controls). Potential pleiotropy was identified, such as the SNP rs79268593 associated with three diagnoses: 'chest swelling, mass, or lump'; 'malignant neoplasm of the bladder'; and 'hypertensive chronic kidney disease'. Further, we found a series of SNPs previously reported to be associated with Crohn's disease and inflammatory bowel disease, that were associated with rheumatoid arthritis in our study. In total, our results showed intriguing relationships between loci and diagnoses, including autoimmune and direct immune system associations. Further work will include seeking replication of the results of this study in an independent EMR-based dataset through the Vanderbilt University Medical Center BioVU repository.

1998T

Bayesian co-estimation of selfing rate and locus-specific mutation rates for a partially selfing population. *B.D. Redelings, M.K. Uyenoyama.* Biology Dept., Duke University, Durham, NC.

We present a novel Bayesian method for inferring the inbreeding rate and locus-specific mutation rates for populations reproducing by partial selfing. We extend the Ewens Sampling Formula (ESF) under the infinite-alleles model to accommodate inbreeding. Our method uses the ESF to determine the likelihood for diploid samples comprising data from patches throughout the genome, accounting explicitly for multi-locus identity disequilibrium generated by inbreeding. Each locus draws a mutation rate from a common mutation rate distribution, which is itself estimated from the data. We construct an MCMC algorithm and apply it to the analysis of androdioecious, gynodioecious, and pure hermaphroditic populations, jointly inferring the frequencies of each reproductive class and rates of inbreeding and locus-specific mutation.

1999F

Geographic structure of an allele associated with blond hair color in western Island Melanesia. *H. Norton, E. Correa.* Department of Anthropology, University of Cincinnati, Cincinnati, OH.

Here we report on the frequency and distribution of an allele in the *TYRP1* gene associated with blond hair color in several populations from across western Island Melanesia. This mutation and its association with hair color were originally discovered by Kenney et al. (2013) in populations from the Solomon Islands, where the blond-hair associated allele occurs at an average frequency of 0.26. However, knowing that the blondism phenotype in Melanesia extends further to the east, we typed this mutation in 515 individuals from several different islands throughout the Bismarck Archipelago. The mean frequency of the blondism allele in the region is much lower (0.125) than observed in Solomon Island populations, although it varies dramatically from island to island, ranging from a frequency of 0.015 on the island of New Britain to a frequency of 0.260 on the island of New Ireland. Linear regression using a recessive model was used to test for an association between genotype at this SNP and quantitatively assessed skin and hair pigmentation (measured as the M index). Associations were tested for in the total sample as well as on three islands separately (New Britain, n = 235; New Ireland, n =152, New Hanover, n = 88). Using age and island as covariates, we demonstrate that genotype at the R932 SNP can explain 8.8% of the variance in hair color ($p < 0.0001$) in the full dataset. A significant association between the R932C genotype and hair M index was only observed on the island of New Ireland ($p < 0.01$), but this may be due in part to reduced power to detect these associations in the smaller island subsamples. We do not observe a significant association between skin pigmentation and R932C genotype in the full dataset or the three islands individually, confirming previous observations (Kenny et al., 2013) that this allele affects hair pigmentation independent of skin pigmentation. The geographic heterogeneity of the blondism allele in western Island Melanesia is consistent with extensive population substructure in the region.

2000W

Test of synergistic interactions among deleterious alleles in humans. *M. Sohail[1], A. Kondrashov[2], P.I.W de Bakker[3], S. Sunyaev[4], GoNL Consortium.* 1) Systems Biology PhD Program, Harvard University, Boston, MA; 2) University of Michigan, Ann Arbor, MI; 3) University Medical Center Utrecht, Utrecht, Netherlands; 4) Brigham and Women's Hospital, Boston, MA.

Understanding the nature of epistatic interactions is crucial to obtaining a complete picture of genetic variation underlying phenotypic diversity. The role of epistatic interactions in the genetic architecture of human complex phenotypes has been widely debated. In evolutionary genetics, epistatic selection may be an important force shaping DNA sequence variation with implications for the adaptive value of sex and reproduction. We developed a new statistical method to detect epistasis in sequencing data, based on testing for variance depletion in the distribution of deleterious mutations. Our statistical test is suitable for performing phenotype-focused studies in medical genetics and addressing key questions in evolutionary theory. We applied our test to detect if synergistic epistasis is the predominant mode of genetic interaction with respect to fitness in human populations. Synergistic interactions mean that multiple mutations have a larger cost on fitness than expected from their individual effects. The mutational deterministic hypothesis for the maintenance of sex postulates that sex is an adaptation to purge the genome of deleterious mutations, and requires synergistic epistasis. We had access to a unique sequencing dataset for this study. The Genome of the Netherlands (GoNL) is an effort to characterize genomic variation in the Dutch population through whole-genome sequencing of 250 families (231 trios, 19 twin quartets) at 12x using Illumina HiSeq. In the GoNL data, we observed that spouses are correlated in the number of private variants for all functional classes of variants, but that the variation of this measure can be primarily attributed to the province of origin, suggesting subtle population substructure within the Netherlands. Indeed, we detected a north-south gradient of increasing heterozygosity due to private variants. After correcting for this structure, our results indicate that the variance in the number of private variants in a specific category decreases with the potential functional importance of the allelic class. However, in the current dataset the variance depletion is not significant compared to the theoretical expectation. This may be due to additional sources of variance inflation, or absence of predominant synergistic epistasis in fitness. We complement our analysis with simulations and analytical approaches to characterize the strength of the variance depletion due to synergistic epistasis as a function of selection.

2001T

Genetic Diversity is a Predictor of Survival in Humans. *N.A. Bihlmeyer[1,2], A. Scaria[3], M. Nalls[4], M. Garcia[5], K.L. Lunetta[6,7], J.M. Murabito[7], D.R. Weir[8], J.A. Smith[9], G. Davies[10], M. Allerhand[10], L. Yu[11], D.A. Bennett[11], S.S. Mirza[12], N. Direck[12], A. Teumer[13], G. Homuth[13], A.V. Smith[14,15], V. Gudnason[14,15], T. Lumley[3], D.E. Arking[1] on behalf of the CHARGE Aging & Longevity Working Group.* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Statistics, University of Auckland, Auckland, NZ; 4) Laboratory of Neurogenetics, National Institute of Aging, National Institutes of Health, Bethesda, MD, USA; 5) Laboratory of Epidemiology, Demography and Biometry, National Institute of Aging, National Institutes of Health, Bethesda, MD, USA; 6) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 7) The National Heart Lung and Blood Institute's Framingham Heart Study, Framingham, MA; 8) Survey Research Center, Institute for Social Research, University of Michigan 426 Thompson Street, 4123 MISQ, Ann Arbor, MI; 9) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 10) Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK; 11) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago IL, USA; 12) Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands; 13) Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; 14) Icelandic Heart Association, Kopavogur, Iceland; 15) University of Iceland, Reykjavik, Iceland.

With the advent of genome-wide association studies (GWAS), and more recently whole-exome and whole-genome sequencing, remarkable progress has been made in elucidating the genetics of complex traits, with numerous genetic variants each explaining a small fraction of the variance. The presence of numerous segregating small effect alleles that influence traits that directly impact health raises the question as to whether global measures of genomic variation are themselves associated with human health and disease. Indeed, increased fitness has been associated with the increase of genetic diversity across many organisms, and is often referred to as Positive Heterozygosity Fitness Correlations (HFCs). Two general mechanisms that act at a genome level to influence fitness have been proposed. The first is compensation for recessive deleterious mutations, whereas the second is a specific advantage of the heterozygous state over either homozygous state (overdominance/heterozygous advantage). In this study, heterozygosity was measured as the sum of all heterozygous loci divided by the expected state given the allele frequency under Hardy Weinberg Equilibrium. To test for the effect of genome-wide heterozygosity on survival, we performed a meta-analysis of 16 cohorts (12 European ancestry, 4 African American ancestry) followed prospectively, with a combined sample size of 45,704 individuals, including a total of 14,785 deaths. Using Stouffer's method to combine Z-scores, weighted by the number of deaths in each cohort, we find a significant association between increased heterozygosity and survival (P=0.03). There was no evidence for heterogeneity across studies, and a direct comparison of European Ancestry to African European ancestry cohorts showed no significant difference in effect estimates (P=0.83). Similarly, no evidence for heterogeneity between sexes (P=0.67) or between causes of death was seen (Cancer, CVD and Other; P=0.85). In summary, this study provides evidence that the protective effect of increased heterozygosity seen in lower organisms functions in humans as well.

2002F

A study on genetic diversity of ADME genes in ethnic groups in Northwestern China. *J. Li, S. Xu.* CAS-MPG Partner Institute for Computational Biology, Shanghai, China.

Genetic polymorphisms in many genes related to drug absorption, distribution, metabolism and excretion (ADME) contribute to the high heterogeneity of drug responses in humans. The minority ethnic groups of northwestern China showed significant anthropological and genetic admixture characteristics with ancestry contribution from both European and East Asian populations. Therefore, the genetic diversity of ADME genes might also show high differentiation between those ethnic groups that could cause the potential clinical risk of drug safety. Here, we investigated the genetic diversity pattern of 285 ADME genes from five Northwestern Chinese minority populations, i.e. Tajik, Uyghur, Kazakh, Kirgiz, and Hui. By integrating the genotyping information of CEU and CHB from Hapmap 3 dataset, we found that the complexities of genetic diversity, such as haplotype diversity, of admixed populations are generally higher than their representative parental populations, i.e. CEU and CHB. Especially, this pattern was apparent in some functional important genes which might be shaped by the natural selection occurred in their ancestral populations. Additionally, we analyze the genetic differentiation associated with clinically relevant, functional polymorphic SNPs, which is important for evaluating potential among-population heterogeneity in drug treatment effects.

2003W

Diversity of lactase persistence alleles in Ethiopia; signature of a soft selective sweep. *D.M. Swallow[1], B.L. Jones[1], T.O. Raga[2], A. Liebert[1], P. Zmarz[1], E. Bekele[2], E.T. Danielson[3], A.K. Olsen[3], N. Bradman[1,4], J.T. Troelsen[3].* 1) Genetics Evolution & Environment, University College London, London WC1E6BT, United Kingdom; 2) Department of Biology, University of Addis Ababa, PO Box 32597, Ethiopia; 3) Department of Science, Systems and Models, Roskilde University, DK 4000 Roskilde, Denmark; 4) Henry Stewart Group, 28/30 Little Russell Street, London WC1A 2HN, United Kingdom.

The persistent expression of lactase into adulthood in humans is a recent genetic adaptation that allows the consumption of milk from other mammals after weaning. In Europe, a single allele (*-13910*T*, rs4988235), located in an upstream enhancer of the lactase gene, *LCT*, increases lactase expression in promoter constructs *in vitro* and is responsible for lactase persistence. The widespread occurrence of this allele on a very extended haplotype suggests strong directional selection in the last 5000 years. In Africa and the Middle East, the situation is more complicated and at least three other alleles (*-13907*G*, rs41525747; *13915*G*, rs41380347; *-14010*C*, rs145946881) in the same *LCT* enhancer region can cause continued lactase expression. We have now examined the *LCT* enhancer sequence in a lactose tolerance tested Ethiopian cohort of more than 350 individuals. We show that a further SNP, *-14009 T>G* is significantly associated with lactose digester status and *in vitro* functional tests confirm that the 14009*G allele also increases expression of an *LCT* promoter construct. Other rarer SNPs are either more frequent in non-digesters or not frequent enough to assess for association. The various derived alleles in the *LCT* enhancer region are spread through several ethnic groups in Ethiopia, and we report a greater genetic diversity in lactose digesters than non-digesters. By examining flanking markers to control for the effects of mutation and demography, we further describe, from empirical evidence, the signature of a soft selective sweep. These observations are of relevance to genome wide disease association studies since such parallel selection of functionally equivalent alleles could potentially hide causal gene regions.

2004T

Characterization of private and highly diverged variants among human populations from the analysis of exome sequencing. *L.R. Botigué[1], D.M. Bobo[1], D. Twigg[2], C.D. Bustamante[3], J.M. Kidd[2,4], B.M. Henn[1].* 1) Ecology and Evolution, Stony Brook, Stony Brook, NY; 2) Human Genetics, University of Michigan, Ann Arbor, MI; 3) Genetics, Stanford University, Stanford, CA; 4) Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

Previous analysis of human genomic variation revealed an excess of private variants present at very low frequencies in Eurasian populations, and that variants in Europeans are enriched for possibly damaging functional effects. These patterns have been related to the Out-of-Africa bottleneck and the exponential growth of European populations over the last 200 generations. However, a recent analogous study found no significant differences between African-American and European populations regarding the burden of potentially deleterious mutations. Here, we analyze exome data from ten populations with different ancient and recent demographic histories (sequences include publically available 1000 Genomes Project data and samples from the Human Genome Diversity Panel). By contrasting populations with and without explosive growth or severe bottlenecks we are able to assess the effect of demography and zygosity on the distribution of potentially damaging alleles. Additionally, we focus on variants showing high frequency differences between populations, and ask whether specific functional categories are driving differentiation among human populations, with the ultimate goal of understanding the relative importance of demography and natural selection as drivers of population divergence.

2005F

Identification of known genes in a novel gene discovery project: Experience of the FORGE Canada Consortium. *S.L. Sawyer[1], C.L. Beaulieu[1], D.E. Bulman[1], F.P. Bernier[2], K.M. Boycott[1] FORGE Canada Consortium.* 1) Medical Genetics, Children's hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 2) Medical Genetics, Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada.

The primary objective of the FORGE (Finding Of Rare disease GEnes) Canada project was the identification of novel disease-causing genes. Over 200 rare diseases with pediatric onset were studied over a 2-year period using whole exome sequencing (WES) and one of three approaches: (1) cohorts of patients with the same rare disease; (2) consanguineous or dominant families with mapping data; and, (3) affected sib-pairs born to unrelated parents. In approximately 1/3 of the 200 FORGE projects analyzed, known genes previously associated with disease were identified as causative for the phenotype despite careful selection of projects prior to WES to enrich for those more likely to identify a novel disease gene. Review of the characteristics of the approximately 60 projects that identified known disease genes highlighted several factors that likely influenced this outcome including: (1) overall significant rarity of the syndrome (ultrarare); (2) disease presentations with high levels of genetic heterogeneity; (3) limited access to exclusionary genetic testing prior to enrollment in the study; and, (4) atypical clinical presentations. Further stratification of our results based on types of clinical presentation indicated that the likelihood of identifying a mutation in a known disease gene in a patient presenting with cerebellar ataxia, for example, was approximately 40%, while the chance of making a molecular diagnosis in a patient with neuropathy was significantly lower. Autosomal recessive inheritance of the disease was the most significant factor positively influencing success in these projects. Given that half of the 120 projects solved through the FORGE Canada initiative were known disease genes we have gained significant insight into the clinical utility of WES and potentially the depth of rare disease.

2006W

Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *S. Petrovski[1,2], Q. Wang[1], E.L. Heinzen[1], A.S. Allen[1,3], D.B. Goldstein[1].* 1) Center for Human Genome Variation, School of Medicine, Duke University, Durham, NC; 2) Departments of Medicine, Austin Health and Royal Melbourne Hospital, University of Melbourne, Australia; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC.

Using empirical polymorphism data from the NHLBI Exome Sequencing Project we introduce a genome-wide scoring system that ranks human genes in terms of their intolerances to functional genetic variation in the human population. We name this score the Residual Variation Intolerance Score (RVIS). It is often inferred that genes carrying relatively fewer common functional variants in healthy individuals may be judged more likely to cause certain kinds of disease. We show that this 'intolerance score' correlates remarkably with genes already known to cause Mendelian diseases (P=10-27). Equally striking, however, are the differences in the relationship between standing genetic variation and disease causing genes for different disease types. Considering disorder classes defined by Goh et al (2007) human disease network, we show a nearly opposite pattern for genes linked to developmental disorders and those linked to immunological disorders, with the former being preferentially caused by genes that do not tolerate functional variation and the latter caused by genes with an excess of common functional variation. We conclude by showing that use of an exome-wide intolerance ranking framework can facilitate interpreting personal genomes and can facilitate identifying high impact mutations through the gene in which they occur.

Application on OMIM disease genes

	OMIM disease genes	OMIM recessive	OMIM Haploinsufficiency	OMIM de novo	OMIM Haploinsufficiency and de novo	Essential Gene List (Georgi et al (2013))
Number of genes	2131	817	175	467	108	2288
Intolerance Score (RVIS): logistic regression	1.4×10-27 β= -0.29	3.3×10-5 β = -0.16	1.6×10-31 β = -0.71	2.7×10-36 β = -0.57	1.4×10-28 β = -0.77	1.3×10-114 β = -0.63
Intolerance Score (RVIS) AUC [95% CI]	0.58 [0.57-059]	0.55 [0.53-0.57]	0.73 [0.70-0.77]	0.65 [0.62-0.67]	0.78 [0.74-0.82]	0.66 [0.65-0.67]

.

2007T

Analysis of nucleotide changes in miRNAs genes along primate evolution.
I. Balcells[1], M. Mele[2], T. Marques-Bonet[1,3], Y. Espinosa-Parilla[1]. 1) Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Spain; 2) Centre for Genomic Regulation (CRG), Barcelona, Spain; 3) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

MicroRNAs (miRNAs) are important regulators of gene expression at post-transcriptional level. Although lineage-specific miRNAs may contribute to shape primate evolution; how it happens and what are the specific miRNA changes that contribute to this process is not well understood yet. Using genomic data, we have compared the 1,595 precursor miRNAs (pre-miRNAs) described in humans (miRBase v19) in several Great Ape populations including humans (n=9), bonobos (n=13), chimpanzees (n= 25), gorillas (n= 27) and orangutans (n=14). Our data showed that a total of 1,782 single nucleotide variants (SNV) occurred in pre-miRNA; 590 in the mature miRNAs and 198 in the miRNA seed region. Surprisingly, although it has been described that miRNA seeds are highly constrained, in this study, SNV densities (SNVD) do not differ between the seed region (SNVD= 0.023), the mature miRNA (SNVD=0.022) and all pre-miRNA (SNVD=0.024). And, interestingly, the distribution of nucleotide changes was seen not to be uniform. They are accumulated in 734 (65.4%) pre-miRNAs, 429 (28.9%) mature miRNAs and 170 (11.9%) seed miRNAs. In the branch going from the chimpanzee-human ancestor to humans, 23 substitutions have been fixed in the seed region of miRNAs. These changes could affect directly miRNA function by modifying the number and spectrum of target genes. Some of these nucleotide changes are predicted to provoke important gain or losses of human miRNA targets according to TargetScan prediction. Nevertheless, expression patterns for these miRNAs and their target genes should be further analyzed to decipher their function. On the other hand, 84 nucleotide changes affected mature miRNAs in the human lineage. Out of them, twenty are expressed in specific regions of the human brain. These nucleotide changes could affect miRNA expression levels and thus, the fine-tuning of brain molecular regulatory networks could be altered. Overall, our work describes that many nucleotide changes have been accumulated during Great Ape evolution in different functional regions of miRNAs. However, the great challenge now is to understand which are the functional implications of these changes.

2008F

Mitochondrial haplogroup B and oxidative stress associations with antiretroviral-associated peripheral neuropathy in Thai individuals.
R. Levinson[1], T. Hulgan[2], M. Gerschenson[4], N. Phanuphak[3, 5], J. Ananworanich[3,4,5,6], A. Baker[2], V. Valcour[7], D.E. LiButti[4], T. Jadwattanakul[8], D. Murdock[1], J. McArthur[9], C. Shikuma[4], D.C. Samuels[1], the SEARCH 003 Study Team. 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Vanderbilt University, Nashville, TN; 3) South East Asia Research Collaboration with Hawaii, Bangkok, Thailand; 4) University of Hawaii, Honolulu, Hawaii, USA; 5) Thai Red Cross AIDS Research Centre, Bangkok, Thailand; 6) HIV-NAT, Bangkok, Thailand; 7) University of California- San Francisco, San Francisco, CA; 8) Queen Savang Vadhana Memorial Hospital, Chonburi, Thailand; 9) The Johns Hopkins University, Baltimore, MD.

SEARCH 003 was a randomized clinical trial in Thailand designed to evaluate mitochondrial (mt) toxicity of first-line antiretroviral (ARV) regimens in HIV-infected persons. mtDNA sequencing from PBMC was performed on 149 subjects with the AffymetrixGeneChip Human Mitochondrial Resequencing Array 2.0. The SEARCH 003 population fell into 3 mitochondrial haplogroup clades, B (15%), M (43%), and F (21%) determined by Phylotree; the remainder were categorized as 'other'. Participants had epidermal nerve fiber density (ENFD) assessed by distal leg skin biopsy at weeks 0 and 24 as a measure of peripheral neuropathy, an ARV-associated mt toxicity. MtDNA levels were measured by RT-PCR. MtDNA oxidative damage was measured by the gene specific repair assay for 8-oxo-deoxyguanine (8-oxo-dG). Analyses included Wilcoxon rank-sum test and logistic regression. Baseline mtDNA levels, ENFD, and treatment arm were not significantly different by mtDNA haplogroup. Haplogroup B individuals had an increase in median ENFD over 24 weeks compared to non-B individuals (p=0.02). Median 24-week PBMC 8-oxo-dG change also increased in haplogroup B individuals compared to non-B individuals (p=0.02). In a logistic regression model, haplogroup B was significantly associated with an increase in ENFD over 24 weeks (OR= 24.3, 95%CI = [4.95, 168]). In this model, other significant covariates were baseline CD4 count, age, and ENFD. These data show that haplogroup B individuals gained ENFD over the first 24 weeks of ARV treatment. However, for individuals with ENFD data the frequency of peripheral neuropathy was significantly higher in the haplogroup B individuals (p= 0.04), though the number of cases was small (N=6). In conjunction with increased mt oxidative damage in haplogroup B, this leads us to hypothesize that the gain in ENFD may be a compensatory response to mt damage. Follow up studies are needed to validate these associations and elucidate mechanisms by which mtDNA variation influences response to therapy. Most haplogroup B individuals in our study were subgroup B5, though any role of subgroup in these responses remains unclear. Haplogroups are closely related to continental origin and their global distribution follows historical human migrations. Haplogroup B is found in individuals of East Asian descent and Native American lineages, including in about 15% of Hispanics. Follow up studies in Southeast-Asian and Hispanic HIV/AIDS populations would be worthwhile.

2009W

Defects in the autophagy pathway contribute to glaucoma caused by mutant myocilin accumulation. *X. Lin[1,2], G. Zode[1,2], C.C Searby[1,2], V.C Sheffield[1,2].* 1) Pediatrics, The University of Iowa, Iowa City, IA; 2) Howard Hughes Medical Institute, MD.

Autophagy is a cellular degradation pathway that involves the delivery of unnecessary or dysfunctional cellular components to the lysosome. Autophagy is often upregulated in circumstances that result in protein aggregate accumulation. Autophagy plays a role in the clearance of aggregate-prone mutant proteins associated with several neurodegenerative diseases in which accumulation of unfolded protein is the major characteristic. Early reports demonstrated that autophagy contributed to the pathogenesis of Alzheimer disease, Parkinson disease and Huntington disease. However, little is known about autophagy in glaucoma, including glaucoma caused by myocilin mutations. It is well documented that mutant myocilin accumulates in the ER of the trabecular meshwork and leads to an increase in intraocular pressure, which is associated with retina gangalion cell degeneration. We have investigated the role of autophagy in clearing mutant myocilin aggregates. We generated stable HEK 293T cell lines expressing mutant and wild-type myocilin. We confirmed that mutant myocilin accumulates and leads to increased ER stress, while wild type myocilin does not accumulate or result in ER stress. In addition, mutant myocilin leads to increased LC3-I levels compared to wild type myocilin and control cells. We also observed increased levels of p-Akt, beclin-1, (responsible for vesicle elongation), and Atg-4 (pivotal to autophagsome membrane formation) in mutant myocilin cells, while no change was observed in p-mTOR (negative regulator of autophagy), Atg 5 (involved in autophagosome formation), or Atg 3 (involved in LC3-I conjugation). When treated with chloroquine, a lysosome basifying agent which blocks the fusion of autophagosomes with lysosomes, we observed exacerbation of myocilin accumulation only in mutant myocilin cells, but an increase in LC3-II/LC3I ratios only in control cells. Increased LC3-I levels in mutant myocilin cells compared to cells expressing WT myocilin indicate that these cells sense increased myocilin aggragate levels, but that autophagosomes could not form properly. These data indicate that autophagy plays a role in the clearance of mutant myocilin, but is ineffective.

2010T

Modeling Mutation Events for Complex Variants in Human Genome. *M. Bhuyan, I. Pe'er.* Columbia University, New York City, NY.

Background: Availability of whole genome sequences paves the way to observing sites that harbor complex, multi-allelic variation. Such sites may be considered as candidates for genetic association studies. By definition, complex variants (CVs) require the occurrence of multiple mutation events, often including indels and repeats besides single nucleotide substitutions. In this study we focus on cataloging and categorizing CVs, modeling implied mutation events, and reconstructing ancestry among alleles. Methods: We use variant information from a cohort of 69 individuals from 12 different ethnicities made publicly available by Complete Genomics. Specific quality control steps were taken to ensure the vast majority of observed complex variants are genuine. We classify CVs into categories guided by likely types of mutation events. We define a probabilistic model for ancestral mutation events, assigning each type a rate of occurrence. We estimate the parameters of this model along with local trees that describe the history of all the alleles associated with corresponding CVs. This is achieved by iteratively finding a minimum weight local trees and updating the model from the mutation events across edges of all local trees. The trees are rooted by alignment to the chimpanzee sequence, aided by an efficient ad hoc procedure to resolve repeat boundaries. Results: The observed frequency of CVs is considerably higher than expected by chance. We identify several common categories of CVs, each providing evidence for particular types of mutation events. (i) Simple repeat regions with variable count of their repeat unit make up most CVs. These include homopolymers (~36% of CVs) as well as microsatellites (~6%). (ii) Short CVs, all of whose aligned alleles are ≤3bp in length. These include tri/tetra-allelic SNPs (~7%) di/tri-nucleotide polymorphisms (~24%). (iii) Block polymorphisms (~1%) can be expressed as combinations of two or more SNPs, usually linked. Mutation events in our model include, in addition to point substitutions and point indels, specific types for period indels, block indels, and more specifically, block indels at the end of the allele sequence. Mutation rates depend on context of mutated and flanking single nucleotides, and are fit to the data as a matrix.

2011F

Modeling variability in the population mutation rate using 1000 Genomes data for discovery and interpretation of rare variation. *V. Aggarwala, B. Voight.* Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA.

An emerging, major challenge is the identification of pathogenic, rare mutations contributing to complex disease, as this class of alleles offers great potential to accelerate our understanding of mechanisms and disease etiology. Despite the increasing evidence of the role of noncoding variation in disease, few computational approaches are designed to aid the interpretation of mutations in this large fraction of the genome. New computational approaches design to understand the biological features that govern where mutations occur, how many occur, and how they persistence over time (i.e., selection) in human populations at broad and fine scales will be essential to the disease discovery and interpretation process. To address this shortcoming, we propose a statistical framework to model and assess the variability in the population mutation rate at broad and fine scales. The generalized statistical model is Bayesian in nature, and uses repeated sampling from standard distributions to predict the mutation spectrum and also assign confidence interval to different types of mutations, in any number of individuals from different populations. Based on fundamental population genetics quantities, we provide a framework to hypothesis test and incorporate new biological features and annotations like DNase hypersensitivity, differential selection, epigenetic marks etc. that have been measured empirically or speculated. As a proof-of-concept, using data from Phase I of the 1000 Genomes Project, we demonstrate that sequencing context is a major feature that explains the variability in the population mutation rate at all genomic scales ($P<2.2E-308$), and confirm expected population-size reductions on the X chromosome ($P<7.3E-182$) and in comparisons of between African and non-African population ($P<5.7E-204$). Furthermore, we show that our analysis framework facilitates discovery of previously unknown features influencing the spectrum of observed genetic variation. Our findings suggest that this basic framework can be used to model and predict the genetic variant spectrum in any region of the genome, information which is critical to building appropriately calibrated rare variant association tests in the non-coding (and coding) genome for complex disease.

2012W

Neanderthal Introgression at Chromosome 3p21.31 was Under Positive Natural Selection in East Asians. *Q. Ding[1], Y. Hu[1], S. Xu[2], J. Wang[1], L. Jin[1,2].* 1) State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 2) CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

Studies on Neanderthal and Denisovan genomes demonstrate archaic hominin introgression in modern Eurasians. In the current study, we present evidence of Neanderthal introgression at 3p21.31 region. Frequency of introgressed haplotypes was high in East Asians, ranging from 49.4% to 66.5%, and was low in Europeans. Signal of strong positive selection was detected in this region only in East Asians. Expansion of the introgressed haplotypes started at 37.4 KYA, suggesting the starting point of selection. We showed that rs12488302-T or its associated alleles might be the candidate targets of selection, among which four are non-synonymous, including rs35455589-G of in *HYAL2*, a gene related to cellular response to UV-B. Furthermore, suggestive evidence supports latitude-dependent selection, implicating the role of UV-B. Interestingly, the distribution of rs35455589-G suggests that this allele was lost during the exodus of modern human from Africa, and reintroduced to Eurasians from Neanderthals.

2013T

Homozygous losses of human-specific neural *SRGAP2C* gene discovered in patients with intellectual disability. *X. Nuttle[1], K. Witherspoon[1], C. Baker[1], B. Coe[1], M. Fichera[2, 3], J. Schuurs-Hoeijmakers[4], R. Bernier[5], J. Gecz[6, 7], B.B.A. de Vries[4], C. Romano[2], J. Shendure[1], E. Eichler[1, 8].* 1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 2) Regional Center for Genetic Rare Diseases with Intellectual Disability or Brain Aging, IRCCS Associazione Oasi Maria Santissima, Via Conte Ruggero, 73, 94018 Troina, Italy; 3) Medical Genetics, University of Catania, Catania, Italy; 4) Department of Human Genetics, Radboud University, Nijmegen, Netherlands; 5) Department of Psychiatry, University of Washington, Seattle, WA 98195, USA; 6) Department of Paediatrics, The University of Adelaide, Adelaide, Australia; 7) Genetics and Molecular Pathology, and South Australian Pathology at Women's and Children's Hospital, Adelaide, Australia; 8) Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, WA 98195, USA.

Incomplete duplications of *SRGAP2*, a gene involved in regulating neuronal migration and development of dendritic spines, are hypothesized to have played an important role in human brain evolution. Several observations support this hypothesis, including the evolutionary emergence of the duplicate genes at the onset of *Homo* encephalization (~2-3 million years ago), fixation of one duplicate (*SRGAP2C*) in the human population, its expression in fetal brain, and striking effects on neuronal development when introduced into mice. Nevertheless, no direct genetic evidence in humans has yet linked disruptive mutations of these human-specific paralogs with neurodevelopmental disease phenotypes. We recently developed a novel high-throughput method using molecular inversion probes to accurately genotype duplicated genes previously intractable to large-scale genetic analysis, such as *SRGAP2*. We surveyed copy number and sequence variation in *SRGAP2* genes in a diversity panel of 1,056 humans and discovered a rare ~38 kbp deletion in *SRGAP2C* observed exclusively in populations of European descent (~2% allele frequency). Using this approach, we genotyped all *SRGAP2* paralogs in 3,153 individuals with intellectual disability and 715 ethnically-matched controls. We identified five individuals having intellectual disability with complete loss of *SRGAP2C* due to frameshifts~four with the same ~38 kbp deletion including exon 2 and one individual with a single base deletion in exon 3. All five individuals are homozygous, likely due to increased consanguinity, and are predicted to have completely lost the human-specific *SRGAP2C*. In contrast, we never observed homozygous loss of *SRGAP2C* in controls. Three individuals with the homozygous exon 2 *SRGAP2C* deletion additionally exhibit microcephaly, consistent with the gene's function in neuronal migration and increasing dendrite density. We are continuing our large-scale genotyping efforts to confirm the pathogenicity of complete *SRGAP2C* loss and gain further insight into phenotypes accompanying it. These results provide further support for the importance of this new human gene for the evolution of unique aspects of human brain development.

2014F

Most Autosomal Recessive Diseases Have No Frequent Mutations that Reflect Increased Carrier Fitness or Mutation Hotspots. *M.AH. Gener, R.V. Lebo.* Pathology, Akron Children's Hospital , Akron, Oh.

Heterozygous advantage results when the carrier of a mutant gene has an increased reproductive fitness over a noncarrier. Then the relative frequency of initially occurring mutant alleles increase more rapidly in subsequent generations than the frequency of subsequently mutated alleles. Increased frequencies of specific mutant alleles have been reported in diseases exhibiting increased carrier fitness as well as in a disease gene with mutation hotspots. Carriers of one cystic fibrosis (CFTR) mutant allele are reported to exhibit increased fitness over noncarriers when infected with diseases like cholera. Carriers of the sickle cell anemia allele and related α-and β-thalassemia alleles survive malarial infection more readily in regions cultivating rice. This selective advantage explains why the ΔF508 mutation comprises ~70% of cystic fibrosis alleles and the sickle cell mutation βs comprises ~90% of sickle cell alleles. Alternatively, spontaneous mutations in the MECP2 Rett syndrome gene occur at mutation hotspots related to specific de novo, lethal gene sequences.(Lebo et al, 2001) The 34 listed autosomal recessive disorders with frequencies exceeding 1 in 100,000 worldwide patients were identified through GeneReviews and further studied through the Human Gene Mutation Database and Online Mendelian Inheritance in Man websites. No evidence of substantial selection was found for any of these disorders that had between 34 mutations in Kyphoscoliotic Ehlers-Danlos Syndrome and 1009 mutations in Leber Congenital Amaurosis. In contrast, hemochromatosis reported in the large northern European Caucasian regional population results from the Cys282Ty mutation in >90% of known alleles, possibly reflecting an as yet unreported selective carrier survival advantage. Furthermore, allelic selection could readily explain the 10% of all individuals who carry the 5T CFTR allele which decreases CFTR gene expression. 5T/5T alleles are found in patients with CBAVD. The lack of common alleles in most diseases makes analysis of multiple disease genes in the worldwide population more complex. In conclusion, inherited heterozygote carrier frequencies in very large patient populations typically reflect the length of time the mutation existed since its occurrence in the human population and whether a founder allele has been selected. Very frequent single mutations are not anticipated in the absence of heterozygous advantage or mutation hotspots.

2015W

Multiple LD-independent signals of extreme sub-population variation at a region associated with type-2 diabetes suggests a non-neutral evolutionary history. *P.L. Babb[1], B.F. Voight[1, 2].* 1) Pharmacology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Type-2 diabetes (T2D) is a complex metabolic disease affecting over 280 million people worldwide. Although >70 loci have been associated with T2D susceptibility, we are only beginning to understand the range of T2D allelic variation across global populations at those loci, which gene(s) and processes are functionally modulated, and how they ultimately contribute to T2D pathogenesis. Furthermore, the evolutionary history of established T2D-associated variation has not been fully evaluated. Recent work has suggested that T2D variants harbor extreme levels of population differentiation (1), consistent with the hypothesis that T2D susceptibility alleles may have once been advantageous for regulating metabolic load in early human populations. In order to comprehensively test this hypothesis using all established loci for T2D, and to elucidate global distribution of risk alleles at T2D-associated loci, we computed Weir and Cockerham's (1984) formulation of F_{ST} for 69 LD-independent T2D loci ($p<5\times10$-8) in 1088 unrelated healthy individuals from 14 populations studied by the 1000 Genomes Project (Phase I). We used different combinations of sub-population membership to calculate global F_{ST} (AFRxAMRxASNxEUR), intra-continental F_{ST} (e.g. intra-ASN: CHBxCHSxJPT), and pairwise-continental F_{ST} (e.g. AFRxEUR) for each SNP. We computed the empirical distribution of F_{ST} (from 3.6 M HapMap SNPs) for all individuals in each type of F_{ST} comparison to examine whether any of the T2D loci fell within the tail of the F_{ST} distribution. In contrast to previous observations, our results suggest that as a group, the 69 T2D loci do not have greater F_{ST} values than a matched-sampling of SNPs (10,000 random draws) from the rest of the genome (Welch 2-sample t-test, p=0.97). However, upon closer examination of individual T2D loci, we noted one locus showing extreme F_{ST} scores across all population comparisons ($p<0.01$), and exhibiting complex patterns of F_{ST} and LD within 1 Mb of the associated SNP nearby the *PRC1* locus. In particular, our investigation revealed at least three LD-independent clusters of high F_{ST}, all within the $p<0.05$ significance cutoff of empirical distribution in the genome (global $F_{ST}>0.26$). We postulate that by triangulating the signals of such LD-independent F_{ST} clusters derived from non-coding disease-associated SNPs, we can help identify the causal gene(s) functionally involved in the manifestation of T2D. 1. Chen, et al., *PLoS Genetics*, 2012.

2016T

Analysis of type 2 diabetes and obesity variants in Mexican Pima Indians: the Maycoba Project. *R.L. Hanson[1], L.O. Schulz[2], P.H. Bennett[1], S. Kobes[1], R.C. Williams[1], M.E. Valencia[3], J. Esparza-Romero[3], R. Urquidez-Romero[3], L.S. Chaudhari[2], E. Ravussin[4], W.C. Knowler[1], L.J. Baier[1].* 1) Diabetes Epidemiology Clin Res, NIDDK, Phoenix, AZ; 2) Northern Arizona University, Flagstaff, AZ; 3) Centro de Investigación en Alimentación y Desarrollo, Hermosillo, Sonora, Mexico; 4) Pennington Biomedical Research Center, Baton Rouge, LA.

The Pima Indians of Arizona have a very high prevalence of type 2 diabetes mellitus (T2DM) and obesity, while Mexican Pima Indians from Maycoba, Sonora, have a much lower prevalence of these disorders. While there are differences in environment between populations, the extent to which they differ at genetic loci for T2DM and obesity is unknown. We genotyped representative population samples of 175 Pimas from Maycoba and 402 from Arizona at 253 putative and established T2DM and obesity variants, along with 96 randomly selected single nucleotide polymorphisms (SNPs). Genetic distance between Mexican and Arizona Pimas, estimated from F_{ST} across the random markers, was 0.031 (95% confidence interval 0.019-0.048). Distance across 46 established T2DM variants or across 32 established obesity variants did not differ significantly from that across random markers (F_{ST}=0.038, p=0.53, and F_{ST}=0.054, p=0.36, respectively). Adjustment for a multiallelic score that sums number of risk alleles across established T2DM variants did not attenuate the increased risk for T2DM in Arizona Pimas, nor did a similar score for obesity attenuate the population difference in body mass index. To identify individual variants that differ between populations, absolute value in allele frequency difference (δf) was calculated and statistical significance was assessed with genomic control, calculated over the random markers, to account for expected differences. Five markers showed significant (false discovery rate<0.05) differences in frequency: 4 in *HLA-DRB1* (δf=0.56-0.75) and one (rs117619140, δf=0.40, $p=3.8\times10^{-4}$) in *TREH*. The largest difference was at the *HLA-DRB1* variant rs9271720 (δf=0.75, $p=1.3\times10^{-7}$). Analysis of data from the Human Genome Diversity Project suggests that this magnitude of allele frequency difference is unusual between human populations of comparable genetic distance: among 55,304,216 allele frequency differences analyzed across 101 pairs of populations with F_{ST} 0.019-0.048, the proportion with δf≥0.75 was 4.0×10^{-7}. These analyses show that established T2DM and obesity-associated genetic variants explain little of the difference in risk between Mexican and Arizona Pimas, and in general do not show greater divergence than expected based on genetic distance. However, variants at *TREH* and *HLA-DRB1*, genes which have previously been associated with T2DM in Arizona Pimas, have greater allele frequency differences than expected, perhaps reflecting effects of selection.

2017F

Evolution, adaptation and disease perspectives of Indian populations. *S. Nizamuddin[1], LVKS. Bhaskar[2], P.B. Gai[3], BG. Venkatesh[2], M. Jyotsna[2], K. Thangaraj[1].* 1) Center for Cellular and Molecular Biology, Hyderabad, Andhra Pradesh, India; 2) Sri Ramachandra University, Porur, Chennai - 600 116, India; 3) Department of Studies in Applied Genetics, Karnataka University, Dharwad-580003, India.

India has been a major corridor for the first modern human, who have migrated out-of-Africa. Since India is inhabited by more than 4,500 anthropologically well-defined populations, who follow strict endogamy marriage practice, they harbors unique genetic architecture compare to populations from rest of the world. India is known not only for its diversity, in terms of social structure, language and culture but also its varied eco-geographical conditions, such as high altitude (Himalayan region: Changkpa population in Ladakh - 5294 meter), sea level (Lakshadweep island: Minicoy population - 10 meter; Andaman and Nicobar: Onge), extreme cold (Himalayan region: Kashmiri Pandit, Changkpa) and moderately warm temperature (Andhra Pradesh: Mala, Madiga, Vysya). In addition, some populations are living in islands while others are in mainland. Since the Indian populations are inhabited in these geographical regions for tens of thousands of years, we hypothesize these populations might have different selective advantages for certain genetic markers. To fetch signal of natural selection in these population, we have genotyped 134 individual on Affymetrix 6.0 platform and did phasing with Beagle software. Further, we calculated XP-CLR and XP-EHH values, and found novel signal. Interestingly, SFRP1, HIRA, HTR2B, ADD1, FGF1, SIM1, SLC7A10, CEBPA, FASLG and ZFPM2 genes are some examples, which are having importance in survival and genetic fitness for continuing progeny. In addition, we have also found novel genetic factors for many diseases, including cleft palate, breast cancer, etc. Details of these finding would be made available during the presentation.

2018W

Whole genome sequencing reveals past population history and signature of natural selection in a Japanese population. *A. Fujimoto[1], T. Abe[1], K. Boroevich[1], K. Nakano[1], A. Sasaki[1], R. Kitada[1], H. Tanaka[2], Y. Nakamura[1], S. Miyano[2], M. Kubo[1], H. Nakagawa[1], T. Tsunoda[1].* 1) Ctr for Integrative Medical Sciences, Riken, Yokohama, Kanagawa, Japan; 2) Human Genome Ctr, The Institute of Medical Science, The University of Tokyo, Japan.

We performed whole genome sequencing of 164 Japanese individuals with 30X coverage, and identified single nucleotide variants, insertions and deletions, copy number variations and rearrangements. We analyzed the pattern of genetic variations and estimated the past population history. Our analysis identified two bottleneck events and a population expansion after the last glacial period. The analysis of the frequency spectrum and the linkage disequilibrium detected signatures of recent positive selection. Significantly higher genetic variation was observed in the HLA and olfactory receptor regions. Our analyses showed the past population history of the Japanese population and identified the signature of adaptation in the local environment of East Asia.

2019T

Integrated natural selection and genome-wide association scan for cholera resistance in Bangladesh. *E.K. Karlsson[1,2], S. Tabrizi[1,2], I. Shylakhter[1,2], F. Qadri[3], J.B. Harris[4,5], R.C. LaRocque[4,6], P.C. Sabeti[1,2,7].* 1) Broad Institute, Cambridge, MA; 2) Center for Systems Biology, Harvard Univ, Cambridge, MA; 3) International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; 4) Div. of Infectious Diseases, Massachusetts General Hospital, Boston, MA; 5) Dept of Pediatrics, Harvard Medical School, Boston, MA; 6) Dept of Medicine, Harvard Medical School, Boston, MA; 7) Dept of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA.

As an ancient disease with high fatality, cholera has exerted strong selective pressure on affected human populations, particularly in the Ganges River Delta, where the disease is endemic. This history of selective pressure makes it easier to identify host variants associated with cholera susceptibility and severity. We previously developed a method, the Composite of Multiple Signals (CMS), which combines different tests for natural selection to provide 20-100x better positional resolution than any individual signal. CMS scores each variant, allowing precise identification of the top functional candidates. We have now combined CMS with genome-wide association to identify selected variants associated with cholera resistance in Bangladesh. We initially performed a genome-wide study of natural selection in a population from the Ganges River Delta and identified 305 candidate selected regions using CMS. The regions were enriched for potassium channel genes involved in cyclic AMP-mediated chloride secretion and for components of the innate immune system involved in NF-κB signaling. Testing the top selected regions in two independent Bengali cohorts found the strongest association with cholera susceptibility in the top region of selection, peaking in the gene SNRNP200. We used these results to develop a model of the human innate immune signaling pathways that respond to V. cholerae infection and have been selected in the Bengali population. In this model, inflammasome activation and the NF-κB signaling pathway play an integrated role in TLR4-mediated sensing of V. cholera - consistent with our in vitro data. We have now expanded this work by combining a genome-wide association scan for cholera susceptibility with a denser selection scan using full genome sequences for the Bengali population provided by the 1000 genomes project, and have identified new host immune factors likely influencing cholera resistance. Our work shows that using publically available genetic datasets to incorporate tests for natural selection into GWAS analyses can help pinpoint functional variants, an approach applicable to other historically prevalent infectious diseases, such as Lassa fever, tuberculosis, leishmaniasis and malaria, and to complex, common diseases, such as inflammatory bowel disease, for which the associated genes may have been historically selected.

2020F

Targeted analysis of immunogenetic diversity in pre-Columbian Central Andean populations. *B. Llamas, G. Valverde-Garnica, A. Cooper, W. Haak.* Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia.

Historic accounts estimate that the South American indigenous population size declined by up to 95% between the period of initial contact with Europeans and the beginning of recovery. Most archaeologists and epidemiologists agree that Old World diseases played a major role in this abrupt population collapse. We used advanced ancient DNA techniques to characterize the genetic diversity of loci involved in immune response in pre-Contact Central Andean populations. We have created immortalized ancient DNA libraries using human remains from various periods ranging from the Archaic Period to the Late Horizon. Results from shotgun sequencing showed that nuclear information and complete mitochondrial genomes could be retrieved from well-preserved samples. We then used a customized targeted enrichment protocol and Next Generation Sequencing to genotype thousands of immunogenetic markers to generate a unique genetic dataset. By contrasting this major genetic survey of pre-Columbian immunogenetic diversity with modern-day populations, which harbour a low (immuno-) genetic diversity, we can test whether indigenous populations were immunologically 'naïve' to Old World diseases, and as a consequence suffered very high levels of mortality.

2021W

Learning Natural Selection from the Site Frequency Spectrum. *R. Ronen[1], N. Udpa[1], E. Halperin[2,3,4], V. Bafna[5].* 1) Bioinformatics and Systems Biology, University of California at San Diego, La Jolla, CA; 2) The Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel; 3) International Computer Science Institute, Berkeley, CA, USA; 4) Department of Molecular Microbiology & Biotechnology, Tel-Aviv University, Tel-Aviv, Israel; 5) Department of Computer Science & Engineering, University of California, San Diego, CA, USA.

Genetic adaptation to external stimuli occurs through the combined action of mutation and selection. A central problem in genetics is to identify loci responsive to specific selective constraints. Many tests have been proposed to identify the genomic signatures of natural selection by quantifying the skew in the the site frequency spectrum (SFS) under selection relative to neutrality. We build upon recent work that connects many of these tests under a common framework, by describing how selective sweeps impact the scaled SFS. We show that the specific skew depends on many attributes of the sweep, including the selection coefficient and the time under selection. Using supervised learning on extensive simulated data, we characterize the features of the scaled SFS that best separate different types of selective sweeps from neutrality. We develop a test, SFselect, that consistently outperforms many existing tests over a wide range of selective sweeps. We applied SFselect to polymorphism data from a laboratory evolution experiment in which populations of Drosophila melanogaster adapted to hypoxia over hundreds of generations. As a result, we identified loci that support the role of the Notch pathway in hypoxia tolerance. Importantly, several of these Notch-related loci (including the Notch gene region itself) were missed by a previous study. We further applied our test to whole genome sequence data from two human populations. We identified many regions evolving under positive selection, some of which are in agreement with earlier studies, but many of which are novel.

2022T

Reconstructing bacterial phylogeny using whole-genome deep-sequencing data. *Y. Lo[1], L. Zhang[2], B. Foxman[2], S. Zöllner[1,3].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Epidemiology, University of Michigan, Ann Arbor, MI; 3) Psychiatry, University of Michigan, Ann Arbor, MI.

Studying the genetic variation among bacterial strains from the human biome gives insight into the origins and transmission of bacterial infections in human communities, and the emergence and spread of virulence and antibiotic resistance in bacterial populations. While sparse genetic data are sufficient to confidently cluster strains into a 'global' phylogeny, a single tree does not fully describe bacterial evolutionary relationships. Ancestries of individual bacterial genes can differ substantially from the global ancestries because of horizontal gene transfer (HGT) and homologous recombination. Comparisons of individual gene trees to the global tree give signals of HGT. Next-generation sequencing methods allow examination of entire bacterial genomes at high resolution and reasonable cost. However, individual gene tree construction often relies on few variant sites on the gene and hence still suffers from lack of data resolution. Here we propose a method to accurately construct and make inferences from gene trees using whole-genome deep-sequencing data of a set of bacterial strains. First we generate the posterior distribution of the global tree using standard Bayesian phylogenetic methods. This global tree posterior distribution serves as prior information for each gene tree. We describe a novel algorithm combining this prior distribution with the likelihood of each gene tree calculated from gene-specific variant calls. Comparing each most probable gene tree to the global tree gives an estimate of the concordance proportion. Bayes factor between each gene tree and the global tree provides a metric for identifying gene phylogenies significantly different from the global phylogeny. We apply the method to a sample of uropathogenic and commensal *Escherichia coli* strains obtained from human patients with urinary tract infections, each sequenced whole genome at >190x. After constructing global and gene trees, we quantify the proportion of the genes having consistent trees with the global tree. Most known virulence genes are not present in all strains. Comparing the gene-trees of these virulence genes to the global tree gives evidence of how often HGT occurs in the sample. Our method provides a novel approach for constructing high confidence gene trees. The results suggest an ongoing circulation of pathogenic genes, and multiple genetic origins of uropathogenic strains in a small geographic region.

2023F

The adaptive variant EDARV370A is associated with straight hair in East Asians. *J. Tan[1], Y. Yang[1], K. Tang[2], P. Sabeti[3,4], L. Jin[1], S. Wang[2,3,4].* 1) Fudan University, Shanghai, China; 2) Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai, China; 3) The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA.

Hair straightness/curliness is a highly heritable trait among human populations. Previous studies have reported European specific genetic variants influencing hair straightness, but the genetic variants influencing hair straightness in East Asians remain unknown. One promising candidate is a derived coding variant of the ectodysplasin A receptor (*EDAR*), *EDARV370A* (*370A*), associated with several phenotypic changes of epidermal appendages. One of the strongest signals of natural selection in human genomes, *370A* has risen to high prevalence in East Asian and Native American populations, while being almost absent in Europeans and Africans. This striking frequency distribution and the pleiotropic nature of *370A* led us to pursue if hair straightness, another epidermal appendage-related phenotype, is affected by this variant. By studying 1,718 individuals from four distinctive East Asian populations (Han, Tibetan, Mongolian, and Li), we found a significant association between *370A* and the straight hair type in the Han ($p = 2.90\times10^{-6}$), Tibetan ($p = 3.07\times10^{-2}$), and Mongolian ($p = 1.03\times10^{-5}$) populations. Combining all the samples, the association is even stronger ($p = 5.18\times10^{-10}$). The effect of *370A* on hair straightness is additive, with an odds ratio of 2.05. The results indicate very different biological mechanisms of straight hair in Europe and Asia, and also present a more comprehensive picture of the phenotypic consequences of *370A*, providing important clues into the potential adaptive forces shaping the evolution of this extraordinary genetic variant.

2024W

Macrophages from African and European populations respond differently to bacterial infection. *Y. Nédélec[1,2], A. Pagé Sabourin[1,2], V. Yotova[1], J.C. Grenier[1], N. Cotta[1], L.B. Barreiro[1,2]*. 1) Research Center, CHU Sainte-Justine, Montréal, Canada; 2) University of Montreal, Montréal, Canada.

Infectious diseases have always been a major health problem throughout the world, imposing strong selective pressure on the human genome. Geographically distinct human populations are postulated to have differing histories of pathogen exposure. Indeed, previous studies demonstrate that people of African and European ancestry differ in their susceptibility to certain infectious diseases like tuberculosis, malaria and sepsis. Differences in infection progression between African and European populations suggest inter-population variation in the immune response, possibly caused by the adaptation of Africans and Europeans to the pathogens of their environment. For the first time, we characterize the immune response of people of African and European ancestry to bacterial infections. We infected monocyte-derived macrophages from 24 African Americans and 24 European Americans with the intracellular pathogens Listeria monocytogenes and Salmonella typhimurium for 4 hours and measured whole genome gene expression of infected and non-infected cells by RNA-sequencing. We assessed macrophage control of bacterial infection at 1 hour and 24 hours by culturing infected cell lysate and counting colony-forming units to approximate the rate of bacterial survival. We found that macrophages derived from people of African ancestry presented fewer intracellular bacteria after 24 hours than people from European ancestry, suggesting that the African Americans better control intracellular bacterial infections. Concordantly with this observation we identified inter-population differences in immune gene pathway expression that might explain this pattern of increased infection control in African American macrophages. Interestingly, multiple genes up-regulated by bacterial infection in people of European ancestry were found to be already highly expressed in the non-infected cells of people from African ancestry. We show that several of these genes appear to have been subject to recent selection, which may explain between population disparities in their expression. Our study identifies multiple candidate genes that may affect the course of L. monocytogenes and S. typhimurium infection in humans. Importantly, our findings suggest that the clinical differences in bacterial infectious disease progression observed in populations of African and European ancestry may be outcome of natural selection.

2025T

Co-evolution of HLA class I with killer cell immunoglobulin-like receptors in a sub-Saharan African population. *P.J. Norman[1], J.A. Hollenbach[2], N. Nemat-Gorgani1[1], L.A. Guethlein[1], H.G. Hilton[1], M.J. Pando[3], K. Koram[4], E.M. Riley[5], L. Abi-Rached[1,6], P. Parham[1]*. 1) Department of Structural Biology, Stanford University School of Medicine, Stanford, CA; 2) Center for Genetics, Children's Hospital Oakland Research Institute, Oakland, CA; 3) Department of Pathology, Stanford University School of Medicine, Stanford; 4) Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Ghana; 5) Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK; 6) Centre National de la Recherche Scientifique, Laboratoire d'Analyse, Topologie, Probabilités, Equipe ATIP, Aix-Marseille Université, Marseille, France.

Human natural killer (NK) cell and cytotoxic T cell responses to pathogens, and the role of NK cells in reproduction, are controlled by diverse receptors that interact with highly polymorphic HLA class I molecules. Although T-cell receptor (TCR) variation is generated somatically with little genetic determination, killer cell immunoglobulin-like receptors (KIR) are highly polymorphic and polygenic and specific combinations of HLA class I and KIR alleles are differentially associated with numerous disease susceptibilities and therapeutic outcomes. Extensive *KIR* gene-content variation is observed between human populations. However, high-resolution analysis of *KIR* allele and haplotype diversity has focused on Asian and Amerindian populations having limited genetic diversity and the more heterozygous sub-Saharan African populations have remained under-represented. We studied KIR and HLA variation in 235 individuals, including 104 mother-child pairs from the Ga-Adangbe ethnic group of Ghana in West Africa. This population has a rich diversity of 175 *KIR* variants that form 208 *KIR* haplotypes and 81 *HLA-A*, *-B* and *-C* variants that form 190 *HLA class I* haplotypes. Balancing selection has maintained this exceptional polymorphism and diversified interaction of HLA class I with KIR, with evidence for their ongoing co-evolution in the Ga-Adangbe. Highlighting the critical role in countering pathogen diversity, every individual studied (N=188) has a unique compound genotype of *KIR* and *HLA class I*. Nevertheless, comparison of the observed frequencies with those simulated under a model of balancing selection revealed the impact of positive selection that has reduced the diversity of functionally interacting sites. Whereas the centromeric region of the Ga-Adangbe *KIR* locus is exceptionally diverse, the telomeric region lacks diversity due to the low frequencies of *Tel B KIR* genes and alleles and correspondingly high frequencies of *Tel A KIR* genes and alleles. As a consequence of the high frequency of HLA-B*53, -B*35 and related allotypes in the Ga-Adangbe, diversity is reduced at residues in the α1 domain of HLA-B that interact with the variable region of T-cell receptor β chains. These features of KIR and HLA-B are consistent with selection by a pathogen endemic to West Africa. The previously reported association of HLA-B*53 with severe malaria and the high prevalence of malaria in the Ga-Adange population strongly suggest *Plasmodium falciparum* as a candidate.

2026F

Host Genetics and Human Adaptation to Lassa Hemorrhagic Fever in West Africa. *S. Tabrizi[1,2,3,10], R. Tariyal[1,2,10], I. Shlyakhter[1,2], S.F. Schaffner[2], S.K. Gire[1,2], M.S. Stremlau[1,2], E.K. Karlsson[1,2], K.G. Andersen[1,2], E. Phelan[2], L.M. Moses[4], W. Omoniwa[5], I. Odia[5], P.E. Ehiane[5], O. Folarin[5,6], A. Tareila[1,2], L.M. Branco[4], J.S. Schieffelin[4], D. Levy[4,7], S. Gunther[8], D.S. Grant[7], G.O. Akpede[5], D.A. Asogun[5], P.O. Okokhere[5], R.F. Garry[4,11], C.T. Happi[5,6,11], P.C. Sabeti[1,2,9,11], Viral Hemorrhagic Fever Consortium.* 1) FAS Center for Systems Biology, Department of OEB, Harvard University, Cambridge, MA; 2) Broad Institute, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Department of Microbiology and Immunology, Tulane University, New Orleans, LA; 5) Institute of Lassa Fever Research and Control, Irrua Specialist Teaching Hospital, Edo State, Nigeria; 6) Department of Biological Sciences, College of Natural Sciences, Redeemer's University, Redemption City, Ogun State, Nigeria; 7) Lassa Fever Laboratory, Kenema Government Hospital, Kenema, Sierra Leone; 8) Lassa Fever Laboratory, Kenema Government Hospital, Kenema, Sierra Leone Department of Virology, Bernhard-Nocht-Institute for Tropical Medicine, Hamburg, Germany; 9) Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA; 10) These authors contributed equally to this work; 11) These authors jointly supervised this work.

Lassa virus has the unique status of being both one of the world's deadliest pathogens, designated bio-safety level 4 (BL-4), and a public health crisis, endemic in large parts of West Africa. It is estimated to have originated over 1000 years ago in Nigeria, and a genome-wide survey of human variation identified the gene *LARGE*, biologically linked to Lassa virus infection, as among the strongest signals of natural selection in the Yoruba population of Nigeria. We pursued the hypothesis that Lassa virus is an ancient selective force, driving the rise of genetic resistance, by conducting a genome-wide association study (GWAS) of Lassa fever cases and controls in Nigeria and Sierra Leone. We report a signal of association at *LARGE*, and show that it overlaps with the previously reported signal of selection. This suggests recent evolution of resistance to Lassa fever in West Africa, making it one of only a handful of known loci under positive selection elucidated to be involved in resistance to infectious disease. Population genetics analysis provides further insights into the origins and spread of the resistance variant in *LARGE*. We show that the same putatively protective haplotype under selection in the Yoruba is also present in the Esan population of Nigeria and the Mende population of Sierra Leone. We find that the frequency is highest in the Yoruba (30%), lower in the Esan (21%) and the Mende (10%), and absent in East African and non-African populations (where Lassa fever is also absent). This is consistent with results from our sequencing of the Lassa genome showing that the virus is oldest in Nigeria and appeared only recently in Sierra Leone, and helps to explain higher fatality rates and greater disease severity in Sierra Leone. Beyond *LARGE*, our other top GWAS signals fall at genes involved in the immune system, including *PTPRE*, which modulates expression of cytokines involved in host response to viral infection, and *IFRD1*, involved in the activation of neutrophils in innate immune response. We also examine signals of adaptation in the Esan and Mende populations and discover numerous novel candidate loci under positive selection, providing new insights into human diversity in West Africa. This work is one of only a few surveys of genetic diversity in West Africa, a region underrepresented in GWA studies. It is also the first ever GWAS on a BL-4 agent, and demonstrates a novel genomics approach to understanding the world's deadliest pathogens.

2027W

Characterizing positive natural selection in Asian urban and indigenous populations. *X. Liu[1], R.TH. Ong[2], W.Y. Saw[3], A. Mohammad[2], Y.Y. Teo[2,3,4,5], HUGO Pan-Asian SNP Consortium.* 1) NUS Graduate School, National University of Singapore, Singapore, Singapore 117456; 2) Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597; 3) Life Sciences Institute, National University of Singapore, Singapore 117456; 4) Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546; 5) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672.

The Pan-Asia SNP consortium is a collaborative work with genetic data from most countries of Asia, and contains data from not only the urbanized populations, but also ethnic minorities and a number of indigenous groups. The indigenous populations represent great diversity of cultural traditions, linguistics, and may possess unique demographic history. However, local adaptations of Asian indigenous groups were not systematically investigated in previous studies. Indigenous people's specific habitant and cultures make them subject to distinct selective pressures. Therefore, we expect to see distinct features of positive selection signals in the indigenous populations, as compared to other Asian urbanized populations. In this study, we aim to find positive natural selection signals in Asian population groups from PanAsia SNP consortium using HaploPS, iHS and XPEHH. We evaluated the power loss of selection methods and phasing accuracy resulting from low SNP coverage of genotyping data. A number of selection signals related to infectious disease and diet were found in 73 Asian population groups. The clustering analysis showed that the sharing of positive selection signals was affected by latitude. Populations residing near the equator tend to share the same positive selection signal, and the populations from further north ie Malaysia, Thailand, China etc also share similar signals. Indigenous populations are enriched with low frequency selection signals, whereas the urban populations have more medium to high frequency selection signals. This observation indicated that the indigenous populations may still undergo strong positive selection pressures and result in recent positive selection events.

2028T

Sensitivity of haplotype signatures of negative selection under different demographic scenarios. *D. Ortega Del Vecchyo[1], M. Ehm[2], M. Nelson[2], S. Zollner[3], K. Lohmueller[1, 4], J. Novembre[5].* 1) Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA; 2) Department of Quantitative Sciences, GlaxoSmithKline, Research Triangle Park, NC; 3) Department of Biostatistics, University of Michigan-Ann Arbor, MI; 4) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA; 5) Department of Human Genetics, University of Chicago, IL.

As documented by recent studies, haplotypic signatures can help distinguish a group of deleterious variants from a set of neutral variants. Differences in haplotype patterns are observed because deleterious variants are expected to have a younger age compared to neutral variants. Here, we explore the sensitivity of haplotypic signatures to distinguish between neutral and deleterious variants under different realistic human demographic scenarios. In order to do this, we developed a novel algorithm based on importance-sampling that allows us to efficiently simulate large numbers of rare variant haplotypes from arbitrary demographic scenarios. We show that the utility of haplotypic signatures is highly dependent on the demographic scenario being simulated. The reason for this is that the ages of deleterious and neutral alleles are also dependent on demography. We present a power analysis to quantify how many variants are necessary to differentiate deleterious variants with different selective coefficients from neutral variants. Additionally, we explore which summary statistics, such as iHS and r^2, provide highest sensitivity. We use those findings to identify deleterious variants in 202 genes analyzed in a large number of samples from Europe (n=12,514, Nelson et al., 2002). The size of the dataset enables us to resolve the haplotypic signatures of a large number of variants that are either putatively neutral or deleterious. We show how we can use these results to identify gene categories that have an enrichment of deleterious variants and hence are stronger candidates for contributing to disease. We also suggest how our results can be used to infer the distribution of selective coefficients.

2029F

Identification of regions under positive selection in the Gullah African American population of South Carolina. *P.S. Ramos[1], S.P. Sajuthi[2], J. Divers[2], Y. Huang[3], U. Nayak[3], W.M. Chen[3], K.J. Hunt[1], D.L. Kamen[1], G.S. Gilkeson[1], J.K. Frenandes[1], I.J. Spruill[1], C.D. Langefeld[2], W.T. Garvey[4], M.M. Sale[3].* 1) Medical University of South Carolina, Charleston, SC; 2) Wake Forest School of Medicine, Winston-Salem, NC; 3) University of Virginia, Charlottesville, VA; 4) University of Alabama, Birmingham, AL.

The Gullahs form a unique population of African ancestry in the U.S. In addition to their relative genetic and environmental homogeneity and low European admixture (less than 11% on average) a shorter genetic distance between the Gullah and Sierra Leonean tribes has also been reported, suggesting that population genetic signals, such as regions under recent selection, may be more easily detected in the Gullah than in other African American (AA) populations. Given the increased prevalence of some complex diseases in AA and the increasing evidence of selection at loci associated with human diseases, identification of alleles under selection may provide insight into disease susceptibility. Since population-specific selection may cause allele frequency differences, the goal of this study was to identify regions with minor allele frequency (MAF) differences between Gullah and Sierra Leoneans. We had available 277 Gullah and 400 Sierra Leonean samples genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0. After stringent QC, 679,513 SNPs with MAF>5% were used to compute the significance of the MAF differences between the two populations. The European component was adjusted for via inclusion of the HapMap CEU principal component in the logistic regression model. In order to exclude spurious MAF differences, only regions where at least two SNPs in LD showed suggestive (P<10-05) MAF differences were considered. The region showing the most significant MAF differences between the Gullah and Sierra Leoneans was at 20q13.13, where multiple SNPs in LD showed higher MAFs in the Gullah and P-values between E-03 and E-07. As shown in the Haplotter and HGDP Selection Browsers, this region shows evidence for selection (|iHS|>2.5) in the YRI HapMap data, as well as in Africans in the HGDP data. Interestingly, the highest iHS and XP-EHH test results in African populations were identified in the Mandenka, who are geographically the closest to Sierra Leone. In summary, we have identified a ~500 kb region at 20q13.13 with significant allele frequency differences between the Gullah and Sierra Leoneans, suggesting that population-specific selective pressures may be operating at this locus. Given the increased prevalence of several complex traits in AA and the homogeneity of the Gullah, identification of these regions in the Gullahs has the potential to elucidate complex disease risks in AA.

2030W

Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *B. Su[1], K. Xiang[1,6], NA. Ouzhuluobu[2], Y. Peng[1], Z. Yang[1,6], X. Zhang[1,6], C. Cui[2], H. Zhang[1], M. Li[1,6], Y. Zhang[1], NA. Bianba[2], NA. Gonggalanzi[2], NA. Basang[4], NA. Ciwangsangbu[4], T. Wu[3], H. Chen[5], H. Shi[1], X. Qi[1].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Inst Zoology, Chinese Academy Sci, Kunming, Yunnan, China; 2) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, China; 3) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining, China; 4) People's Hospital of Dangxiong County, Dangxiong, China; 5) Department of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, USA; 6) University of Chinese Academy of Sciences, Beijing, China.

Tibetans are well adapted to high-altitude hypoxic conditions, and in recent genome-wide scans, many candidate genes have been reported involved in the physiological response to hypoxic conditions. However, the limited sequence variations analyzed in previous studies would not be sufficient to identify causal mutations. Here we conducted re-sequencing of the entire genomic region (59.4 kb) of the hypoxic gene EGLN1 (one of the top candidates from the genome-wide scans) in Tibetans, and identified 185 sequence variations including 13 novel variations (12 substitutions and one in-del). There is a non-synonymous mutation (rs186996510, D4E) showing surprisingly deep divergence between Tibetans and lowlander populations (FST = 0.709 between Tibetans and Han Chinese). It is highly prevalent in Tibetans (70.9% on average), but extremely rare in Han Chinese, Japanese, Europeans and Africans (0.56%-2.27%), suggesting that it might be the causal mutation of EGLN1 contributing to high-altitude hypoxic adaptation. Neutrality test confirmed the signal of Darwinian positive selection on EGLN1 in Tibetans. Haplotype network analysis revealed a Tibetan-specific haplotype, which is absent in other world populations. The estimated selective intensity (0.029 for the C allele of rs186996510) puts EGLN1 among the known genes undergone the strongest selection in human populations, and the onset of selection was estimated to have started at the early Neolithic (~8,400 years ago). Finally, we detected a significant association between rs186996510 and hemoglobin levels in Tibetans, suggesting that EGLN1 contributes to the adaptively low hemoglobin level of Tibetans compared to acclimatized lowlanders at high altitude.

2031T

Fluctuating And Geographically Specific Selection Characterize Rapid Evolution Of The Human Killer Immunoglobulin-Like Receptor (KIR) Locus. *J.A. Hollenbach[1], P.J. Norman[2], J.I. Rotter[3], E.A. Trachtenberg[1], P. Parham[2], K.D. Taylor[3].* 1) Center for Genetics, Children's Hospital Oakland Research Institute, Oakland, CA; 2) Departments of Structural Biology and Microbiology & Immunology, Stanford University School of Medicine, Stanford, CA; 3) Medical Genetics Institute, Cedars Sinai Medical Center, Los Angeles, CA.

Killer cell immunoglobulin-like receptors (KIR) are expressed on human natural killer (NK) cells and a small percentage of cytotoxic T-cells, and regulate immune responses during infection. Diversity of KIR is known to impact NK and T cell function and has been associated with infectious and autoimmune diseases, as well as cancer and transplant outcome. The KIR gene complex of human chromosome 19q13.4 displays extensive allelic and gene-content variability, with between 7-14 highly polymorphic genes. To examine KIR evolution in the context of the entire human genome, gene-content diversity and 125 SNPs in the KIR and flanking regions were compared to >650,000 genome-wide SNPs in 852 individuals from 52 populations of the human genome diversity panel (HGDP). KIR allelic diversity was further examined in a subset of 60 individuals. In both African and Oceanic populations, SNP diversity and linkage disequilibrium (LD) patterns in the telomeric segment of the KIR region show strong evidence for purifying selection in response to local pressures. Prior to emergence of modern humans, KIR3DL1/S1 diverged into three linages that encode high and low expressing inhibitory or activating receptors, respectively. In Africa, ongoing selection specifically favors the highly expressing inhibitory receptors. In contrast, in Oceanic populations there is evidence of strong and recent selection for the activating form of the receptor. In the centromeric segment of the KIR region, high diversity worldwide is consistent with balancing selection, particularly in Oceania. Highlighting the considerable fluctuation of selection pressures, East Asians exhibit ongoing purifying selection on the centromeric KIR that began before divergence of Amerindians; populations in which centromeric KIR diversity subsequently rebounded under balancing selection. In all populations exons 7-9 of KIR3DL3 and KIR3DL2 mark the centromeric and telomeric boundaries of LD within the KIR region. Contrasting with the instability of the centromeric and telomeric KIR regions is a segment of strong LD that extends from 40kb 5' of the KIR locus to intron 5 of KIR3DL3 that remains relatively conserved worldwide. We show that this segment is likely maintained by epistatic selection to maintain co-inheritance of functionally dependent polymorphisms. In conclusion, we find strong evidence of fluctuating and geographically specific selection in the KIR region that is independent of demographic history.

2032F

Inference of Natural Selection and Demographic History for African Pygmy Hunter-Gatherers. *P.H. Hsieh[1], K.R. Veeramah[2], J. Lachance[4], S.A. Tishkoff[4], J.D. Wall[5], M.F. Hammer[1,2], R.N. Gutenkunst[1,3].* 1) Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ; 2) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, AZ; 3) Department of Molecular and Cellular Biology, University of Arizona, AZ; 4) Department of Biology and Genetics, University of Pennsylvania, Philadelphia, PA; 5) Institute for Human Genetics, University of California, San Francisco, CA.

African Pygmies are hunter-gatherers primarily inhabiting the Central African rainforests, where they are exposed to high temperatures, high humidity, and a pathogen and parasite-enriched woody habitat. These factors undoubtedly influenced their evolutionary history as they adapted to this environment. Many Pygmy populations have historically been in socio-economic contact with neighboring Niger-Kordofanian speaking farmer populations, particularly since the agriculture expansion in sub-Saharan Africa that began five thousand years ago (kya). To look for the true signatures of adaptation to the rainforest habitat of pygmies we must control for this complex demographic history. We sequenced and combined 40x whole genome sequence data from 3 Baka pygmies from Cameroon, 4 Biaka pygmies from the Central African Republic, and 9 Niger-Kordofanian speaking Yoruba farmers from Nigeria. We used ∂a∂i, a model-based demographic inference tool, to infer the history of these populations. Our best-fit model suggests that the ancestors of the farmer and pygmy populations diverged 150 kya and remained isolated from each other until 40 kya. This divergence is more ancient than estimated by previous studies that included fewer loci, but is consistent with a PSMC analysis, a separate inference tool that uses different aspects of the genomic data than ∂a∂i. Interestingly, our analysis shows that models with bi-directional asymmetric gene flow between farmers and pygmies are statistically better supported than previously suggested models with a single wave of uni-directional migration from farmers to pygmies. To identify possible targets of positive selection, we conducted a genomic scan using complementary methods, including the frequency-spectrum based G2D test, the population differentiation based XP-CLR test, and the haplotype based iHS test. We performed 10,000 simulations based on the above best-fit demographic model in order to assign statistical significance to each reported target of natural selection. Our results reveal that genes involved in cell adhesion, cellular signaling, olfactory perception, and immunity were likely targeted by natural selection in the pygmies or their recent ancestors. Our analysis also shows that genes involved in the function of lipid binding are enriched in highly differentiated non-synonymous mutations, suggesting that this function may have acted differently on the Pygmies and farmers after their divergence from their common ancestor.

2033W

Forward simulations of recurrent selection and demographics with rescaled parameters. *L.H. Uricchio[1], R.D. Hernandez[2].* 1) Bioengineering Graduate Group, UC Berkeley & UCSF, San Francisco, CA; 2) Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA.

It is well known that positive selection impacts patterns of diversity in linked regions of a genome. When a population experiences frequent positive selection, the resulting patterns of genetic variation in linked regions can be quite skewed, but theoretical treatments of this process (often referred to as 'recurrent hitchhiking') are mostly limited to simple models. Forward simulation can provide insights into more sophisticated models that include arbitrary demographic scenarios, rates of selection, and distributions of selection coefficients, but the computational burden of forward simulation is often prohibitive. A remedy to this computational challenge is to rescale the relevant parameters (e.g., population size) in a way that conserves the underlying dynamics. However, *ad hoc* approaches to parameter scaling in recurrent hitchhiking may not always provide sufficiently accurate dynamics, potentially skewing patterns of diversity in simulated DNA sequences. Here, we perform a detailed theoretical analysis of the recurrent hitchhiking model that relaxes some simplifying assumptions, and present a simple method for parameter rescaling under the model. We thoroughly test the robustness of rescaling across the parameter space, and describe the conditions under which rescaling provides accurate results. We find that our approach enables us to perform large-scale simulations in a fraction of the computational time. We apply this rescaling method to simulations of interference among selected sites as well as demographic models of growth and contraction.

2034T

Variation in Bone Mineral Density (BMD) in Children of Different Ethnic Backgrounds is Explained By Genetic Profiling: The Generation R Study. *M. Medina Gomez[1,2,3], D.H.M. Heppe[2,3,4], K. Estrada[1,2,5,6,7], L. Oei[1,2], A. Hofman[1,2,3], M. Kayser[8], A.G. Uitterlinden[1,2,3], V.W. Jaddoe[2,3,4], O. Lao[8], F. Rivadeneira[1,2,3].* 1) Internal Medicine, Erasmus MC University, Rotterdam, Netherlands; 2) Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 3) The Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands; 4) Pediatrics, Erasmus Medical Center, Rotterdam, The Netherlands; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 7) Harvard Medical School, Boston, MA. USA; 8) Forensic Molecular Biology, Erasmus Medical Center, Rotterdam, The Netherlands.

Aim: To study genetically-determined differences in BMD in children across ethnicities, we constructed a genetic score (GS) of BMD SNPs in children from multiple ethnic background and tested for association with total-body (TB-) BMD. **Methods:** We included 4,009 children from The Generation R study, a single-center prospective multiethnic birth cohort in Rotterdam, The Netherlands, with GWAS data and TB-BMD measured by DXA at a mean age of 6.2 years. The admixed population includes a wide spectrum of ethnic origins including blends between continental groups. Clustering based on genetic data (using ADMIXTURE) was employed assigning children to one of three transcontinental ancestral groups: predominantly, Caucasian (n=3,513), Asian (n=159) and African (n=337). Differences in TB-BMD were assessed by least-squares means using the Caucasian population as reference and adjusting for age, sex, height, fat and lean mass. The GS consisted of the sum of BMD-increasing alleles (0,1,2) in 63 genetic variants (from 56 BMD loci identified in adults) and was analyzed across quintiles. **Results:** Children of African descent had higher TB-BMD (0.585 g/cm2) than those of Asian (0.559 g/cm2; P=2E-6) and Caucasian descent (0.552 g/cm2; P=1E-6). The GS explained 5.2% of the variance in BMD and 2.6% after principal component correction. As compared to children in the GS middle quintile (54% of the population; n=2,179), individuals in the GS highest quintile (2.6% of the population; n=106) had 0.69 SDs higher BMD (P<1E-6), while those in the GS lowest quintile (1.9% of the population; n=77) had 0.53 SDs (P=4E-6) lower BMD. The number of children of African descent were overly represented (P=3E-104) in the two highest quintiles of the GS with 65.6%, as compared to 18.8% in children of Caucasian and 15.7% in those of Asian descent. Positive differences of >10% in BMD-increasing allele frequencies were seen between children of African and Caucasian descent for 21 variants and differences of > 5% for 34 variants of the score. Data from the Human Genome Diversity panel also indicated higher frequency of BMD-increasing alleles in African populations, with South-west Bantuans displaying the highest frequencies. **Conclusions:** Genetic ancestry of African descent is correlated with a higher frequency of BMD-increasing alleles. Ethnic differences in bone accrual are already evident at childhood and are partially explained by allele frequency differences in variants from known BMD loci.

2035F

Non-Random Geospatial Distributions of HLA-Haplotypes in the United States. *N.P. Leahy[1], A. Chatterjee[2], M. Albrecht[1], M. Maiers[1].* 1) Bioinformatics Research, Be The Match, Minneapolis, MN; 2) School of Statistics, University of Minnesota, Minneapolis, MN.

A key predictor of hematopoietic cell transplantation success is using a donor with a fully HLA-matched donor for the patient. Donor registries such as the Be The Match® require knowledge of human population structure to target recruitment to subpopulations with rare variants. We used data from the Be The Match® registry to test for spatial heterogeneity within self-identified race and ethnic groups (SIRE). In total there were four broad SIRE groups (African-American, Asian/Pacific Islander, Caucasian, Hispanic) and 18 subcategories of the broad SIRE groups. The high genetic variability at HLA and low haplotype frequencies presented challenges for traditional analysis methods. Therefore, we opted for a permutation methodology to establish heterogeneity. For each of the SIRE groups, we generated a matrix, ***H***, of the occurrences of the top 100 haplotypes in 50 US states, District of Columbia, and Puerto Rico. Our metric was the maximum eigenvalue for the variance-covariance matrix, λ_{max}. To test for spatial heterogeneity in our data, we performed 10,000 permutations of the data to generate a probability distribution of λ_{max}. Significance was determined by the position of the metric to the distribution. This was repeated for the top 1000 haplotypes for 1000 permutations. The permutation analysis found for the 100 most common haplotypes, Asian/Pacific Islander ($p < 0.0001$) and Hispanic ($p = 0.0036$) had significant special heterogeneity and Caucasian was marginally insignificant. When the analysis was repeated with the 1000 most common haplotypes, Caucasian was not significant ($p < 0.0001$) and Hispanic was no longer significant ($p = 0.412$). Outcomes of the subcategories mostly reflected those of the broad SIRE groups. African-American was not significant. These results were consistent with historical founder events. That not all SIREs can be treated as homogenous subpopulations has implications for recruitment strategies of donor registries. To better serve patients, it is necessary to resolve the patterns within spatially heterogeneous SIRE groups to maximize diversity within registries. That increasing the number of haplotypes shifted some SIRE groups from significant to non-significant is problematic. The expectation was adding additional genetic information should have provided better resolution of population structure, requiring a more nuanced hypothesis.

2036W

Proportion of African ancestry in *Helicobacter pylori* is associated with increased severity of gastric lesions in human hosts with high Amerindian ancestry. *N. Kodaman[1], A. Pazos[2,8], B.G. Schneider[2], R. Sobota[1], M.B. Piazuelo[2], C.L. Shaffer[3], J. Romero-Gallo[2], T. de Sablet[2,7], L.E. Bravo[4], K. Wilson[2,5,7], T. Cover[3,6,7], S.M. Williams[9], P. Correa[2].* 1) CHGR, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Pathology, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Pathology, Universidad del Valle School of Medicine, Cali, Colombia; 5) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN; 6) Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN; 7) Veterans Affairs Tennessee Valley Healthcare System, Nashville, TN; 8) Universidad de Nariño, Pasto, Colombia; 9) Department of Genetics, Dartmouth University, Hanover, NH.

Helicobacter pylori is the principal cause of gastric cancer, the second leading cause of cancer mortality worldwide. In certain Colombian populations, over 90% of individuals are infected with *H. pylori*, but infection rate does not generally predict cancer prevalence. In particular, residents of the Andean mountain region are 25 times more likely to develop gastric cancer than their coastal counterparts, despite similar rates of infection. We determined the ancestry of *H. pylori* isolates from the gastric biopsies of 275 Colombian subjects from both the mountain and coastal regions. Most isolates contained genomic regions from four ancestral *H. pylori* populations: Africa1 (AA1), Europe1 (AE1), Europe2 (AE2), and East Asia (AEA), but these proportions varied with geography. The AA1 cluster was more common in coastal samples (mean=47.9%), and AE2 in mountain samples (mean= 50.7%). The human ancestry of the biopsied individuals also varied with geography, with mean proportions of 57.9% African, 22.6% Amerindian, and 19.5% European ancestry in the coastal region, and 67% Amerindian and 30.4% European ancestry in the mountain region. All pairwise correlations between *H. pylori* ancestry and human host ancestry were significant. While African *H. pylori* ancestry correlated negatively with Amerindian human ancestry ($r=-0.60$), the interaction between the two had a strongly deleterious effect on histopathology. In a multivariate linear model, the marginal effect of Amerindian ancestry on histopathology became non-significant ($p=0.63$) when the interaction was included, indicating that Amerindian ancestry *per se* was not responsible for the increased severity of the lesions. Importantly, region of origin (coast or mountain) was also not significant when added as a covariate to the multivariate model ($p=0.89$). Our findings indicate that AA1 ancestry is relatively benign in humans of African ancestry, but potentially deleterious in Amerindians, making the interaction between human and *H. pylori* ancestry a potential confounder of univariate analyses on histopathology. Co-evolution has likely modulated disease risk, and the disruption of this relationship may account for the risk discrepancy in Colombian populations.

2037T

Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia. *D. Lu, S. Xu.* Partner institute for Computational Biology, Shanghai, China.

The 1000 Genomes Project (1KG) aims to provide a comprehensive resource on human genetic variations. With an effort of sequencing 2,500 individuals, 1KG is expected to cover the majority of the human genetic diversities worldwide. In this study, using analysis of population structure based on genome-wide single nucleotide polymorphisms (SNPs) data, we examined and evaluated the coverage of genetic diversity of 1KG samples with the available genome-wide SNP data of 3,831 individuals representing 140 population samples worldwide. We developed a method to quantitatively measure and evaluate the genetic diversity revealed by population structure analysis. Our results showed that the 1KG does not have sufficient coverage of the human genetic diversity in Asia, especially in Southeast Asia. We suggested a good coverage of Southeast Asian populations be considered in 1KG or a regional effort be initialized to provide a more comprehensive characterization of the human genetic diversity in Asia, which is important for both evolutionary and medical studies in the future.

2038F

The Iranian Genomes Project. *R. Daneshjou[1], M. Ronaghi[2], C.D. Bustamante[1], P.C. Sabeti[3], R.B. Altman[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Illumina Inc., San Diego, CA; 3) Department of Human Evolutionary Biology, Harvard University, Cambridge, MA.

While large-scale full-genome studies, such as the 1000 Genomes Project, have examined genetic variation for many of the world's populations, Middle Eastern populations have remained little studied. This group's unique history and geographic location, at the presumed bottleneck of the out of Africa migration, make them an important population to investigate. Here, we present the Iranian Genome Project, an effort to deeply sequence (at 30x coverage) over 50 individuals of Iranian descent. Our cohort includes conserved and understudied sub-populations from Iran, such as Zoroastrian, Bakhtiari, and Jewish populations. We explore the relationship between the Iranian population to other 1000 Genomes populations and the relationships between the Iranian sub-populations using EIGENSOFT software and the HAPMIX algorithm to compute FST and Principle Components Analysis (PCA). We further identify variants that are unique to the Iranian population and predict the functional impact of these variants. Because ancestry is an important consideration in clinical genetics, we identify the frequency of known or predicted damaging mutations in genes identified by the recent American College of Medical Genetics and Genomics (ACMG) guidelines on incidental findings in genome sequencing. Additionally, we calculate frequencies for important known clinical pharmacogenetic variants and identify new variants in known pharmacogenetic genes, which may play an important role in this population. Given the importance of diverse reference sequences for both population and clinical genetics, the Iranian Genome Project serves to fill in the gap by providing deeply sequenced genomes from a key but unexplored Middle Eastern population.

2039W

Morphometric and ancient DNA study of human skeletal remanants in Indian Subcontinent. *N. Rai[1], M. Mirazon Lahr[2], L. Singh[1], K. Thangaraj[1].* 1) Evolutionary and Medical Genetics, Centre for Cellular and Molecular Biology, Hyderabad, India; 2) Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, U.K.

Recovery and sequencing of mtDNA from ancient human remnants is a daunting task but provides valuable information about human migrations and evolution. Our present study is the first to recover, amplify and sequence (HVR and coding regions of mtDNA) inadequately preserved and highly degraded (1.5 Ky to ≤1.0 Ky ago) hominids mitochondrial DNA of three most intriguing and indigenous ancient population of South and South-East Asia (Myanmar=20 Buried individuals, Nicobar Islands=15 and Andaman Island=6). Following all parameters and to avoid the chance of contamination we independently extracted and sequenced the DNA in two different labs and measured the cranial variability in all hominid skulls using 128 cranial landmarks, compiled 3D morphometrics, genetic data of ancient DNA samples and analyzed the admixture and genetic affinities of above three populations. Results showed the predominant frequency of F1a1 and complete absence of 9bp deletion in ancient Nicobarese. Unlike in previous reports on modern Nicobarese, the high frequency of F1a1 haplogroup in ancient Nicobarese show the probable migration of Nicobarese from South East Asia and the complete absence of 9bp deletion suggests the different events of settlement. This study failed to detect genetic affinities of Burmese with Nicolbarese even though their phenotype and language appears to be same. We first time report any kind of population study on Burmese populations and with the genetic affinity of Burmese with East Asian, East Indian (Including Gadhwal region of Himalaya) and Bangladeshi populations, we found significant admixture with West Eurasians. Our study strongly supports the West Eurasian and East Asian route of migration and settlement of early Burmese population. The three populations in the present study are quite different in their genetic structure but 3D morphometric study using huge number of landmarks explains a close homology among these populations and this can be explained by the role of climatic signature on these populations.

2040T

No significant differences in the accumulation of deleterious mutations across diverse human populations. *R. Do[1,2], D. Balick[1,3], I. Adzhubey[1,3], S. Sunyaev[1,3], D. Reich[1,4].* 1) Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142; 2) Center for Human Genetics Research, Massachusetts General Hospital, Boston, Massachusetts, 02114; 3) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115; 4) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115.

Differences in demographic history across populations are expected to cause differences in the accumulation of deleterious mutations because natural selection works less efficiently when population sizes are small. Surprisingly, however, the relative burden of deleterious mutations has never been directly measured across human populations on a per-haploid genome basis, despite the fact that this is what matters biologically in the absence of dominance and epistasis. Here we empirically measure the relative accumulation of deleterious mutations in 13 diverse populations (Yoruba, Mandenka, San, Mbuti, Dinka, Australian, French, Sardinian, Han, Dai, Mixe, Karitiana and Papuan) along with one archaic population (Denisova). All the present-day populations have statistically indistinguishable accumulations of coding mutations. We highlight two examples. First, we find no evidence for a lower mutational load in West Africans than in Europeans despite the approximately 30% higher genetic diversity in West Africans: the accumulation of nonsynonymous mutations in West Africans is 1.01±0.02 times that in Europeans, and for 'probably damaging' mutations, the ratio is 1.03±0.04. Second, we find no evidence for a lower mutational load in populations that have experienced agriculture-related expansions over the last 10,000 years and those that have not: the ratio in Chinese to Karitiana hunter gatherers from Brazil is 0.99±0.07. We determined that these null results are not an artifact of insensitivity of our method to differences in demographic history. As a positive control, we also analyzed archaic Denisovans who are known to have had a small population size for hundreds of thousands of years since separation from modern humans. We show that the Denisovan lineage has accumulated 'probably damaging' mutations 1.33±0.06 times more rapidly than modern humans since they split. These analyses are important because of the new constraints they place on the distribution of selection coefficients in humans. Given the currently estimated demographic histories of West Africans and Europeans, combined with the fact that we do not detect a lower accumulation of deleterious mutations in West Africans than Europeans, we can conclude that only a small proportion of nonsynonymous mutations have selection coefficients in the range s=-0.01 to -0.001, which is the range of selection coefficients which would be expected to show a lower accumulation in West Africans than in Africans.

2041F

Y chromosomes of ancient Hunnu people and its implication on the phylogeny of East Asian linguistic families. *LL. Kang[1,2], TB. Jin[1], F. Wu[1,2], X. Ao[2], SQ. Wen[2], CC. Wang[2], YZ. Huang[2], XL. Li[1,2], H. Li[1,2].* 1) Key Laboratory of High Altitude Environment and Gene Related to Disease of Tibet Ministry of Education,School of Medicine, Tibet University for Nationalities, Xianyang, Shaanxi, China; 2) MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China.

The Hunnu (Xiongnu) people, also called Huns in Europe, were the largest ethnic group to the north of Han Chinese until the 5th century. The ethnolinguistic affiliation of the Hunnu is controversial among Yeniseian, Altaic, Uralic, and Indo-European. Ancient DNA analyses on the remains of the Hunnu people had shown some clues to this problem. Y chromosome haplogroups of Hunnu remains included Q-M242, N-Tat, C-M130, and R1a1. Recently, we analyzed three samples of Hunnu from Barköl, Xinjiang, China, and determined Q-M3 haplogroup. Therefore, most Y chromosomes of the Hunnu samples examined by multiple studies are belonging to the Q haplogroup. Q-M3 is mostly found in Yeniseian and American Indian peoples, suggesting that Hunnu should be in the Yeniseian family. The Y chromosome diversity is well associated with linguistic families in East Asia. According to the similarity in the Y chromosome profiles, there are four pairs of congenetic families, i.e., Austronesian and Tai-Kadai, Mon-Khmer and Hmong-Mien, Sino-Tibetan and Uralic, Yeniseian and Palaesiberian. Between 4,000-2,000 years before present, Tai-Kadai, Hmong-Mien, Sino-Tibetan, and Yeniseian languages transformed into toned analytic languages, becoming quite different from the rest four. Since Hunnu was in the Yeniseian family, all these four toned families were distributed in the inland of China during the transformations. There must be some social or biological factors induced the transformations at that time, which is worth doing more linguistic and genetic researches.

2042W

Admixture in the Pre-Columbian Caribbean. *J.C. Martinez-Cruzado[1], E.P. Tascón-Peñaranda[1], F. Curbelo-Canabal[1], T. Porrata-Doria[1], C. Eng[2], E.G. Burchard[2].* 1) Department of Biology, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico; 2) Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, California.

The biological origin of the Caribbean aborigines that greeted Columbus is one of the most controversial issues regarding the population history of this region. Genome studies suggest an Equatorial-Tucanoan origin, consistent with the Arawakan language spoken by most natives of the region. However, the archaeological evidence suggests an early arrival from Mesoamerica, and their admixture with the more recent Arawak-speaking group stemming from the Amazon remains a possibility. The lineages comprehending most Puerto Rican samples belonging to haplogroups B1 and C1, which in turn encompass 44% of all Native American mtDNAs in the island, have an unambiguous South American origin. However, none of those belonging to haplogroup A2, encompassing 52% of all Native American mtDNAs, have been related to South America or any other continental region. To augment the scarce data from Mesoamerican countries other than Mexico, we present the complete mtDNA sequence of 6 Honduran samples belonging to distinct control region lineages in addition to 3 from the Dominican Republic and 3 from Puerto Rico. Interestingly, maximum likelihood phylogenetic reconstruction including 40 published haplogroup A2 sequence haplotypes from Mesoamerica, Central America and South America clusters 8 out of 10 Mesoamerican and Andean haplotypes in a deep rooted group, separate from, and excluding all Costa Rican, Panamian and Brasilian haplotypes, suggesting a relatively recent origin for Chibchan-Paezan and Amazonian groups. Furthermore, 4 of the 5 Greater Antillean A2 haplotypes are included in the deeply rooted Mesoamerican-Andean cluster. Moreover, the only Cuban haplotype in the literature and the remaining A2 haplotype from the Dominican Republic form even more deeply rooted private branches. Similarly, the only haplogroup C1d sample sequenced from the Dominican Republic forms a private branch with the deepest root in a maximum likelihood tree containing 19 additional C1d haplotypes from Mexico to Brasil plus the CRS. In conclusion, our preliminary results suggest that a substantial proportion of the Native American mtDNA lineages from the Greater Antilles do not share an Amazonian origin with the language their people spoke in 1492. Furthermore, the position of two Dominican lineages at the earliest split in both their respective trees suggests an early origin that could be explained by extensive lineage extinctions in Mesoamerica and the Andes or an origin in North America.

2043T

Resquencing of Australian Aboriginal mtDNA and Y chromosomes. *Y. Xue[1], M. Cerezo Fernandez[1], Y. Chen[1], S. McCarthy[1], MO. Pollard[1], Q. Ayub[1], N. Nagle[2], P. McAllister[3], RJ. Mitchell[2], C. Tyler-Smith[1].* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Genetics, La Trobe University, Melbourne, Australia; 3) Griffith University, Gold Coast, Queensland, Australia.

Modern humans originated in Africa and spread across the rest of the globe 50-70 thousand years ago. The first identified divergence outside Africa was between the ancestors of the Australian Aborigines and some nearby populations on the one hand, and the ancestors of Asians, Europeans and other non-Africans on the other hand. The genetic characterization of this divergence and subsequent events in Australian Aboriginal history before the colonial era remain poorly described. Seven Australian Aboriginal males requested sequencing of their mitochondrial DNA (mtDNA) and Y chromosomes, and we generated high depth Illumina 100 bp paired-end sequence data. The average coverage is ~15-20x for Y chromosome and ~3000x for mtDNA. The mtDNAs of the samples belong to new branches of haplogroups P (5 individuals), M (1 individual) and O (1 individual). These haplogroups are known from Australia, but the many new variants illustrate their diversity and distinct origins. The Y chromosomes belong to haplogroups C (3 individuals), K* (3 individuals), both previously known from Australia, and M (1 individual). This last haplogroup has a restricted geographical distribution centred on Papua New Guinea, and illustrates a genetic link between Australia and that region. The ~3,000 new Y-SNPs present in these samples are permitting refined estimates of the ancient coalescence times between Australian lineages and those in the rest of the world, and re-examination of the hypothesis of more recent links between Australia and South Asia, originally based on Y-chromosomal similarities.

2044F

The possible role of social selection in the distribution of the "Proto-Mongolian" haplotype in Kazakhs, Kyrgyz, Mongols and other Eurasian populations. *M. Zhabagin[1,2], P. Tarlykov[3], Zh. Sabitov[4], H. Dibirova[5], A. Bogunova[6], I. Tazhigulova[7], S. Frolova[5], Zh. Isakova[8], A. Nimadava[9], I. Zakharov[2], O. Balanovsky[2,5].* 1) Center for Life Sciences, Nazarbayev University (Astana, Kazakhstan); 2) Vavilov Institute of General Genetics RAS (Moscow, Russia); 3) Research Centre of Medical Genetics RAMS (Moscow, Russia); 4) Gumilyov Eurasian National University (Astana, Kazakhstan); 5) The Komsomolsk-on-Amur State University (Komsomolsk-on-Amur, Russia); 6) National Center for Biotechnology (Astana, Kazakhstan); 7) Forensic science centre of the Ministry of Justice of the Republic of Kazakhstan (Astana, Kazakhstan); 8) Research Institute of Molecular Biology and Medicine (Bishkek, Kyrgyz); 9) Mongolian Academy of Sciences (Ulan Bator, Mongolia).

Social factors may be important contributors to reproductive success and determination of the selective survival of individuals. Therefore, social selection and other social factors are important for understanding population structure and its formation. The role of social selection on the distribution and formation of Y-chromosomal gene pool has been studied. There is a strong connection between social selection and birth rate of the descendants, whose fathers had achieved high social status during the expansion of the Mongol Empire and associated historical events. A total of 783 haplotypes, including 687 newly obtained and 96 retrieved from the literature were assigned to the haplogroup C3*-M217 (xM48) based on genotyping 17 Y-chromosomal STR markers. These haplotypes represent 11 populations of Eurasia: Kazakhs, Mongols, Kyrgyz, Telengits, Circassians, Balkar, Temirgoys, Karachai, Evenki, Kizhi and the Pashtuns. As the result, a major haplotype 13-16-25-15-16-18-14-10-22-11-10-11-13-10-21 (DYS389a-DYS389b-DYS390-DYS456-DYS19-DYS458-DYS437-DYS438-DYS448-GATA4-DYS391-DYS392-DYS393-DYS439-DYS635, N=94) was found to have 12.00% frequency within haplogroup C3*. This haplotype includes and extends the previously described 'star-cluster' haplotype. Noteworthy, the frequency of this major haplotype within haplogroup C3* was 16.80% in Kazakhs, 10.13% in Mongols and 2.63% in Kirgiz who are not considered as direct descendants of Genghis Khan. 35.10% of the major haplotype was represented by Kazakh tribe Ashamayly-Kerey, 17.02% by the Khalkh Mongols and 7.44% by the Barguts. Therefore, we suppose this major ancestral haplotype to be the "proto-Mongolian haplotype", inherited by Genghis Khan and his descendants. It is important to mention that Temujin belongs to Kiyat-Borjigin tribe that in turn is a branch of the bigger Borjigin tribe, part of the Khalkh Mongols. Thus, Genghis Khan might be considered as a carrier rather than founder of the star-cluster haplotype. He and his descendants are the ones who contributed to a positive effect of social selection in the distribution of this haplotype. Other examples are the Barguts, who had Genghis Khan's credit and were granted with a number of privileges, or the Kerey, based on the fact that Temujin had been brought up at the court of the Togrul Khan, belonging to the Kerey tribe.

2045W

Y-chromosome diversity in Mayan, Ch'ol, and Chontal populations from Campeche and Tabasco. *A. Quinto[2], M.A. Meraz[2], R. Camacho[1], T. Schurr[3], M. Vilar[3], G. Noris[4], C. Santana[4], J.B. Gaieski[3], A.C. Owings[3], R. Gómez[1].* 1) Dep. de Toxicología, Cinvestav-IPN, Mexico City, Mexico; 2) Dep. de Biomedicina Molecular, Cinvestav-IPN, Mexico City, Mexico; 3) Dept. Anthropology, Univ. Pennsylvania, Philadelphia, PA; 4) Laboratorio BIMODY, Querétaro, Qro., Mexico.

Mayan populations are the descendants of a Mesoamerican civilization that survived the Classic period collapse, the arrival of Spanish conquerors, and the European colonization of the Americas. Today, Mayan speakers (≈1.3 million inhabitants) present one of the two largest indigenous populations in Mexico, the other being the Náhua (≈1.6 millions inhabitants). The Mayan language family includes fourteen languages that are spoken in seven states of Mexico (Chiapas, Tabasco, Yucatán, Quintana Roo, Campeche, San Luis Potosí, and Veracruz). Although archeologically well understood, this complex cultural and linguistic region remains genetically understudied. Therefore, to assess the genetic diversity in Mayan populations and reconstruct historic patterns of interactions among Mesoamerican groups, we genotyped 112 males from 16 communities in the states of Tabasco (Ch'ol and Chontal) and Campeche (Mayans). The pattern of Y-chromosome haplogroup diversity in these populations was analyzed using published SNP and Y-STR genotyping methods, and the results compared with data from others Native American populations. Our Y-chromosome analysis revealed high haplotypic diversity in the study populations, with Mayans being the most diverse (0.9905 ± 0.0083), followed by Ch'ol (0.9316 ± 0.0386) and Chontal (0.9291 ± 0.0230). Haplogroup Q-M3 was the most frequent paternal lineage in all groups, suggesting a common ancestral origin for them. However Mayan populations also had significant European (~32%) admixture represented by haplogroups R1a, R1b, E1b1b, and T. In addition, Chontal and Ch'ol populations showed some admixture with Europeans with the presence of haplogroups E1b1b (11%) and J2 (5%), respectively. Our results suggest the Mayans of Campeche have experienced greater European admixture, whereas Chontal and Ch'ol populations have a much higher frequency of Amerindian Y-chromosomes.

2046T

Juxtapositions of short IBD blocks can cause biased estimation in inferences based on the length of IBD blocks.
C.W.K. Chiang[1], *J. Novembre*[2]. 1) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA; 2) Human Genetics, University of Chicago, Chicago, IL.

Blocks of identity-by-descent (IBD) play an important role in many modern genetic applications, including long-range phasing, imputation, genetic mapping, detection of natural selection, and demographic inferences. One commonly used definition of IBD blocks is that they are contiguous segments of the genome inherited from a recent shared common ancestor without intervening recombination. With programs like Beagle's fastibd, long IBD blocks (>1cM) can be efficiently detected using high-density SNP array data of a population sample. However, all programs detect IBD based on contiguous segments of identity-by-state (IBS). As such, detected IBD blocks could often be due to the juxtaposition of smaller IBD blocks inherited from different common ancestors. Here, we show the juxtaposition of small IBD blocks leads to an error in estimating the length distribution of IBD blocks and can affect downstream inferences. To demonstrate the prevalence of subsegment juxtapositions, we used coalescent simulations where we know the precise genealogy of the sample and found that >35% of the detected IBD segments of 1cM or longer are composed of at least two subsegments. In particular, 11% of the detectable segments consist of at least 1 other subsegment >25% of the total length, and this effect was more pronounced for detectable segments between 1 to 2cM long, compared to segments >2cM long. To demonstrate that the juxtaposition can lead to practical problems, we investigated the impact on a novel estimator of the de novo mutation rate using IBD blocks. We observed accurate estimates of the input mutation rate when true IBD blocks are used, but overestimates of the mutation rate by ~15 fold using inferred IBD blocks. When the effect of juxtaposition on the estimated age of the block was modeled, the mutation rate estimate improved greatly. Our results suggest that identifying IBD blocks based on extended IBS can inflate the length of IBD blocks, and in this case results in an inflated estimate of the de novo mutation rate, unless properly accounted for. This effect should be carefully considered as methods to detect shorter IBD blocks using sequencing data are being developed.

2047F

Improved detection of ancient hominin admixture in modern humans.
S.R. Browning[1], *B. Vernot*[2], *B.L. Browning*[1,2,3], *J.M. Akey*[2]. 1) Biostatistics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA; 3) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA.

The S* statistic [1] detects genomic segments with variants in unusually strong linkage disequilibrium (LD) indicative of ancient admixture with now-extinct hominins such as Neandertals. The detected introgressed haplotypes can be used to reconstruct archaic genomes which will facilitate the inference of important population genetics characteristics such as the effective population sizes of archaic populations, timing and parameters of introgression, and phylogenetic relationships of archaic groups.

S* looks for near complete LD between a subset of SNPs in a region, with penalties incurred when small numbers of haplotypes enter or leave the introgressed state [1]. Recombination events in humans since admixture are inherited which results in groups of haplotypes entering or leaving the introgressed state together. Instead of counting haplotypes, we count haplotype clusters in the Beagle model [2]. The haplotype cluster approach is particularly advantageous when the number of individuals being analyzed is large, because a large number of haplotypes entering/leaving the introgressed state may reflect a single recombination event that is represented by one haplotype cluster. A single recombination event should be penalized lightly, whereas multiple recombination events should be penalized heavily in the statistic.

To evaluate our method, we simulated sequence data using the model of [3]. In the simulated data, we find that haplotype clusters correlate closely with introgression status. Each haplotype cluster typically represents only introgressed or only non-introgressed haplotypes, even when haplotype phase information is not known a priori. In simulated data with 100 African reference samples and 500 European samples, we find 20% more introgressed regions using our haplotype clustering approach than with the standard S* approach, for the same false discovery rate. Incorporation of haplotype clustering into an S*-type statistic is useful for exploiting the increasing numbers of sequenced human genomes to find introgressed haplotypes.

[1] Plagnol, V. and J.D. Wall, PLoS Genetics, 2006. 2:e105. [2] Browning, S.R., Am J Hum Genet, 2006. 78:903-13. [3] Fu, W., et al.,Nature, 2013. 493:216-20.

2048W

The theoretical accuracy of deterministic approximations to coalescent formulas.
E.M. Jewett, N.A. Rosenberg. Biology, Stanford University, Stanford, CA.

Many coalescent distributions and expectations can be derived by conditioning on the number, n_t, of alleles at time t in the past that are ancestral to a data set of n_0 alleles sampled in the present. However, summing over the conditional distribution of n_t, given n_0, can be computationally challenging when n_0 is large. Thus, such formulas can be difficult to evaluate on modern genomic datasets with hundreds or thousands of sampled lineages. One alternative to conditioning on all possible values of n_t is to use an approximation in which n_t is assumed to equal its expected value $E[n_t]$ with probability one (Slatkin, 2000). This approximation greatly reduces the number of terms in conditional expressions, significantly reducing their computational complexity. However, despite the utility of the approximation, its theoretical accuracy is not known. Instead, the accuracy of any given version of the approximation must be evaluated empirically by comparing it with the true distribution. As a result, the accuracy of a given approximation cannot be known outside the range of parameter values over which the true distribution can be computed. Here, we show that approximate distributions converge uniformly to the true distributions under certain simple assumptions, and we derive an expression for the asymptotic approximation error. Our results provide a theoretical basis for understanding the ranges of parameter values over which any given approximation is accurate, facilitating the application of the approximation $n_t = E[n_t]$ to reduce the complexity of computing coalescent formulas on large genomic data sets.

2049T

DNA-based detection of Glucose 6-Phosphate Dehydrogenase (G6PD) deficiency alleles in an Eastern Caribbean population.
C. Gupta, L. Deschênes, C. Headland, I. McIntosh. American University of the Caribbean School of Medicine, Cupecoy, St. Maarten.

G6PD deficiency is the most common enzyme deficiency worldwide (~400 million cases) with highest prevalence in populations where malaria is endemic. G6PD is required to produce NADPH and maintain glutathione in the reduced state in red blood cells, and deficiency predisposes to haemolytic anaemia and neonatal jaundice in hemizygous males. The frequencies of deficiency alleles have been reported as 0.05 - 0.25 in African populations but have not been assessed in the Eastern Caribbean. We developed a noninvasive method to detect the major deficiency alleles through extraction of DNA from buccal cells, PCR amplification and identification of the deficiency alleles via restriction endonuclease digestion and gel electrophoresis. In a cohort self-reporting as African or Afro-Caribbean we identified two hemizygous males and two heterozygous females giving an overall G6PD deficiency allele frequency of 0.048. This result is at the lower end of the range reported previously for which we propose two possible explanations: (1) admixture in the population introducing European & Amerindian alleles (of note, the incidence of Hb-S and Hb-C alleles in this cohort was 0.086 c.f. 0.13 in previous studies of this population); (2) mutation-specific analysis would not detect other deficiency alleles which may be more common in this population.

2050F

Insights on the evolutionary history of Tibetans from whole-genome sequence data.
H. Hu[1], T. Simonson[2], G. Glusman[3], J. Roach[3], G. Cavalleri[4], M. Brunkow[3], M. McCormack[4], N. Petousi[5], P. Lorenzo[6], R. Gelinas[3], L. Jorde[7], J. Prchal[6], P. Robbins[5], C. Huff[1]. 1) Department of Epidemiology, UT MD Anderson Cancer Center, Houston, TX; 2) University of California, San Diego La Jolla, Division of Physiology, San Diego, CA; 3) Institute for Systems Biology, Family Genomics Group, Seattle, WA; 4) Royal College of Surgeons in Ireland, Molecular and Cellular Therapeutics, St Stephen's Green, Dublin 2, Ireland; 5) University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford, OX1 3PT, United Kingdom; 6) University of Utah, Hematology, Salt Lake City, UT; 7) University of Utah, Department of Human Genetics, Salt Lake City, UT.

The Tibetan people inhabit the world's highest plateau, with an average altitude exceeding 4,500 meters. Over thousands of years Tibetans have evolved unique adaptations to this extreme environment, such as low oxygen levels. Here we report the first study of Tibetan evolutionary history from whole-genome sequence data. Using high-coverage whole-genome sequence data from 17 Tibetan and 9 Han Chinese individuals, we characterize recent admixture events and ancient demographic history of Tibetans from SNV allele frequency spectra data in the context of the Out of Africa and European/Asian divergence events. We restrict this analysis to genomic regions at least 10 kb away from exons or conserved regions to minimize the confounding effects of natural selection. Our results indicate that the Tibetan population diverged from the Han Chinese population much earlier than previously estimated from exome sequence analysis. We also detect evidence of an ancient admixture event from Han Chinese to Tibetans around 25,000 years ago, as well as substantial subsequent gene flow between Tibetan and Han Chinese populations. We utilize this demographic model to detect genomic signals of recent positive selection using the Composite of Multiple Signals (CMS) test. The CMS test identifies several genes previously implicated in Tibetan high-altitude adaptation, including EGLN1 and EPAS1. Our results provide a high-resolution map of the potential functional targets of recent positive selection in Tibetan genomes.

2051W

Identification of signals of recent selection in the Sea Island Gullah African Americans.
S.P. Sajuthi[1], J. Divers[1], Y. Huang[3], U. Nayak[3], W. Chen[3], K.J. Hunt[2], D.L. Kamen[2], G.S. Gilkeson[2], J.K. Fernandes[2], I.J. Spruill[2], C.D. Langefeld[1], W.T. Garvey[4], M.M. Sale[3], P.S. Ramos[2]. 1) Wake Forest University Health Sciences, Winston-Salem, NC; 2) Medical University of South Carolina, Charleston, SC; 3) University of Virginia, Charlottesville, VA; 4) University of Alabama, Birmingham, AL.

PURPOSE: Many common diseases are more prevalent in specific ethnic groups. Given the growing number of disease-associated loci in regions that show evidence of selection, identification of alleles under selection may provide insight into disease susceptibility. Relative to other African-Americans (AA), the AA Gullah population has lower European admixture and higher ancestral homogeneity from the Sierra Leone area in West-West Africa. We sought to capitalize upon the relative closeness between the Gullah and Sierra Leoneans to identify regions that differentiate both populations and may hence be under recent population-specific selective pressures. METHODS: We computed a linear regression model of the HapMap YRI principal component (PC2) as a quantitative outcome, using 277 Gullah and 400 Sierra Leonean samples genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0. We adjusted for European admixture via inclusion of the CEU component (PC1) as a covariate. In total, 679,513 SNPs with MAF>5% were used in this analysis. In order to exclude spurious loci, only regions where at least one SNP met genome-wide significance (P<5E-10) and a second significant SNP (P<1E-07) in LD with it were considered. RESULTS: Nine regions met our criteria as those that best differentiate the Gullah from the Sierra Leonean. The most significant was a ~2 Mb region at Xq22.2-q22.3 around the IL1RAPL2 gene, where 4 SNPs had P<5E-10. Other regions included 3q12.3, 4q35.1, 6p23, the extended HLA at 6p22.1-21.32, 7p15.3, 10q11.22, 10q25.1, and 14q24.2. Four additional regions with at least 3 SNPs in LD with P<1E-07 within 1 Mb were identified. Consistent with the literature, our scan identified the HLA and other regions harboring multiple immune-related genes. CONCLUSION: We have identified several regions that differentiate the Gullah from the Sierra Leoneans, suggesting that recent selection may be operating at these loci. Given the relative homogeneity of the Gullah and their genetic proximity to Africans from Sierra Leone, identification of regions that might be under selection in the Gullah has the potential to elucidate disease risks in AA.

2052T

Population structure and selection pressures: short indels and structural variants from NGS in 250 Dutch trios.
A. Abdellaoui[1], V. Guryev[2,3], L. Francioli[4], J.Y. Hehir-Kwa[4], W. Kloosterman[5], T. Marschall[6], A. Schoenhuth[6], E. Lameijer[7], S. Koval[8], F. Hormozdiari[9], J. de Ligt[5], N. Amin[10], F. van Dijk[11,12], L. Karssen[10], H. Mei[13], E.E. Eichler[9], D.I. Boomsma[1], K. Ye[7], Genome of the Netherlands Consortium. 1) Biological Psychology, Vrije Universiteit (VU) Amsterdam, Amsterdam, Noord Holland, Netherlands; 2) European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 3) Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 6) Life Sciences Group Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands; 7) Section Molecular Epidemiology, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands; 8) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; 9) Department of Genome Sciences, University of Washington, Seattle, Washington; 10) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 11) Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands; 12) Genomics Coordination Center, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands; 13) Netherlands Bioinformatics Centre, Nijmegen, The Netherlands.

While microarray data have contributed much to population genetics, the higher resolution of whole-genome sequence data is expected to yield new insights about population stratification, population history, the prevalence of selection pressures, and the identification of functional variants under selection. Ancestry differences in the Netherlands show clear geographic distributions, as previously mapped using principal component analysis (PCA) on microarray SNPs. Ancestry-informative PCs can reveal the consequences of population history, and can also be used to detect selection pressures and traces of migration (Abdellaoui et al, 2013). The Genome of the Netherlands (GoNL: http://www.nlgenome.nl/) is a whole-genome-sequencing project in a representative sample consisting of 250 family trios from all provinces in the Netherlands. Sequencing was done on the Illumina HiSeq 2000 platform at the Beijing Genomics Institute (BGI) on blood-derived DNA from uncultured cells and accomplished coverage was 14-15x. Several methodologies (read depth, read pair, de novo assembly, split-read alignment, and combined approaches) were applied for detection of short indels and different types and size ranges of structural variants (SVs), including deletions, large insertions, tandem duplications, inversions, mobile element insertions and translocations. PCA on short common indels revealed two PCs that show regional differences that were not previously observed, in addition to two PCs showing regional differences previously captured with microarray SNPs (differentiating between North and South, and between East and West). Functional variants under selection (and thus with important phenotypic consequences) can be identified by comparing the distribution of alleles between the subpopulations identified by the PCs. Similar analyses are currently being undertaken for additional types of rare and common SVs as well as a combined set of short indels and SVs. [Reference: Abdellaoui A et al: Population structure, migration, and diversifying selection in the Netherlands. Eur J Hum Genet 2013; e-pub ahead of print 27 March 2013.].

2053F

Reconciling Migration Models to the Americas with the Variation of North American Native Mitogenomes. *A. Achilli[1], U.A. Perego[2,3], H. Lancioni[1], A. Olivieri[3], F. Gandini[3], B. Hooshiar Kashani[3], V. Battaglia[3], V. Grugni[3], N. Angerhofer[2], M.P. Rogers[4], R.J. Herrera[5], S.R. Woodward[6], D. Labuda[7], D. Glenn Smith[8], J.S. Cybulski[9], O. Semino[3], R.S. Malhi[4,10], A. Torroni[3].* 1) Department of Cellular and Environmental Biology, University of Perugia, Perugia, Perugia, Italy; 2) Sorenson Molecular Genealogy Foundation, Salt Lake City, UT 84115, USA; 3) Dipartimento di Biologia e Biotecnologie, Università di Pavia, Italy; 4) Department of Anthropology and Institute for Genomic Biology, University of Illinois Urbana-Champaign, USA; 5) Department of Biological Sciences, Florida International University, Miami, Florida 33199, USA; 6) AncestryDNA, Provo, USA; 7) Département de Pédiatrie, Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Canada; 8) Department of Anthropology, University of California, USA; 9) Canadian Museum of Civilization, Gatineau, Canada; 10) Institute for Genomic Biology, University of Illinois Urbana-Champaign, IL 61801, USA.

The old tripartite linguistic subdivision of Native Americans into three major groups has been revived by recently reported patterns of nuclear genome diversity indicating that Native Americans descend from three streams of Asian/Beringian gene flow. Although this scenario was suggested by early mitochondrial DNA (mtDNA) data, neither this model, nor the alternative scenarios have been fully evaluated by employing entire mitochondrial genomes. In this study we focused our attention on two North American mtDNA haplogroups, known as A2a and B2a. Analyses of B2a mitogenomes, which are absent in Eskimo-Aleut- and northern Na-Dene-speakers, revealed that this haplogroup arose in North America ≈11-13 thousand years (ka) ago from one of the founder Paleo-Indian B2 mitogenomes. In contrast, haplogroup A2a, which is typically found among Eskimo-Aleuts and Na-Dene, but also present in the easternmost Siberian groups, originated only 4-7 ka ago in Alaska, led to the first Paleo-Eskimo settlement of northern Canada and Greenland, and contributed to the formation of the Na-Dene gene pool. However, mitogenomes also show that Amerindians from northern North America, without distinction between Na-Dene and non-Na-Dene, were heavily affected by an additional and distinctive Beringian genetic input. In conclusion, most of the contemporary mtDNA variation observed along the double-continent stems from the first wave that from Beringia followed the Pacific coastal route and is dated to ≈15-18 ka ago based on the 16 mitogenome founders identified so far. This was accompanied or followed by a second inland migratory event, marked by haplogroups X2a and C4c, which affected all Amerindian groups of Northern North America. Considerably later, the ancestral A2a carriers spread from Alaska undertaking both a westward migration to Asia and an eastward expansion into the circumpolar regions of Canada. Thus, the First American founders left the greatest genetic mark but the original maternal makeup of North American Natives was subsequently reshaped by additional inputs and local population dynamics, making even a three-wave view too simplistic.

2054W

Data from extended 1000 Genomes phase I populations refine comparison of X-linked and autosomal population genetic patterns. *L. Arbiza[1], S. Gottipati[1], A. Siepel[1], A.G. Clark[1,2], A. Keinan[1].* 1) Dept. of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Dept. of Molecular Biology and Genetics, Cornell University, Ithaca, NY.

Contrasting the patterns of variability on human chromosome X and the autosomes has grown in power and informativeness with the rapid expansion of whole-genome sequence data. Such a comparison can shed light on demographic processes, differences in the histories of males and females, and the past action of natural selection. Previous studies were based on either genotyping data sets, a small set of resequenced loci in six populations, or genomic sequencing data from at most two populations. These studies leave several open questions. Most notably, to what extent do genome-wide patterns generalize to additional populations? Previous studies also differed in their approaches and population genetic measures, which has led to apparent discrepancies among results. For example, estimates of population differentiation (F_{ST}) reflect more recent patterns of variation ~since the time of split of the specific populations compared~ than nucleotide diversity (π), which captures more ancient epochs. To address these questions and extend our previous results, we capitalized on the 1000 Genomes phase I data, consisting of 14 populations from four ancestry-based continental groups. We improved on our methodology for normalizing X-linked and autosomal π by genetic divergence from different outgroups. To obtain refined resolution of different epochs and disentangle previous results, we also estimated F_{ST} between each pair of populations and compared the level of differentiation of X-linked and autosomal SNPs along each branch in the population tree. These analyses point to three results: (1) While less sensitive to more recent demographic events than estimates of population differentiation, estimates of π support a reduction of the ratio of X-linked to autosomal effective population size (X/A) in non-African compared to African populations. (2) A wide array of models of human demographic history compiled from the literature only partially predict the magnitude of this observed X/A reduction, supporting the need to invoke additional factors such as sex-biased events during the out-of-Africa dispersal. (3) F_{ST} results provide further evidence and accentuate these observations, additionally pointing to variability across different branches, and suggesting changes in mating patterns through human history. Finally, beyond demographic inference, we characterize the relative effect of selection on X and A in different epochs by studying F_{ST} as a function of distance from genes.

2055T

Population differentiation of two Brazilian populations inferred from non-genic Alu insertions and microsatellites of the Class I Major Histocompatibility Complex. *A.C. Arcanjo[1], H.A. Sakata[3], R.C.P. Toledo[1], E.C. Castelli[2], H.R. Magaldi[5], J.A. Peña[4], S.F. Oliveira[1,3].* 1) Biologia Animal Post-Graduate Program, Institute of Biology, University of Brasília, Brasilia, Distrito Federal, Brazil; 2) Department of Pathology, School of Medicine of Botucatu, University of the State of Sao Paulo, Botucatu, Sao Paulo, Brazil; 3) Department of Genetics and Morphology, Institute of Biology, University of Brasilia, Brasilia, Distrito Federal, Brazil; 4) Department of Genetics, Physical Anthropology and Animal Physiology, Faculty of Science and Technology, University of the Basque Country, Bilbao, Spain; 5) Biologia Molecular Post-Graduate Program, Institute of Biology, University of Brasilia, Brasilia, Distrito Federal, Brazil.

Genetic markers located in important genomic regions, such as genes or regulatory sequences, are most commonly affected by natural selection and thereby behave very differently from neutral regions of the human genome. The human MHC (6p23.1) is a region of great importance due to the presence of genes related to action and regulation of the human immune system. Several studies have reported that natural selection might be acting upon its genes, but few have reported the behavior of non-genic elements in such region. Aiming to infer whether this region is under selection, and not solely its genes, this study intended to evaluate the behavior of the Class I MHC region based on Alu insertions and STR markers located nearby but not within the genes in two Brazilian populations. These populations have shown to be very different regarding neutral markers of the genome. 122 samples from two distinct populations, an urban admixed population (Brasília, 65 samples) and an afro-derived rural population (Kalunga, 57 individuals) were studied for the loci AluMICB, AluTF, AluHJ AluHG e AluHF and D6S272, D6S2690, D6S2705, D6S478 and D6S2707. The fragments were amplified by PCR and statistics based on allelic frequencies were obtained with the Arlequin 3.5 and Genepop 4.1 softwares. Deviations from neutrality were inferred with the Slatkin's Implementation of Ewens-Waterson Neutrality Test within the PyPop 0.7.0 software. Both populations showed deviations from Hardy-Weinberg's Equilibrium for the locus D6S2705, probably due to the biggest quantity of missing data for this locus. AluHG showed heterozygote excess even in accordance with HWE, which may indicate influence of evolutionary mechanisms due to its proximity to HLA-G gene. Even considering the different rates of mutation among microsatellites and Alu insertions, both sets of markers indicate that the populations are similar and show little population divergence, in contrast to previous studies regarding neutral markers and Class III MHC microsatellites. The Ewens-Watterson Neutrality test indicated that the D6S478 and D6S2707 loci do not behave neutrally when analyzed and this could be due to the proximity of HLA-G gene, which has already shown deviations from neutrality. These results indicate that the Class I MHC is probably under selection and the genetic markers studied behave very differently than neutral markers of the genome in the studied populations.

2056F

Y chromosomal sequences from southern Africa allow direct comparison of paternal and maternal prehistories. *C. Barbieri[1], S. Lippold[1], R. Schröder[1], S.W. Mpoloka[2], M. Stoneking[1], B. Pakendorf[3].* 1) Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; 2) Department of Biological Sciences, University of Botswana, Gaborone, Botswana; 3) Laboratoire Dynamique du Langage, UMR5596, CNRS and Université Lyon Lumière 2, Lyon, France.

Comparisons of mtDNA and non-recombining Y-chromosome (NRY) variation can be quite informative about sex-biased differences in human demographic history such as population size variation, unbalanced gene flow after contact, and different origins and migration routes; yet NRY analyses are hampered by the common use of ascertained haplogroup defining SNPs, which precludes the possibility to perform many demographic analyses of NRY variation. Recently we developed a capture-based method for enriching Illumina sequencing libraries for about ~500 kb of NRY sequence which overcomes limitations of SNP-based typing of NRY variation and allows direct comparisons to complete mtDNA genome sequences. We applied this method to a broad DNA collection of ~300 southern African samples predominantly centered on the so-called Khoisan, which was previously analyzed for complete mtDNA genome sequences as well as for autosomal SNPs. Preliminary results show a high level of NRY variation and the presence of highly divergent haplotypes. Our dataset will be compared to the complete mtDNA genome sequences from the same dataset as well as to comparable Y chromosome sequence data generated for other African populations of the CEPH panel, to frame the variability of Khoisan groups within the continent. In particular, we will investigate the extent of sex-biased admixture among groups, the potential contribution of East African pastoralists to southern African groups, and the comparative demographic history of maternal and paternal lineages in southern Africa.

2057W

Population structure in five Mennonite communities. *K.G. Beaty[1], P.E. Melton[2], M.J. Mosher[3], M.H. Crawford[1].* 1) Department of Anthropology, University of Kansas, Lawrence, KS; 2) Center for Genetic Origins of Health and Disease, University of Western Australia, Perth, Australia; 3) Department of Anthropology, Western Washington University, Bellingham, WA.

Mennonites are a branch of the Anabaptists religious groups that formed in Europe during the time of the Reformation. Those who followed Menno Simmons, the Dutch-North German branch, were later known as Mennonites, and this group later fled from the Netherlands and Germany during the 16th Century to Prussia, and in the 1850s came to the United States. The movement and genealogical history of this group has been well documented and indicate population fissions along familial lines after arrival in the United States. This project examines Y-STR profiles consisting of 17 loci from 94 males from the Kansas Mennonite communities of Alexanderwohl (n=13), Lonetree (n=20), Gardenview (n=15), and Meridian (n=25), as well as a community of Mennonites from Henderson, Nebraska (n=21). The most prevalent haplotype in all communities was R1b, which represented 56% of the lineages in the entire sample, with other haplogroups represented in frequencies below 10% including E1b, G2a, I1, I2, J2, L, Q, and R1a. There are no shared Y-STR haplotypes between the communities and haplogroup frequencies vary considerably between populations, supporting the historical evidence that original migrating groups split along familial, and specifically, paternal lines upon arrival to the United States. The Alexanderwohl and Henderson communities, which spilt off from group that migrated from the Ukraine in 1874, are the only two communities with Haplogroups G2a and J2a1. The Meridian community, a heterogenous group believed to be an amalgamation of Swiss, German, and Dutch Mennonites as well as Amish, is represented by 8 haplgroups, including haplogroups L found only in Meridian and a high frequency of E1b1b (12%) when compared to the other populations. It also has a mean number of pairwise differences of 9.3090 (+/- 4.2857) and gene diversity of 0.9621 (+/- 0.0007), higher than any other Mennonite community. Lonetree, differs from the other communities in its high frequency of haplogroups R1a (20%), while Gardenview differs from the other communities with a high frequency of haplgroup I2a (33.4%). Overall, the Y-STR profiles of these groups support the documented movement, fission, and in the case of Meridian, fusion, of Mennonite communities after their arrival to the United States.

2058T

Identity by descent segment detection in sequence data. *B.L. Browning[1,2,3], S.R. Browning[2].* 1) Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Biostatistics, University of Washington, Seattle, WA; 3) Genome Sciences, University of Washington, Seattle, WA.

Chromosome segments shared by inheritance from a recent ancestor without recombination are said to be identical by descent. Segments of identity by descent (IBD) are ubiquitous in population samples. Detection of these segments is useful for a variety of applications including analysis of fine-scale population structure and inference of recent effective population size. Existing methods for IBD segment detection were designed for SNP array data. Sequence data can differ strongly from SNP array data in density of variants, allele frequency spectrum and genotype error rates. We investigate the performance of IBD detection methods with sequence data; we present a new IBD detection method, IBDseq, that is specifically designed for sequence data; and we apply IBDseq to autosomal whole genome sequence data for 2400 individuals from the UK10K project.

With IBDseq we do not model LD or haplotype phase, but instead thin variants to eliminate strong inter-variant LD. This is because much of the information for IBD detection in sequence data comes from the large number of low frequency variants which are difficult to phase and tend not to be in high linkage disequilibrium (LD) with other variants. Genotype error is explicitly modeled in IBDseq because current sequence data can have relatively high error rates compared to SNP array data, and because recent mutations can disrupt allelic identity in IBD segments. To reduce computation time, IBDseq calculates maximal subsequence scores rather than employing a full hidden Markov model, and it uses multi-threading to take advantage of multi-core compute nodes. We compare IBDseq to Refined IBD, Germline and PLINK on realistic simulated sequence data. IBDseq has higher power than the other methods, particularly for detection of very short (<1 cM) segments. In UK10K project sequence data on 2400 individuals, IBDseq finds 444 million IBD segments with a median length of 0.49 cM.

2059F

Synthesizing genetic and genealogical data to trace historical waves of European and African immigration to the United States. *J.K. Byrnes[1], J.M. Granka[1], K. Noto[1], R.E. Curtis[2], Y. Wang[1], M.J. Barber[1], N.M. Myres[2], C.A. Ball[1], K.G. Chahine[2].* 1) Ancestry.com DNA L.L.C., 153 Townsend St., Ste. 800, San Francisco, CA; 2) Ancestry.com DNA L.L.C., 360 West 4800 North, Provo, UT.

Census data can reveal waves of immigration, with immigrants from a single source population migrating to roughly the same U.S. locations within a short period of time. A refinement of the timings and locations of migratory waves using genetic and genealogical data would be useful not only for understanding the peopling of the U.S., but also for understanding patterns of genetic disease risk. To examine migratory waves using genealogical and genetic data, we first identify 'IBD-enriched groups': groups of U.S. AncestryDNA customers who share elevated amounts of DNA identity-by-descent (IBD) with a reference panel of individuals with known deep ancestry from a single location. Elevated IBD with individuals from one location could indicate a source population from which ancestors of a customer originated. As a proof of principle, for individuals in a group that is IBD-enriched for a particular location, we search for an enrichment of ancestors born in this location across the collection of their pedigrees. Then, for each IBD-enriched group, we search the collection of pedigrees for enrichment of more recent birth locations, representing migratory destinations within the U.S. Among over 100,000 U.S. AncestryDNA customer samples, we identify IBD-enriched groups who have elevated IBD with more than 20 European countries and seven West African population groups. For each IBD-enriched group, we present maps showing patterns of ancestor birth location enrichment through time. As expected, customers in an IBD-enriched group to a particular location have an enrichment of ancestor births in that location, though this enrichment is dependent on pedigree completeness (e.g. it is rare for African Americans to have known ancestors from West Africa in their pedigrees). For some punctuated migratory waves, such as the arrival of Norwegian immigrants in the Midwest during the 19th century, clear signals of both source and migratory destinations are visible. These locations are temporally ordered, with ancestors from the migratory destination appearing more recently in the pedigrees than ancestors from the source location. Finally, we are able to pinpoint the age and origin of particular haplotypes defining IBD-enriched groups. With our approach, we demonstrate that we can estimate individual ancestral origins and detail human migratory history by jointly studying estimates of genetic relatedness along with genealogical data.

2060W

Population demography and maternal history of Oceania. *A.T. Duggan[1], B. Evans[2], M. Kayser[3], R.J. Trent[4], D.A. Merriwether[5], G. Koki[6], F.R. Friedlaender[7], J.S. Friedlaender[7], M. Stoneking[1].* 1) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Sachsen, Germany; 2) College of Asia and the Pacific, Australian National University, Canberra, Australia; 3) Department of Forensic Molecular Biology, Erasmus MC - University Medical Center Rotterdam, The Netherlands; 4) Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 5) Department of Anthropology Binghamton University Binghamton, NY, USA; 6) Institute for Medical Research, Goroka, EHP, Papua New Guinea; 7) Department of Anthropology, Temple University, Philadelphia, PA, USA.

We present a large-scale study of mtDNA diversity across Near and Remote Oceania with whole-genome mtDNA sequencing and a sample collection of more than 1,300 individuals spanning from the Bismarck Archipelago in the west to the Cook Islands in the east. As the location of at least two major migration events (initial colonization over 40,000 years ago, followed by an expansion of Austronesian-speaking migrants around 3,500 years ago), Oceania provides a unique opportunity to study the effects of population admixture. Our results support the idea of sex-biased admixture between the resident populations and the migrants of the Austronesian expansion. We find that haplogroups of putative Asian origin which are thought to have spread with the Austronesian expansion are found at high frequency in all but two populations and, in general, we see little evidence of distinction between Papuan and Austronesian speaking populations. Santa Cruz, which is part of the Solomon Islands but geographically distinct from the main island chain and considered part of Remote Oceania, has long been considered a linguistic oddity and is now accepted to represent a very deep branch in the Oceanic language family. We find that it is also a genetic outlier, with potential direct connections to the Bismarck Archipelago not evident in the main Solomon Islands chain. In this expanded dataset, we find additional evidence of instability and increased heteroplasmy at the 'Polynesian motif' position 16247, further confirming previous findings restricted to the Solomon Islands.

2061T

The Saudi Arabian Genome Reveals a Two Step Out-of-Africa Migration. *J.J. Farrell[1], A.K. Al-Ali[2], L.A. Farrer[1], A.N. Al-Nafaie[3], A.M. Al-Rubaish[4], E. Melista[1], Z. Naserullah[5], A. Alsuliman[6], P. Sebastiani[7], M.H. Steinberg[1], C.T. Baldwin[1].* 1) Department of Medicine, Boston University School of Medicine, Boston, MA; 2) Department of Clinical Biochemistry, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia; 3) Department of Hematology, College of Medicine, University of Dammam, Dammam, Kingdom of Saudi Arabia; 4) Department of Internal Medicine, College of Medicine, University of Dammam Dammam, Kingdom of Saudi Arabia; 5) Department of Pediatrics, Maternity & Child Hospital, Dammam, Kingdom of Saudi Arabia; 6) Department of Hematology, King Fahd Hospital, Hafof, Al-Ahsa, Kingdom of Saudi Arabia; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA.

Here we present the first high-coverage whole genome sequences from a Middle Eastern population consisting of 14 Eastern Province Saudi Arabians. Genomes from this region are of interest to further answer questions regarding 'Out-of-Africa' human migration. Applying a pairwise sequentially Markovian coalescent model (PSMC), we inferred the history of population sizes between 10,000 years and 1,000,000 years before present (YBP) for the Saudi genomes and an additional 11 high-coverage whole genome sequences from Africa, Asia and Europe.

The model estimated the initial separation from Africans at approximately 110,000 YBP. This intermediate population then underwent a long period of decreasing population size culminating in a bottleneck 50,000 YBP followed by an expansion into Asia and Europe. The split and subsequent bottleneck were thus two distinct events separated by a long intermediate period of genetic drift in the Middle East. The two most frequent mitochondria haplogroups (30% each) were the Middle Eastern U7a and the African L. The presence of the L haplogroup common in Africa was unexpected given the clustering of the Saudis with Europeans in the phylogenetic tree and suggests some recent African admixture. To examine this further, we performed formal tests for a history of admixture and found no evidence of African admixture in the Saudi after the split. Taken together, these analyses suggest that the L3 haplogroup found in the Saudi were present before the bottleneck 50,000 YBP. Given the TMRCA estimates for the L3 haplogroup of approximately 70,000 YBP and the timing of the Out-of-Africa split, these analyses suggest that L3 haplogroup arose in the Middle East with a subsequent back migration and expansion into Africa over the Horn-of-Africa during the lower sea levels found during the glacial period bottleneck.

These results are consistent with the hypothesis that modern humans populated the Middle East before a split 110,000 YBP, underwent genetic drift for 60,000 years before expanding to Asia and Europe as well as back-migration into Africa. Examination of genetic variants discovered by Saudi whole genome sequencing in ancestral African populations and European/Asian populations will contribute to the understanding human migration patterns and the origin of genetic variation in modern humans.

2062F

The CARTaGENE Genomics Project : Population structure, local ancestry contributions and relatedness analysis of the French Canadian founder population. *H. Gauvin[1,2], Y. Idaghdour[1], J. Hussin[1], J.-P. Goulet[1], J.-C. Grenier[1], M. Capredon[1], A. Hodgkinson[1], T. de Malliard[1], V. Bruat[1], E. Gbeha[1], E. Hip-Ki[1], P. Awadalla[1,3].* 1) Sainte-Justine Hospital Research Center, Université de Montréal, Montreal, Canada; 2) Department of Social and Preventive Medicine, Université de Montréal, Montreal, Canada; 3) Department of Pediatrics, Université de Montréal, Montreal, Canada.

The province of Quebec was colonized four hundred years ago by settlers coming from France. In the early settlement of Quebec, some First Nations populations were allied to French settlers and exchanged genetic material as the result of mixed unions between an aboriginal mother and a French father. After the British Conquest, the rapidly expanding French Canadian population generally grew in relative isolation with limited exchange with British, and other incoming populations such as Acadians and Loyalists. The colonization of the territory took place in successive waves leading to regional founder effects contributing to the uniqueness of the French Canadian population. Today, about 80% of the province's 8 million inhabitants are French speaking. Starting in 2010, over 30,000 people from the province of Quebec were recruited to be a part of the CARTaGENE project. Genotyping data (Illumina Omni2.5M) was generated for ~1000 participants sampled in three distinct regions of the province: the Montreal area, the Quebec City area and the Saguenay region. We use the genotypes to study regional differentiation due to demographic history, admixture and migration patterns within Quebec. We analyze identity by descent (IBD) segments to infer relatedness between participants and trace back regional populations flow inside Quebec. Inferred IBD segments are also analyzed to attest their ancestral origin and their breakpoints are compared to recent and ancestral recombination breakpoints across the genome. We also investigate local ancestry of participants with a panel consisting of Native American and European populations. Specifically, we (i) identify regions of the genome shared between individuals and inherited from common ancestors; (ii) characterize IBD tracts in relation to recombination map and (iii) infer local ancestry across the genome to quantify the relative contributions of the different populations. This work refines previous analysis of population structure in Quebec, provides a more accurate picture of how the different populations contributed to the actual French Canadian genetic pool and gives insight on how their contribution is linked to mutations causing differences in disease prevalence throughout regions of Quebec.

2063W

Reconstruction of Ancestral Human Genomes from Genome-Wide DNA Matches. *J.M. Granka[1], R.E. Curtis[2], J.K. Byrnes[1], M.J. Barber[1], N.M. Myres[2], K. Noto[1], Y. Wang[1], C.A. Ball[1], K.G. Chahine[2].* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Provo, UT.

Individuals who lived long ago may still have much or all of their genome present in modern populations. The genomes of these individuals exist in small segments broken down by recombination and inherited in part by his or her descendants. If such an individual had many children, leading to a large number of descendants today, much of the ancestral genome will be present in modern populations. For the pairs of descendants with the 'target' ancestor as their most recent common ancestor (MRCA), any region of their genomes shared identical-by-descent (IBD) most likely represents the corresponding region of the ancestor's genome. Given a set of pairs of individuals linked to the same MRCA, we develop a novel computational approach to reconstruct the haplotypes of the MRCA from the IBD segments and haplotypes of the descendants. With simulated data we assess the performance of our method, affected by factors such as quality of genealogical trees used to infer the MRCA, reliability of inferred IBD, coverage of IBD segments, number of descendants of the MRCA, and number of sampled descendants. To demonstrate the utility of our method, we examine over 125,000 individuals in the AncestryDNA database with phased genome-wide single nucleotide polymorphism data and detailed genealogical information. After first identifying regions of the genome shared IBD between all individuals, we selected one group of several hundred individuals with an 18th century couple as a known MRCA. Using our method to tile together these individuals' IBD segments, we are able to reliably construct the ancestral couple's four haplotypes in large genomic regions with high coverage of IBD segments. In regions of the genome with lower IBD coverage, we are unable to identify and construct all haplotypes with certainty. Our study demonstrates the possibility of reconstructing the genomes of human ancestors, with large family sizes and a large number of living descendants, who lived one to even 12 generations ago. The ability to reconstruct the genomes of human ancestors using genetic and genealogical data has exciting implications in the fields of population genetics, medical genetics, and genealogy research.

2064T

The population history of African and Caribbean vervet monkeys inferred from 130 whole-genome-sequenced samples. *Y. Huang[1], H. Svardal[2], C.A. Schmitt[1], A.J. Jasinska[1], J. Wasserscheid[3], N. Juretic[3], Y.J. Jung[1], M. Muller-Trutwin[4], B. Jacquelin[4], M. Antonio[5], M. Dione[5], J.P. Grobler[6], R.K. Wilson[7], K. Dewar[3], W. Warren[7], G. Weinstock[7], T.R. Turner[6,8], M. Nordborg[2], N.B. Freimer[1], First two authors contributed equally.* 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA; 2) Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, Austria; 3) McGill University and Genome Quebec Innovation Centre, Montréal, Canada; 4) Virology Department, Institut Pasteur, Paris, France; 5) Medical Research Council, Fajara, The Gambia; 6) Department of Genetics, University of the Free State, Bloemfontein, South Africa; 7) The Genome Institute at Washington University in St. Louis, St. Louis, MO, 63108; 8) Department of Anthropology, University of Wisconsin-Milwaukee, Milwaukee, WI, 53201.

Vervet monkeys (genus *Chlorocebus*; African green monkey) are Old World monkeys widely distributed in sub-Saharan Africa and subsequently introduced to the Caribbean. Vervets are broadly used in biomedical research, but there has been relatively few genetic studies and their taxonomy is disputed. Understanding their population genetic history, including genetic relationship among vervet subspecies, is essential for designing association studies of quantitative traits and for elucidating host-pathogen relationships, e.g., with simian immunodeficiency virus (SIV). To investigate vervet population genetics, we have carried out whole genome sequencing of 130 monkeys representing five African subspecies and three Caribbean populations. Our results, from both global ancestry analysis and comparisons of allele frequencies are concordant with the historical literature, in suggesting that the Caribbean populations derive from West African *sabaeus* monkeys. The significantly slower decay in linkage disequilibrium in the Caribbean samples, compared to those from their presumed ancestral population (from the Gambia), support the hypothesis that currently large Caribbean populations experienced a severe bottleneck at the time of their founding. On the African continent, as predicted by the isolation by distance model, the global ancestry result using whole-genome polymorphism data revealed a pattern of relatedness between vervet subspecies that reflects their current geographic distribution. In the classification of the subspecies *cynosuros*, the genetic data suggest a differing phylogeny from that proposed by most standard taxonomies. Additionally, these data indicate patterns of genetic clustering between monkeys sampled at different locations corresponding closely to patterns of variation in SIV sequences observed in samples from the same monkeys, suggesting longstanding co-evolution between host and pathogen. Past population size changes of the subspecies, inferred from the time to the most recent common ancestor distribution along the autosomes, suggested that they started to diverge from each other ~500,000 - 2,000,000 years ago. The results improve our understanding of vervet population history, providing a foundation for investigation of biomedically important traits.

2065F

The Sherpa and Tibetans share a common genetic history and adaptations to high-altitude. *C. Jeong[1], D.B. Witonsky[1], G. Alkorta-Aranburu[1], B. Basnyat[2], M. Neupane[2], J.K. Pritchard[1,3], C.M. Beall[4], A. Di Rienzo[1].* 1) Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America; 2) Nepal International Clinic, Kathmandu, Nepal; 3) Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America; 4) Department of Anthropology, Case Western Research University, Cleveland, Ohio, United States of America.

The Sherpa are an ethnic group in Nepal, traditionally living in high-altitude (HA) regions of the Himalayas. Linguistic and historical evidence indicates a close relationship with Tibetans. Furthermore, they share with Tibetans a low concentration of hemoglobin, a key physiological adaptation to HA hypoxia. These findings suggest that the Sherpa and Tibetans constitute closely related branches of the East Asian (EA) populations sharing phenotypic HA adaptations. To develop this idea further, we studied the genetic history of the Sherpa in relation to other Asian populations. First, we asked if the Sherpa are genetically distinct by comparing genome-wide genotyping data of 69 Sherpa native residents at 3800m with the Human Genome Diversity Panel data using principal component analysis (PCA) and ADMIXTURE. These show that the Sherpa genome contains a large portion of a genetic component clearly distinguishable from lowland EA populations. Sherpa individuals have neither South nor Central Asian genetic components and they are tightly clustered with other EA individuals in the global PCA plot. Second, we inferred a past population size change of the Sherpa by applying the pairwise sequential Markovian coalescence method to whole genome sequences (27-30x) of two males. We compared the Sherpa results with those of Han and Dai individuals, which were sequenced at similar depth (Meyer et al, 2012). This comparison suggests an early split of the Sherpa and lowland EA populations about 20,000-30,000 years ago and no evidence for the recent population expansion that characterizes the history of other EA populations. We estimate the effective population size of the Sherpa after the split at about 4,000 individuals. Third, a maximum-likelihood tree of EA populations supports a branch consisting of the Sherpa and Tibetans (from Yunnan Province of China; Beall et al, 2010). Last, we found that 21 of 36 EPAS1 (endothelial PAS domain-containing protein 1) gene SNPs with a significant association with hemoglobin level in a previous study of Tibetans (Beall et al, 2010) also show an association with nominal $p < 0.05$, all of which have the same direction of allelic effect as in Tibetans. The estimated effect sizes also match the Tibetan results. To summarize, our findings support a shared genetic history and genetic adaptation to HA of the Sherpa and Tibetans.

2066W

The Genetic Architecture Of Skin Pigmentation In Southern Africa. *A.R. Martin[1], J.M. Granka[2], C.R. Gignoux[3], M. Möller[4], C.J. Werely[4], J.M. Kidd[5], M.W. Feldman[6], E.G. Hoal[4], C.D. Bustamante[1], B.M. Henn[7].* 1) Stanford University, Genetics Department, Stanford, CA; 2) Ancestry.com, San Francisco, CA; 3) University of California, Pharmaceutical Sciences, San Francisco, CA; 4) Stellenbosch University, Health Sciences, Tygerberg, South Africa; 5) University of Michigan, Department of Human Genetics, Ann Arbor, MI; 6) Stanford University, Department of Biological Sciences, Stanford, CA; 7) SUNY, Deptartment of Ecology and Evolution, Stony Brook, NY.

Skin pigmentation is one of the most recognizably diverse phenotypes in humans across the globe, but its highly genetic basis has mainly been studied in northern European and Asian populations. The Eurasian pigmentation alleles are among the most differentiated variants in the genome, suggesting strong positive selection for light skin pigmentation. Light skin pigmentation is also observed in the far southern latitudes of Africa, among KhoeSan hunter-gatherers of the Kalahari Desert and other populations. The KhoeSan hunter-gatherers, believed to have diverged from other populations 100,000 years ago, maintain extraordinary levels of genetic diversity, but it is unknown whether light skin pigmentation represents convergent evolution or the ancestral human phenotype. We have collected saliva samples, ethnographic information, and pigmentation phenotypes from 123 individuals in the ≠Khomani San from the Kalahari. To understand the genetic basis for light skin pigmentation, we have genotyped and exome sequenced 91 ≠Khomani San individuals to high coverage, generating one of the largest indigenous African exome datasets sequenced outside of 1000 Genomes. Because linkage disequilibrium decay is rapid in this population, we have assessed parameters influencing phasing and imputation accuracy since ideal reference panels do not exist. We have also pursued multiple genotype/phenotype mapping methods, including a mixed model approach, admixture mapping, and linkage mapping. After controlling for admixture from European and Bantu-speaking populations, we find that globally common variants are not significantly associated with pigmentation. Rather, our results indicate that there are a multitude of rare variants in known pigmentation genes, and suggest that previously unidentified genes acting in canonical pigmentation pathways may be involved. Our results highlight the strength of diverse population studies to explain phenotypic variation in the context of human evolutionary history.

2067T

A Novel Likelihood Ratio Test for Sex-Bias and the Effect of Cryptic Sex-Bias on the Estimation of Demographic Parameters. *S. Musharoff[1], S. Shringarpure[1], C.D. Bustamante[1], S. Ramachandran[2].* 1) Genetics, Stanford Univ Sch Medicine, Stanford, CA; 2) Ecology and Evolutionary Biology, Brown University, Providence, RI.

Sex-bias is defined as an unequal number of breeding males and females in a population. This can be caused by variance in reproductive success, demographic events involving unequal numbers of males and females, and/or differential selection at sex-linked genomic loci. We present a novel likelihood ratio test for sex-bias in a single population based on the Poisson random field model. Our test has more power to detect a sex-biased demographic event from unlinked or partially linked sites than the commonly-used test statistic Q (the ratio of neutral genetic diversity estimated from the X chromosome to that estimated from the autosomes). We further develop a test for homogeneity of sex-bias in a set of populations and compare its power to the test statistic Q (here, defined in terms of Fst). We apply our tests for sex-bias to whole-genome, exome, and SNP data from globally distributed human populations and compare our results to those from previous studies. In addition to being of fundamental interest, the presence of sex-bias affects demographic inference. Sex-bias, either in the male or female direction, decreases the effective population size of the X chromosome as well as the autosomes of a population. If this reduction in effective population size is not accounted for, demographic parameters estimates (e.g., bottleneck times or divergence times) will be inflated. We assess the effect of cryptic sex-bias on the estimation of demographic parameters using simulated genomic data. We then estimate demographic parameters jointly from the X chromosome and the autosomes from the dataset described above and compare the parameter estimates to those obtained from autosomes alone. These analyses give us a more complete picture of the presence and effect of human sex-biased demography.

2068F

Y-chromosomal variation in native South Americans: bright dots on a gray canvas. *M. Nothnagel[1,2], L. Roewer[3], L. Gusmão[4,5], V. Gomes[4], M. González[4], D. Corach[6], A. Sala[6], E. Alechine[6], T. Palha[5], N. Santos[5], A. Ribeiro-dos-Santos[5], M. Geppert[3], S. Willuweit[3], M. Nagy[3], S. Zweynert[7], M. Baeta[8], C. Núñez[8], B. Martínez-Jarreta[8], F. González-Andrade[9], E. Fagundes de Carvalho[10], D. Aparecida da Silva[10], J. José Builes[11,12], D. Turbón[13], A. Maria Lopez Parra[14], E. Arroyo-Pardo[14], U. Toscanini[15], L. Borjas[16], C. Barletta[17], S. Santos[5], M. Krawczak[2].* 1) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 2) Institute of Medical Infomatics and Statistics, Christian-Albrechts University, Kiel, Germany; 3) Institute of Legal Medicine and Forensic Sciences, Department of Forensic Genetics, Charité - Universitätsmedizin Berlin, Berlin, Germany; 4) Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal; 5) Universidade Federal do Pará, Laboratório de Genética Humana e Médica, Belém, Pará, Brazil; 6) Universidad de Buenos Aires, Facultad de Farmacia y Bioquimica, Servicio de Huellas Digitales Geneticas, Buenos Aires, Argentina; 7) Department of Psychiatry and Psychotherapy, Charité~Universitätsmedizin Berlin, Berlin, Germany; 8) Department of Forensic Medicine, University of Zaragoza, Zaragoza, Spain; 9) Science and Technology Department, Ministry of Public Health, Quito, Ecuador; 10) Laboratorio de Diagnósticos por DNA, Instituto de Biologia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil; 11) GENES Ltda., Laboratorio Genetica Forense y Huellas Digitales del DNA, Medellín, Colombia; 12) Instituto de Biología, Universidad de Antioquia, Medellín, Colombia; 13) Unitat d'Antropologia, Departamento de Biologia Animal, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain; 14) Laboratorio de Genetica Forense, Departamento de Toxicología y Legislación Sanitaria, Facultat de Medicina, Universidad Complutense de Madrid, Madrid, Spain; 15) PRICAI-Fundación Favaloro, Buenos Aires, Argentina; 16) Laboratorio de Genetica Molecular, Unidad de Genetica Medica, Facultad de Medicina, Universidad del Zulia, Maracaibo, Venezuela; 17) Facultad de Ciencias Biológicas, UNMSM-Universidad, Nacional Mayor de San Marcos, Lima, Peru.

While human populations in Europe and Asia have often been reported to reveal a concordance between their extant genetic structure and the prevailing regional pattern of geography and language, such evidence is lacking for native South Americans. In the largest study of South American natives to date, we examined the relationship between Y-chromosomal genotype on the one hand, and male geographic origin and linguistic affiliation on the other. We observed virtually no structure for the extant Y-chromosomal genetic variation of South American males that could sensibly be related to their inter-tribal geographic and linguistic relationships, augmented by locally confined Y-STR autocorrelation. Analysis of repeatedly taken random subsamples from Europe adhering to the same sampling scheme excluded the possibility that this finding was due to our specific scheme. Furthermore, for the first time, we identified a distinct geographical cluster of Y-SNP lineages C-M217 (C3*) in South America, which are virtually absent from North and Central America, but occur at high frequency in Asia. Our data suggest a late introduction of C3* into South America no more than 6,000 years ago and low levels of migration between the ancestor populations of C3* carrier and non-carriers. Our findings are consistent with a rapid peopling of the continent, followed by long periods of isolation in small groups, and highlight the fact that a pronounced correlation between genetic and geographic/cultural structure can only be expected under very specific conditions.

2069W

Analysis of haplotype sharing and recent demographic history in the Netherlands. *P. Palamara, I. Pe'er, The Genome of Netherlands Consortium.* Columbia University, New York City, NY.

Chromosomal segments that are identical by descent (IBD) were recently shown to convey information about population-level features such as demography, natural selection and heritability of common traits. In a recent work [1], we have developed analytical models for the relationship between haplotype sharing and demography, and shown that IBD sharing provides an effective way for reconstructing demographic events of the recent millennia, where classical methods are typically underpowered. We now extend the developed models to accommodate the simultaneous analysis of multiple demes, providing insight into recent migration rates as well as population size fluctuations. Using this approach we analyzed sequencing data for 498 unrelated individuals from 11 Dutch provinces (the Genome of Netherlands Project). Pairs of individuals from all the analyzed provinces are found to share several IBD segments of length greater than 1 centimorgan (cM), suggesting recent common ancestry of these groups. We observe a north-to-south gradient of declining IBD sharing frequency. While the chance of sharing long (>7 CM), extremely recent IBD segments correlates with modern-day geographic distance, shorter segments are more frequently shared with individuals currently residing in the north of the country, regardless of the individuals' modern location. Using the developed analytical methods, we reconstruct coalescent distributions and migration rates across the analyzed provinces. In all cases we find evidence for recent exponential growth at different rates for different provinces, with substantial recent gene flow between these demes. Using the retrieved model, we estimate the average haploid pair of Dutch individuals in the studied dataset to find a common ancestor ~1600 years before present, with earlier common ancestors typically found in northern provinces and variation that depends on modern geographic location.

[1] Pier Francesco Palamara, Todd Lencz, Ariel Darvasi, Itsik Pe'er. Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. The American Journal of Human Genetics, Volume 91, Issue 5, 2 November 2012, Pages 809-822.

2070T

Limitations to determining genetic history in the recombinant genome and connection to demographic events from samples of modern populations. *D. Platt, F. Utro, L. Parida.* IBM T. J. Watson Research Center, Yorktown Heights, NY.

We report on our study of the reconstructable genetic history and connections to demographic events impacting genetic history in the recombinant genome derived from modern population samples. We used COSI, which allows for specification of demographic events, such as the out-of-Africa migrations, glacial isolation, and post glacial and agricultural expansions, along with parameterizations for African, European, Asian, and African-American populations, and which allows for the stochastic construction of multiple population histories consistent with these populations' demographic parameters. In prior work, we constructed Ancestral Recombination Graphs (ARGS) to represent the full genetic history of these multiple simulations of these populations. We mathematically defined and characterized Minimal Descriptor (MD) ARGs capturing the essential substructure of an ARG which preserves the genetic landscape of the extant samples, including the topology and edge lengths of the marginal trees. As such, these MDs derived from full COSI output, contain the largest amount of information uniquely specified by extant samples. We identified a measure of 'history' as the number of nodes in a MD due to recombination and coalescence in the minimal descriptor, and measured the number of nodes within any given epoch that were reconstructable from the entire modern population. We showed that coalescence causes the loss of information defining the recombination history, which information loss can include which genome segments had evolved together. This information loss is similar to that observed for equivalent SNPs in non-recombinant phylogenies, where the order in which the SNPs accumulated is lost. Up to 35% of the reconstructable nodes are lost. Further, the reconstructable histories consistently revealed underlying demographic events, even across radically different histories. Here we report, for a given specific history, the reconstructability of the genetic record from samples drawn from the modern population. We generated a representative history, and resampled with varying sample sizes, seeking to quantify how many of the historical nodes were reconstructable from each of the samples, giving a representation of how much of the history could be expected to be recovered from any given sample. We have found the older ARG nodes are broadly supported across multiple resamplings, while more detail in recent events become available with larger sample sizes.

2071F

Analyses of exome variants suggest enrichment for functionally-important variants during human expansion into the Americas. *S. Ramachandran[1,2], J.J. Yang[3].* 1) Ecology & Evolutionary Biol, Brown Univ, Providence, RI; 2) Center for Computational Molecular Biology, Brown Univ, Providence, RI; 3) Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN.

Genetic variation bears signatures of demographic processes such as population divergences, and a modest number of non-coding variants is adequate to classify human individuals into clusters corresponding to their ancestry. In contrast, the population structure of coding variants has not been well characterized, due to (1) low sampling of non-European individuals for exome- and whole-genome studies and (2) controlling for ancestry in genome-wide association (GWA) studies to avoid spurious correlations. However, controlling for ancestry overlooks the fact that haplotypic and allelic diversity decreases with genetic distance from Africa, meaning derived alleles and variants with functional importance may be particularly enriched on East Asian and Native American ancestral backgrounds. We set out to comprehensively examine the relationships between genetic ancestry and coding variant frequency in admixed US Hispanic populations. In a preliminary analysis of 60 unrelated Mexican individuals from the 1000 Genomes (MXL), we evaluated associations between East Asian, African, Native American, and European ancestry coefficients with the number of copies of the reference allele at polymorphisms on the Illumina Exome chip. African and East Asian ancestry is least often associated (<5%) with number of copies of the reference allele. In contrast, the proportion of SNPs associated with levels of Native American ancestry is 10.3%, slightly lower than those associated with European ancestry (11.2%), and Native American ancestry is significantly correlated with copies of the non-reference allele (and the derived allele) in almost all of these cases. The differences observed in allele frequencies at exome SNPs on different ancestral backgrounds suggests that phenotypically-important variants are enriched on Native American backgrounds, which may explain differential incidence in certain diseases ~ such as acute lymphoblastic leukemia (Xu et al. JNCI 2013) ~ for individuals with >10% Native American ancestry.

2072W

Whole exome sequencing of 126 Northeast Asian individuals including Korean, Chinese, Japanese, and Mongolian ethnicities. *A. Rhie[1,2,3], S. Lee[1,2,3], W. Roh[1], J. Shin[1], J.I. Kim[1,2,3], J.S. Seo[1,2,3,4].* 1) Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul 110-799, Korea; 2) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; 3) Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul 110-799, Korea; 4) Macrogen Inc., Seoul 153-023, Korea.

Along with the development of massively parallel sequencing method, human DNA sequencing data have been accumulated at a very rapid rate these days. However, most of them were to focus on target diseases or phenotypes, and genomic information of normal individuals has not been sufficient so far for each population. Here we provide the sequence information of Northeast Asian exomes, along with some sequence-level and population-level analyses. We sequenced 126 exomes of Asian individuals; 37 Koreans, 25 Chinese, 25 Japanese, and 39 Mongolians. Variations including single nucleotide polymorphism (SNP), insertion/deletion, and copy number variation were called for these samples, and the genotype frequencies of sequenced bases were provided including wild type bases (http://asap.gmi.-ac.kr). The bases with low genotype qualities were explored genome-wide, and bases with ≥50% low quality genotypes were found to be largely overlapped with segmental duplications (37.99%), which include several genes such as NBPF10, FLG, FRG1, MUC6, and TBC1D3C. In addition, we tried to identify genetic diversity and population structure of the Northeast Asians and reconstruct the demographic history of these populations. As a result, we showed evidences of the Northern route migration in the East Asia and suggested some Mongolian-specific functional loci (rs1453544: OR4D6, rs1453547: OR5A2, rs148138101: ARHGAP3, rs202130413: ARSB, etc.). In this study, we provided the exomic sequence information of the Northeast Asian populations. We hope that our data provide fundamental information to those studying on Asian populations in genetic or medical research fields.

2073T

The impact of background selection on fine-scale population structure in humans. *R. Torres[1,3], RD. Hernandez[2,3].* 1) Biomedical Sciences Graduate Program; 2) Department of Bioengineering and Therapeutic Sciences,; 3) UCSF, San Francisco, CA.

Patterns of observed genetic polymorphism across human populations have provided great insight about those populations' respective demographic histories and their underlying population structure. Due to geographic and cultural barriers, reduced gene flow has resulted in different distributions of allele frequencies between human populations through the process of genetic drift. Population genetics methods and measures, such as principle components analysis (PCA) and Fst, have uncovered the genetic structure that exists between human populations, especially at the intercontinental scale. However, because of the processes of natural selection and recombination, we should not expect that the effective population size of an individual will be homogenous across the entire genome. Rather, diversity-reducing selection (e.g., genetic hitchhiking and background selection) near functional loci has resulted in a mosaic of different effective population sizes across the human genome. Population genetics theory predicts that the strength of drift is a function of both effective population size and time. Leveraging this fact, we have demonstrated through PCA that loci contributing to principle components of populations across European populations of the POPRES dataset are strongly correlated with background selection in 7 of the top 10 PCs (Kendall's tau, significance assessed by permutation). Similar patterns were replicated with respect to Fst as a function of background selection across the same populations. By controlling for conservation, we will present new analysis to demonstrate that such patterns are not the result of direct selection at each site but due to linkage to selected sites. Our findings will highlight the fine-scale population structure that exists at the intra-continential scale and will illuminate valuable targets for the modeling of subtle population demographic histories.

2074F

Evaluating the impact of sex-biased demography and selection on genomic patterns of human diversity. *A. Walia[1], W. Fu[2], S. Ramachandran[1,3].* 1) Ecology & Evolutionary Biology, Brown University, Providence, RI 02912, USA; 2) Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA; 3) Centre for Computational and Molecular Biology, Brown University, RI 02912 USA.

Disentangling the effects, and relative roles, of demography and selection is central to human population genetics. Due to its mode of transmission, the X chromosome is an ideal system with which to gain insight into both natural selection and sex-biased demographic processes in the human lineage. Recessive deleterious mutations on the X are exposed to selection in males, and the smaller effective population size of the X means the X chromosome experiences drift more strongly than do the autosomes. The ratio of the genetic diversity normalized by divergence between the X chromosome and the autosomes (π_x/π_{Aut}) is expected to be 0.75 . However, past studies (Hammer et al. 2009, Keinan et al. 2009, Hammer et al. 2010, Gottipati et al. 2011) report equivocal estimates for this ratio, and offer different explanations for observed deviations from (π_x/π_{Aut})=0.75. Here, we analyze whole genome data of 567 female individuals from 14 populations, genotyped in Phase 1 of the 1000 Genomes Project. To study the effect of selection on observed genetic diversity, we partition polymorphisms by distance to their nearest gene. We then calculate the ratio of genetic diversity normalized by divergence on the X chromosome to that on the autosomes in different distance intervals to the nearest gene. We compare the (π_x/π_{Aut}) ratio for all pairs of 1000 Genomes populations. Gottipati et al. 2011 suggest that male-biased migration out-of-Africa led to a smaller effective population size of females in non-African populations based on their observation of $(\pi_x/\pi_{Aut})_{CEU}/(\pi_x/\pi_{Aut})_{YRI}$ <1. Their analysis focused on Utah Residents from the CEPH panel with Northern and Western European ancestry(CEU) and Yoruba in Ibadan, Nigeria(YRI). We analyze all 1000 Genomes Phase 1 populations to identify genomic signatures of sex-biased demographic events and social structures in human evolutionary history. This study allows us to examine the relative impact of locally- acting selection and sex-biased demographic processes on human genome-wide variation.

2075W

Estimation of Effective Population Size and Divergence Time of Modern Human Populations Suggest Recent Gene Flow between European and African Populations. *S. Xu, P. Qin, D. Lu, L. Zhao.* Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.

Effective population size (Ne) and population divergence time (TD) are two fundamental parameters to understand demographical history of modern human populations. Taking advantage of genome-wide high density single nucleotide polymorphism (SNP) data, we estimated Ne and TD in 53 populations worldwide based on linkage disequilibrium (LD) information across genome. The estimation by LD-based approach makes use of large number of closely linked markers and makes it possible to show changes of Ne traced over time. Our results revealed the stable African populations and the expansion of non-African populations. We further used two independent approaches to estimate the divergence time between populations, which gave consistent estimation. Interestingly, our estimation of divergence time between European and African was much lower than that between Asian and African, suggesting more gene flow between European and African populations since the initial population divergence. The results in this study will advance our understanding of population demographical history, especially, the "out of Africa" model in the early history of human migration and recent interaction between populations since the initial divergence.

2076T

Spatial Localization At Each Locus In the Genome. *W. Yang[1], A. Platt[2], C. Chiang[2], E. Eskin[1,3,4], J. Novembre[5,], B. Pasaniuc[3,6].* 1) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA; 2) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA; 3) Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA; 4) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA; 5) Department of Human Genetics, University of Chicago; 6) Department of Pathology and Laboratory Medicine, Geffen School of Medicine at University of California, Los Angeles, Los Angeles, CA.

Ancestry analysis plays an important role in studies of human disease and evolution. Traditionally, ancestry inference from genetic data has been focused on modeling populations as discrete sources. As a result, ancestry prediction at each locus in the genome of any given individual can only be performed in terms of discrete population assignments or ancestry proportions. We have recently proposed methods for ancestry inference in a geographic continuum, by explicitly modeling the spatial distribution of each allele at any locus in the genome as a continuous function of geographic space (Yang et al Nat Genet 2012, Baran et al AJHG 2013). A major drawback of these approaches is that they only model individuals as having ancestry from one or at most two geographical locations (whereas, for example, 20% of the individuals collected in the POPRES data have recent ancestors from at least two country origins). In this work we propose approaches for continuous localization from genetic data for individuals that have recent ancestors from multiple geographical locations. That is, our methods seek to localize on a continuous map the ancestry of the recent ancestors (parents, grandparents and in general all ancestors up to several generations) using the genetic data of a given individual. In addition, we leverage the inheritance pattern to assign geographical locations for each segment in the genome according to the location of its ancestors. We devise an efficient Expectation Maximization approach within hidden Markov models of ancestry in conjunction with forward-backward algorithm that assigns continuous local ancestry for a given individual efficiently (e.g. an average of 2 minutes per individual). We use real data from POPRES as well as admixed individuals of Latino ancestry to show that our method yields high accuracy in ancestor localization as well as in assigning continuous ancestry at each locus in the genome. In simulations starting from POPRES data, our method accurately infers the locations of the 2 grandparents of a given individual's phased haplotype within an average of around 400Km of their true location and assigns continuous local ancestry within around 400Km of the true location at each segment in the genome. Software package implementing our methods is freely available in our website http://genetics.cs.ucla.edu/spa.

2077F

Separation of the largest eigenvalues in eigenanalysis of genotype data from discrete subpopulations. *K. Bryc[1], W. Bryc[2], J.W. Silverstein[3].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Department of Mathematical Sciences, University of Cincinnati, Cinciinati, OH; 3) Department of Mathematics, North Carolina State University, Raleigh, NC.

Principal component analysis (PCA) has been a powerful and efficient method for analyzing large datasets in population genetics since its early applications by Cavalli-Sforza and others. In particular, PCA of single nucleotide polymorphism genotype data can be used to illuminate population structure. We present a mathematical model, and the corresponding mathematical analysis, that justifies and quantifies the use of principal component analysis of biallelic genetic marker data for a set of individuals to detect the number of subpopulations represented in the data. A good estimate for the number of populations, K, is needed in Bayesian clustering algorithms such as STRUCTURE (Falush et al. 2003) or ADMIXTURE (Alexander et al. 2009), where one must specify a priori the number of clusters in the data, which affects the inferred relationships among individuals. In contrast to previous work, our results describe behavior of the eigenvalues of the sample covariance matrix without centering or normalization, taking into account both the number of individuals and the number of markers. The raw unprocessed covariance matrix is more amenable to mathematical analysis, and the singular values of such raw data exhibit quantifiable properties that can be used directly to determine the number of populations in the data in an almost deterministic fashion, at least when the number of individuals in the study is sufficiently large. We show that for large data sets of individuals from K well-differentiated subpopulations, with overwhelming probability the un-centered sample covariance matrix has K large eigenvalues. We demonstrate in two proof-of-principle simulations that we are able to obtain evidence of population structure when the number of individuals is large enough. We indicate that the power of the technique relies more on the number of individuals genotyped than on the number of markers.

2078W

Detection of Identity-by-Descent in Whole-Exome Sequencing Data. *W. Fu, M.J. Bamshad, D.A. Nickerson, J.M. Akey, NHLBI Exome Sequencing Project.* Department of Genome Sciences, University of Washington, Seattle, WA.

The distribution of genomic regions that are shared Identical by Descent (IBD) among individuals can provide insight into population history, facilitate the identification of adaptively evolving loci, and is an important tool in disease gene mapping. Next-generation sequencing technology has enabled the creation of large exome sequencing datasets from thousands of individuals. However, accurately inferring IBD segments from exome data is difficult because of the sparsity of the data. Here, we describe an accurate and robust computational framework to detect IBD in exome data, which we evaluated through comprehensive simulations. The key insight of our approach is to identify and exclude exomic loci that are refractory to accurate IBD calling because of insufficient exon density. We applied this computational framework to detect IBD segments in high-coverage exome sequences from 4,298 European-American individuals studied as part of the NHLBI Exome Sequencing Project. Patterns of IBD sharing reveal significant evidence for cryptic and spatial population structure, and allow us to infer the average number of meiosis that separate two randomly selected European-American individuals. Overall, our results enable the power of IBD analyses to be applied to exome data and reveal novel insights into the genetic structure of European-Americans in the United States.

2079T

The timing and history of Neandertal gene flow into modern humans. *S. Sankararaman[1,2], N. Patterson[2], S. Pääbo[3], D. Reich[1,2].* 1) Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Max Planck Institute for Evolutionary Anthropology, Leipzig.

Previous analyses of modern human variation in conjunction with the Neandertal genome have revealed that Neandertals contributed 1-4% of the genes of non-Africans with the time of last gene flow dated to 37,000-86,000 years before present. Nevertheless, many aspects of the joint demographic history of modern humans and Neandertals are unclear. We present multiple analyses that reveal details of the early history of modern humans since their dispersal out of Africa.

1.We analyze the difference between two allele frequency spectra in non-Africans: the spectrum conditioned on Neandertals carrying a derived allele while Denisovans carry the ancestral allele and the spectrum conditioned on Denisovans carrying a derived allele while Neandertals carry the ancestral allele. This difference spectrum allows us to study the drift since Neandertal gene flow under a simple model of neutral evolution in a panmictic population even when other details of the history before gene flow are unknown. Applying this procedure to the genotypes called in the 1000 Genomes Project data, we estimate the drift since admixture in Europeans of about 0.065 and about 0.105 in East Asians. These estimates are quite close to those in the European and East Asian populations since they diverged, implying that the Neandertal gene flow occurred close to the time of split of the ancestral populations.

2.Assuming only one Neandertal gene flow event in the common ancestry of Europeans and East Asians, we estimate the drift since gene flow in the common ancestral population. We show that an upper bound on this shared drift is 0.018. Because this is far less than the drift associated with the out-of-Africa bottleneck of all non-African populations, this shows that the Neandertal gene flow occurred after the out-of-Africa bottleneck.

3.We use the genetic drift shared between Europeans and East Asians, in conjunction with the observation of large regions deficient in Neandertal ancestry obtained from a map of Neandertal ancestry in Eurasians, to estimate the number of generations and effective population size in the period immediately after gene flow. These analyses suggest that only a few dozen Neandertals may have contributed to the majority of Neandertal ancestry in non-Africans today.

2080F

A method for computing the exact distribution of the genealogical history of a sample derived from a structured population. *M. Uyenoyama, S. Kumagai.* Biology, Duke Univ, Durham, NC.

Likelihood-based inference methods designed to analyze variation in genes sampled from structured populations require a means of accommodating changes in rates of evolutionary processes, particularly coalescence and migration. Population structure increases the dimensions of the genealogy of a sample, and migration modifies the subsets of lineages among which coalescence can occur. Some inference methods require detailed simulation of migration events to accommodate such changes. Here, we present a simple method for determining the probability of a gene genealogy under a specified model. This probability requires specification of the probability of all possible states of ancestral lineages corresponding to the nodes in a gene genealogy and the density of the time interval between nodes. We describe the qualitative features of the distribution of gene genealogies, including factors that influence the geographical location of the most recent common ancestor and the nature of the distribution of internode lengths.

2081W

Rare Variant Stratification in small geographic areas. *S. Zöllner[1,2], M. Mueller-Nurasyid[3,4,5], M. Zawistowski[1], M. Reppell[1], J. Novembre[6], K. Wolf[7], A. Peters[7], H. Grallert[8], K. Strauch[4,5].* 1) Dept Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Dept Psychiatry, Univ Michigan, Ann Arbor, MI; 3) Dept of Medicine I, LMU, Munich; 4) Chair of Genetic Epidemiology, LMU, Munich; 5) Institute of Genetic Epidemiology, Helmholtz Zentrum München Germany; 6) Dept Human Genetics, Univ Chicago, IL; 7) Institute of Epidemiology II, Helmholtz Zentrum München Germany; 8) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München Germany.

The geographic distribution of rare variants provides important insights into human population history and evolutionary forces affecting our genome. Moreover, differences in rare variant counts may affect population-based association studies of rare variants; accumulation tests that jointly analyze multiple rare variants are vulnerable to confounding by very subtle population stratification. Population genetic theory suggests that rare variants are affected by small differences in migratory patterns, such as the differences between urban and rural human populations. However, little data is available to assess the degree of population differentiation in small geographic areas. To explore this question, we analyze samples from the KORA study collecting ~2400 individuals of German ancestry from the German city of Augsburg and 16 surrounding sampling locations. The sampling locations are towns and villages with census sizes between 3.000 and 30.000 located in a geographic region of ~ 1,200 sqkm. We evaluate the genetic differentiation between sampling locations using allele sharing, the normalized probability that two individuals sharing an allele come from a different sampling location. We show that for common and uncommon variants (minor allele frequency >0.005), allele sharing in the study area is near 1, indicating that genotype frequencies do not differ between sampling locations. However, for rare variants, some of the same sampling locations differ from all other sampling locations and the level of differentiation does not depend on the geographic distance between sampling locations. Instead, the average allele sharing is clearly correlated with the census size of a sampling location; locations with small census size are more genetically differentiated than locations with large census size. The degree of population differentiation observed in this sample is predicted to be sufficient to generate inflated Type I error rate in accumulation tests. Tests that model causal and protective variants are especially susceptible to this type of population differentiation. Moreover, we explore the ability of typical correction strategies such as PCA and case-control matching to correct for the observed stratification.

2082T

Response of carrier frequency to a population bottleneck can quantify the amount of recessive variation. *D.J. Balick[1,2], R. Do[2,3], D.E. Reich[2,4], S.R. Sunyaev[1,2].* 1) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115; 2) Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142; 3) 2 Center for Human Genetics Research, Massachusetts General Hospital, Boston, Massachusetts, 02114; 4) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115.

Given the rich demographic history of humans, understanding the population dynamics in light of these differences has often been a confounding factor in identifying human evolutionary parameters, such as the strength and mode of selection. Here we present a rare example in which we can exploit disparate population histories to extract a detailed understanding of crucial evolutionary parameters. In the presence of a recent bottleneck and subsequent re-expansion, the founder's effect responds differently to distinct modes of selection, for example equilibrating more quickly from an increased additive burden than a comparable recessive burden. The genetic variation in the pre-bottleneck equilibrated population, as measured by the average number of deleterious alleles per individual, is exceeded by the bottlenecked population when all deleterious mutations act additively, but exhibits a relative depletion under recessive selection. As a result, one can extract the dominant form of selection by comparing an easily measurable feature of a population. We introduce a methodology for probing the fraction of polymorphic alleles in a population that act recessively. We analyze the transient dynamics of the site frequency spectrum of a population in response to a bottleneck, and find a qualitative distinction between additive and recessively acting alleles. By comparing the average number of polymorphic mutations per individual subjected to a founder event to the same statistic in an equilibrium population, we are able to determine the fraction of the variation that acts recessively. The transient dynamics in response to a bottleneck are explored in detail, and a crossover is identified in the recessive fraction as a function of the dominance coefficient and the strength of selection. Under the simplest model with no additional epistasis, we determine bounds on this ratio. This provides a simple statistical test for the presence of recessive variation on the whole genome level, or among a collection of functionally or medically relevant loci, and can be used to parse the mode of selection among existing regional variation on the local level.

2083F

Whole genome sequencing informs genealogy and the search for genetic modifiers in a PSEN1 E280A early-onset familial Alzheimer's disease cohort. *H.C. Cox[1], M. Lalli[2, 3], D.E. Mauldin[1], H. Li[1], M. Brunkow[1], M.L. Arcilia[2, 4], G. Garcia[4], L. Madrigal[5], S. Moreno[5], K.S. Kosik[2, 3, 4], L. Hood[1], J. Roach[1], G. Glusman[1], F. Lopera[5].* 1) Institute for Systems Biology, 401 Terry Ave N, Seattle WA 98109, USA; 2) Neuroscience Research Institute, University of California at Santa Barbara, CA, USA; 3) Program in Biomolecular Science and Engineering, University of California at Santa Barbara, Santa Barbara, California; 4) University of California at Santa Barbara, Department of Molecular, Cellular, and Developmental Biology, Santa Barbara, California; 5) Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellín, Colombia.

Mutations in the presenilin-1 gene (PSEN1) can promote progressive cerebral deposition of APP, resulting in Alzheimer's disease (AD). In particular, carriers of the 'Pasia mutation', a fully penetrant and dominantly inherited amino acid change (p.Glu280Ala) in PSEN1, suffer early-onset familial AD. This mutation is highly prevalent in Antioquia, Colombia, where over 1000 carriers have been identified. The mean age of onset of mild cognitive decline and dementia in carriers is 44 and 49 years, respectively. The variation in these traits implicates additional genetic factors influencing disease progression. This population provides a unique opportunity to identify genetic modifiers underlying AD. The origin of this mutation is believed to be a single early 17th century male founder during Spanish colonization of Colombia. A comprehensive genealogy with over 3000 descendents of this male founder has been previously described. The genealogy is characterized by geographic isolation, consanguinity and admixture of native Amerindian woman with Spanish Conquistadors and individuals of native African ancestry due to the Atlantic slave trade. We have obtained whole-genome sequences for 102 individuals [from this isolated population, including 74 carriers]. We are analyzing these using custom workflows and the Ingenuity Variant Analysis platform (www.ingenuity.com/variants). We computed pairwise relationships among all sequenced individuals. We determined mitochondrial haplotypes for all individuals and Y-chromosome haplotypes for all males. We observed mitochondrial haplogroups A, B, C, L and T, which are frequent in South and Central America, Africa and Eastern Europe. We used these lines of evidence to reconstruct the extended genealogy and characterize ancestry. This study provides a unique opportunity to investigate the etiology of AD in one of the largest known AD kindreds. The reconstruction of the founder genealogy provides a comprehensive and powerful resource for whole genome sequence analysis that will inform the search for modifiers of early-onset AD. [Funding: Université du Luxembourg, wwwen.uni.lu/lcsb].

2084W

Integration of low coverage whole genome sequence from the Kuusamo Finnish isolate with 1000 Genomes Project data provides an improved population reference panel. *AP. Sarin[1,2], K. Palin[3], I. Zara[4], K. Rehnstrom[1,4], M. Perola[1,2,5], V. Salomaa[6], A. Palotie[1,4,7], S. Ripatti[1,2,4,8], R. Durbin[4].* 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 3) Genome-Scale Biology Program, University of Helsinki, 00014 Helsinki, Finland; 4) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 5) Estonian Genome Center, University of Tartu, Estonia; 6) Department of Chronic Disease Prevention, National Institute for Health and Welfare, Finland; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 8) Hjelt Institute, University of Helsinki, Helsinki, Finland.

Understanding the genetic variation in population isolates is a valuable resource for genetic studies aiming to map rare and common variation associated to complex traits and common diseases. The reduced genetic complexity in such populations leads to better genotype imputation accuracy when using a population-specific reference panel and the presence of variants rare in the general population and pushed to higher frequency by genetic drift in the isolate offers the opportunity to better represent those variants in reference panels. With the aim of evaluating the potential for imputation of including samples from an isolated sub-population in a more general population reference panel, we sequenced the whole genome of 400 healthy individuals and the whole exome of 24 additional individuals from Kuusamo, a population sub-isolate founded roughly 350 years ago in northeastern Finland. The Kuusamo population was settled by 34 families in 1680s, and reached the present-day populations size of ~20,000 individuals without significant immigration until recently. Our simulations suggested that 400 individuals cover over 97% of all chromosomes in Kuusamo. Whole genome sequencing (WGS) was done at 4.8x on average with 100bp paired end protocol with the Illumina platform, and whole exome sequencing was done at ~37x coverage on average with the Agilent Sure Select Enrichment Kit at the Wellcome Trust Sanger Institute. Of 9,719 coding variants found in the median exome sequence, 9,595 (98.5%) are present in 1000 Genomes phase 1 (1000GP), and 98 (1.0%) more in the 400 WGS call set. Fst between Kuusamo and Finnish and between Kuusamo and Europeans from 1000GP are 0.006 and 0.013 respectively. We also evaluated the proportion of IBD segments shared across individuals and the percentage of individuals with at least one surrogate parent, as described in Kong at. al, Nature Genetics 2008, using the Systematic Long Range Phasing (SLRP) software (Palin et al, Genetic Epidemiology 2011). We evaluated imputation quality of 41 additional Kuusamo samples and 40 Helsinki samples using IMPUTE2. In general imputation is better into the Kuusamo samples than the Helsinki samples. Moreover, in Kuusamo samples a combined reference panel with both Kuusamo and 1000GP haplotypes provides the best performance. Our study demonstrates the value of combining outbred, isolated and sub-isolated population data to increase the amount of variability captured through imputation.

2085T

Genealogical evidence of allele frequency shuffling explains regional genetic structure in a founder population. *C. Bherer[1], L. Excoffier[2,3], M.H. Roy-Gagnon[4], J. Hussin[1], H. Vézina[5], D. Labuda[1,6].* 1) Centre de recherche du Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Québec, Canada; 2) CMPG, Institute of Ecology and Evolution, University of Bern 6, Baltzerstrasse, CH-3012 Bern, Switzerland; 3) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 4) Department of Epidemiology and Community Medecine, University of Ottawa, Ontario, Canada; 5) BALSAC Project, Université du Québec à Chicoutimi, Québec, Canada; 6) Department of Pediatrics, Université de Montréal, Québec, Canada.

Founding events promote allele frequency changes, and even deleterious alleles may rise in frequency by chance. Founder effects may thus account for the greater prevalence of some rare monogenic diseases, like among the French Canadians of Quebec, who are a prime example of founder population. Here we ask whether the demographic history of Quebec is responsible for the increase in frequency of rare disease-causing variants, and if this demography has affected the site frequency spectrum (SFS). By simulating the Mendelian transmission of alleles in the Quebec population genealogy, we studied allele frequency changes following the initial founding settlement and its subsequent regional spatial expansion. We used ascending genealogies of 2,221 individuals representing eight Quebec regions. The total genealogy includes 153,447 ancestors linked to 8,834 founders. Allele dropping (AD) simulations, starting from an equilibrium SFS among founders, led to significant loss for alleles below 1%, but no changes for the larger frequency classes. Using Tajima's D equilibrium test, a significant deficit of rare alleles was found in the whole Quebec sample, in the North-Eastern and Eastern regions, whereas an excess of rare alleles was observed in the western regions, suggesting that regional demographic histories had opposite effects. FST analysis showed significant genetic structuring of the regional populations. We further quantified frequency changes for different initial allele frequencies among founders. For both rare and common alleles, the variance in frequency changes was significantly different between regions, demonstrating differential shuffling effect of allele frequencies among regional populations. We observed extensive variance in genetic contribution among founders. In the North-Western, North-Eastern and Eastern regions, we find that a unique allele carried by a single founder could have increased in frequency up to 5% until now, thus potentially explaining the clinical founder effect of specific Mendelian disorders observed in the North-East. Our results demonstrate that regional demographic histories resulted in regional founder effects and genetic structure over a very short evolutionary time.

2086F

Genetic characterisation of two Greek population isolates. *K. Hatzikotoulas[1], K. Panoutsopoulou[1], I. Tachmazidou[1], L. Southam[1,3], D. Xifara[3,4], L. Moutsianas[3], A. Farmaki[2], A. Matchan[1], N.W. Rayner[1,3], Ch. Kiagiadaki[5], E. Tsafantakis[5], I. Ntalla[2,7], M. Karaleftheri[6], G. Dedoussis[2], E. Zeggini[1].* 1) Human Gentetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Harokopio University Athens, Athens, Greece; 3) Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 4) Department of Statistics, Oxford, United Kingdom; 5) Anogia Medical Centre, Anogia, Greece; 6) Echinos Medical Center, Xanthi, Greece; 7) Department of Health Sciences, University of Leicester, Leicester, United Kingdom.

Genetic association studies of low-frequency and rare variants can be empowered by focusing on isolated populations. It is important to genetically characterize population isolates for substructure and recent admixture events as these may give rise to spurious associations. Under the auspices of the HELlenic Isolated Cohorts study (HELIC; www.helic.org) we have collected >3,000 samples from two isolated populations in Greece: the Pomak villages (HELIC Pomak), a set of religiously-isolated mountainous villages in the North of Greece; and Anogia and surrounding mountainous villages on Crete (HELIC MANOLIS). All samples have information on anthropometric, cardiometabolic, biochemical, haematological and diet-related traits. 1,500 individuals from each population isolate have been typed on the Illumina OmniExpress and Human Exome Beadchip platforms. Multidimensional scaling analysis with the 1000 Genomes Project data shows similarities of the two population isolates with Mediterranean populations such as the Tuscans from Italy and Iberians from Spain. We also observe evidence for structure within the isolates, with the Kentavros village in the Pomak strand demonstrating high levels of differentiation. To characterise the degree of isolatedness in these populations we estimated the proportion of individuals with at least one 'surrogate parent' (using only the subset of samples with pairwise pi-hat<0.2) and compared this to an outbred Greek population from the TEENAGE study, which comprises 707 unrelated adolescents from the Attica district. We find that for random regions in the genome the proportion of individuals with at least one surrogate parent in the MANOLIS isolate is >60% and in the Pomak isolate is >65% compared to ~1% in the outbred Greek population. Our results establish these populations as isolates and provide some insights into the genomic architecture of Greek populations, which have not been previously characterised.

2087W

Risks for Mendelian Disorders in the Bronx. *K. Upadhyay[1], H. Ostrer[1,2], C. Oddoux[2].* 1) Pathology, Albert Einstein College of Medicine, Bronx, NY; 2) Pathology, Montefiore Medical Center, Bronx, NY.

The Bronx, a borough of 1.4 million people, is the third most genetically diverse county in the United States. Of the current residents, 51% are Hispanic-Latino (of whom Puerto Ricans and Dominicans are the largest ethnic groups) and 35% are African-American. To understand the risks of Mendelian disorders among these populations, Affymetrix Axiom Exome 319K arrays were analyzed for 192 members of each group, followed by variant calling using Affymetrix Genotyping Console software. Following quality control, calls were made for 305,519 probes. The called probes were mapped for annotated variants by cross-referencing to Online Mendelian Inheritance in Man, resulting in 1440 known alleles, of which 818 were polymorphic and 338 were pathogenic for Mendelian disorders. These alleles were also cross referenced against 14 other World populations, including an independent set of Puerto Ricans (demonstrated to overlap with Bronx Puerto Ricans by PCA analysis) Among these alleles, 157 were shared among all 3 populations, 32 were shared by Puerto Ricans and Dominicans, 22 were shared by Puerto Ricans and African Americans and 12 were shared by Dominicans and African Americans. In singleton populations, 37 were found only in Puerto Ricans, 45 were found only in Dominicans and 30 were found only African Americans. Among the conditions previously described in these populations were prostate cancer susceptibility and progression (EPHB2 p.A279S), hemochromatosis (HFE p.C282Y), and hereditary amyloidosis (TTR p.I112V). Among the conditions with allele frequencies ≥1% in at least one population were cystic fibrosis (CFTR p.D1270N), spastic paraplegia 44 (GJC2 p.I36M), susceptibility to diffuse gastric cancer (CDH1 p.A617T), Usher syndrome 1C (USH1C p.V130I), cone-rod dystrophy 3 (ABCA4 p.L1201R), fructose intolerance (ALDOB p.A150P), Fuchs endothelial 6 corneal dystrophy (ZEB1 p.G841P) and susceptibility to intestinal carcinoid tumors (SDHD p.H50R). Thus, these studies provide insights into previously unrecognized mutations for Mendelian disorders among the populations of the Bronx.

2088T

***MAK* gene Alu element insertion: high carrier frequency among Ashkenazi Jews most likely due to founder effect.** *L. Shi, M. Luo, L. Edelmann, R. Kornreich.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Background and Purpose: Retinitis pigmentosa (RP) is a group of highly heterogeneous inherited disorders characterized by abnormalities of the photoreceptors of the retina which lead to progressive visual loss. It is estimated to have an incidence of 1 in 3,500 to 1 in 4,000 in the United States and Europe. Whole exome sequencing identified *male germ cell-associated kinase* (*MAK*) (OMIM 154235) as a novel RP disease-cause gene (RP62, OMIM 614181) (Turker *et al. PNAS* 2011). A recurrent homozygous Alu insertion in *MAK* exon9 was identified in 21 unrelated Jewish RP patients, which accounted for approximately 1.2% of 1,798 unrelated RP individuals (Ozgul *et al. Am J Hum Genet* 2011). Population screening of this Alu insertion among Ashkenazi Jewish (AJ) individuals revealed a carrier frequency of 1/55 (Stone *et al. IOVS* 2011). The purpose of this study was to further investigate the carrier frequency of the *MAK* Alu element insertion in the New York metropolitan AJ population and perform haplotype analysis to see if this mutation exhibited a founder effect. Methods: We screened 1,494 unrelated AJ individuals for the Alu insertion by polymerase chain reaction (PCR) amplification of *MAK* gene exon9 and agarose gel electrophoresis. All positive samples identified were confirmed by Sanger sequencing from both directions. Genetic markers flanking the *MAK* gene were analyzed by capillary electrophoresis and haplotype construction was performed by PHASE software. Results: Among our 1,494 AJ samples, 52 showed an extra PCR band of increased size, suggesting a population carrier frequency of 1 in 29 (95% confidence interval 1 in 44 to 1 in 21). Sequencing this extra band identified an identical insertion sequence of an Alu element plus a 13-bp genomic sequence duplication in exon 9 of *MAK* gene among all 52 individuals. Flanking markers revealed a shared haplotype in over 85% of the carriers suggesting a founder effect. Conclusion: *MAK* gene exon 9 Alu element insertion has a carrier frequency of 1 in 29 among Ashkenazi Jews from the NY metropolitan area, most likely from a founder effect. Genetic screening may be appropriate for this mutation in the AJ population due to its high prevalence, however as this form of RP is adult onset with slow progression, it may not be appropriate for prenatal screening.

2089F

Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans.

A. Hodgkinson[1], F. Casals[1], J. Hussin[1], Y. Idaghdour[1], V. Bruat[1], T. de Malliard[1], JC. Grenier[1], E. Gbeha[1], F. Hamdan[1], S. Girard[2], JF. Spinella[1], M. Lariviere[1], V. Saillour[1], J. Healy[1], I. Fernandez[1], D. Sinnett[1], J. Michaud[1], G. Rouleau[1], E. Haddad[1], F. Le Deist[1], P. Awadalla[1]. 1) Sainte-Justine Research Centre, University of Montreal, Montreal, Quebec, Canada; 2) Centre d'Excellence en Neuromique de l'Université de Montréal, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada.

Whole-exome or gene targeted resequencing in hundreds to thousands of individuals has shown that the majority of genetic variants are at low frequency in human populations. Rare variants are enriched for functional mutations and are expected to explain an important fraction of the genetic etiology of human disease, therefore having a potential medical interest. In this work, we analyze the whole-exome sequences of French-Canadian individuals, a founder population with a unique demographic history that includes an original population bottleneck less than 20 generations ago, followed by a demographic explosion, and the whole exomes of French individuals sampled from France. We show that in less than 20 generations of genetic isolation from the French population, the genetic pool of French-Canadians shows reduced levels of diversity, higher homozygosity, and an excess of rare variants with low variant sharing with Europeans. Furthermore, the French Canadian population has accumulated a large number of putative damaging functional variants, that may impact upon the fitness of the population and could partially explain the increased incidence of genetic disease in the province. Our results highlight the impact of population demography on genetic fitness and the contribution of rare variants to the human genetic variation landscape, emphasizing the need for deep cataloguing of genetic variants by resequencing worldwide human populations in order to truly assess disease risk.

2090W

Whole exome sequencing in families with rare perinatal onset immunodysregulatory diseases present with fever and systemic inflammation.

Z. Deng, Y. Liu, A. Almeida de Jesus, H. Sun, M. Gadina, R. Goldbach-Mansky. Intramural Research Program, NIAMS, National Institute of Health, Bethesda, MD.

WES (Whole Exome Sequencing) has increasingly become the tool of choice in translational research, providing molecular diagnosis in Mendelian diseases and identifying important genes in key biological pathways. Here we report using WES to investigate the molecular basis of a group of rare immunodysregulatory diseases that are characterized by perinatal onset fever and systemic inflammation. We have developed a bioinformatics pipeline to process WES data and an integrated workflow to analyze variants in family trios or quartets. Using the pipeline, we were able to assess the quality of WES data and identify discrepancies in sample gender and family relatedness. The number of coding variants per sample in our WES studies ranged from 19,000 to 25,000, correlating with exome coverage and sample ethnicity. The Transition to Transversion ratio (Ti/Tv) varied from 2.98 to 3.27 with a median of 3.15 after excluding a poorly performed batch. Other QC metrics such as heterozygous to homozygous ratio, synonymous to nonsynonymous ratio and indel percentage were all within expected ranges for WES. Amongst 19 parent-child trios, the Mendelian Inheritance Error (MIE) rates varied from 0.68% to 3.71% with a median of 1.69%. Concordance in two pairs of technical replicates and one pair of identical twins were from 97.25% to 97.82%. On average, the novel variants (defined as not present in dbSNP137) accounted for about 2% of all variants in each sample. The variant annotation, analysis and filtering workflow has allowed us to successfully identify *de novo* mutations in 5 trios. All of the mutations were confirmed by Sanger sequencing, along with a number of other variants in biologically interesting genes. In two of the trios, the mutations were located in genes (PSMB8, NLRP3) known to be responsible for the diseases (CANDLE, NOMID). Functional studies are currently underway in the other families to confirm the pathogenicity of the mutations in these families. Our results suggest that WES can be used as an effective tool in translational research of rare immunodysregulatory diseases.

2091T

The impact of population demography and selection on the genetic architecture of complex traits.

K.E. Lohmueller. Ecology and Evolutionary Biology, University of California, Los Angeles, CA.

The medical genetics community is moving into an era of using exome sequencing to identify low-frequency (<1%) coding region variants associated with complex traits. At the design stage of such experiments, it is critical to understand how many and which individuals to sequence. One important factor that has not been explored fully in previous studies is the demographic history of the sequenced population. Recent studies of thousands of individuals have found evidence of extreme recent population growth in several populations. Here I investigate the effect of this growth on the power to identify low-frequency variants associated with complex disease. In particular, I perform forward simulations using realistic demographic models for different human populations that take into account various levels of recent population growth. The simulations also include a distribution of selective effects for deleterious mutations. Phenotypes are simulated using a liability threshold model where a mutation's effect on the phenotype is proportional to its effect on fitness. First, my simulations show that recent population growth leads to a proportional increase in deleterious amino acid changing polymorphisms, especially if the growth occurred so recently that purifying selection has not had sufficient time to remove the deleterious mutations. Second, if a mutation's effect on fitness is correlated with its effect on the trait, then rare variants account for more of the additive genetic variance of the trait in a recently expanded population than in a population that did not expand. Third, in line with the previous result, I show that recent population growth can increase the genetic heterogeneity of disease. Specifically, a sample of diseased individuals from a population that has undergone recent growth is expected to contain more distinct causal mutations than a sample of the same number of cases from a population that has not expanded. If the causal mutations are scattered across many distinct genes, this increased heterogeneity due to population growth will reduce the power of commonly used gene-based rare variant association tests. These findings suggest that careful consideration of recent population history is essential for designing optimal medical sequencing studies.

2092F

Regional rare allele sharing in a homogeneous, geographically clustered population.

S.L. Pulit[1,2], A. Menelaou[1], C.C. Elbers[1], L.C. Francioli[1], P.I.W. de Bakker[1,2], The Genome of the Netherlands Consortium. 1) University Medical Center Utrecht, Utrecht, The Netherlands; 2) Medical and Population Genetics Program, Broad Institute, Boston, MA, USA.

Next-generation sequencing has yielded unprecedented cataloguing of variants across the frequency spectrum. Although population structure in common variation (CV) is well studied, the exploration of population structure due to rare variants (RVs) has only just begun. Some studies suggest that RV structure is distinct from that of CV while others have found that it can be captured by principal components (PCs) calculated using common SNPs. A broader understanding of population structure is needed as we increasingly focus on the role of RVs in complex traits.To elucidate the architecture of RVs, we used Genome of the Netherlands (GoNL) data of 250 Dutch trios sequenced at 14x coverage. We specifically analyzed f_2 variants, mutations appearing exactly twice in GoNL (N=330 unrelateds evenly sampled from 11 provinces). Since f_2s are rare and likely to be newer mutations, they can illuminate demographic history. We annotated f_2 allele carriers and their province of birth and found that within-province sharing was strongest; on average, both f_2 alleles were 1.46x more likely to be found in the same province. After dividing the Netherlands into northern, central, and southern regions, we tested if within-region sharing of f_2s was stronger in the north or south, areas with historically less migration, compared to the central provinces, areas with more migratory populations, the country's largest city (Amsterdam) and biggest seaport (Rotterdam). We observed strong sharing in the northern vs. central regions ($p<10^{-200}$), but no excess sharing for the southern vs. central regions (p=0.11). These observations are consistent with Dutch history: for centuries, flooding in the northern regions prevented migration into or out of the north and populations primarily settled in the more prosperous central regions.We also explored to what extent CV and RV structure overlap by projecting the samples onto CV PCs. While PC1 captures a North-South cline, the northern samples (with much stronger f_2 sharing) are no more tightly clustered in PC space (mean distance between samples=0.072) than the central (mean dist=0.065, p=0.93) or southern samples (mean dist=0.069, p=0.99). Our results suggest that RV structure, even in a homogeneous population from a small geographic area (41,000 km2), is not fully captured by PCs, which has important implications for RV disease studies such as those employing the exome chip.

2093W

Estimation of Migration Rates and Patterns Based on Distributions of Rare Variants.
R. Rothwell[1], M. Ehm[2], P. St. Jean[2], M. Nelson[2], J. Novembre[3], S. Zöllner[1]. 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Quantitative Sciences, GlaxoSmithKline, RTP, NC; 3) Human Genetics, University of Chicago, Chicago, IL.

Careful modeling of the relationships between populations is essential to our understanding of population diversity. Most current models, however, ignore historical changes in migration rate, modeling gene flow as a constant parameter. The growing availability of sequencing data allows estimating temporal changes in these relationships. Identifying such changes may quantify the impact of historical events on human diversity. In this study, we develop a method for using rare variants to identify changing migration patterns. The distribution of rare alleles among multiple populations depends on the migration rate between those populations in the historic time since the mutation event that generated the variant. Rare variants that arose recently are only affected by recent migration events and thus allow an estimate of the migration rate in the recent past. Common variants that arose further in the past allow an estimate of the average migration rate over the time since they arose. Using this intuition, we develop a likelihood function based on the distribution of alleles, where we obtain one estimate of the migration rate each from all variants with a given minor allele count. By comparing different estimates of migration across variants with different minor allele count, we generate a temporal outline of migration rate over time. We evaluate the performance of our method on simulated data of populations with changing migration patterns. For changes as recent as 20 generations in the past, our algorithm identifies a clear temporal pattern. We also show that this methodology can be adjusted to model population characteristics including changes in effective population size and is robust to minor model misspecifications. We apply this methodology to sequence data from the exons and flanking regions of 202 target genes (2220 exons total), generated from European and African American samples. In the African American population, we observe sharply increased migration in recent years from European populations, reflecting a pattern consistent with the transatlantic slave trade. In the European populations, we find generally high gene flow with slightly lower migration rates from the more geographically distant populations. In the Nordic-Baltic regions, we observe temporal trends indicating widespread Scandinavian expansion across Europe and subsequent reduction in migration. In Central and Western Europe, we find increasing European movement in recent years.

2094T

Exome sequencing revealing unique genetic profile of Quebec Nunavik Inuit population.
S. Zhou[1], L. Xiong[1,2], A. Ambalavanan[1,3], A. Dionne-Laporte[1], D. Spiegelman[1], E. Henrion[1], O. Diallo[1], C. Bourassa[1], N. Dupré[4], M.-P. Dubé[5], P. Dion[1], G. Rouleau[1,6]. 1) Centre of Excellence in Neuroscience of Université de Montréal(CENUM), CHUM Research Center and the Department of Medicine, Montréal; 2) Centre de recherche Fernand-Seguin, HÔpital Louis H. Lafontaine, Montréal; 3) Department of Human Genetics, McGill University; 4) Department of Neurology, Université Laval, Québec; 5) Pharmacogenomics Centre, Montreal Heart Institute, Université de Montréal, Montréal; 6) Montreal Neurological Institute and Hospital, McGill University, Montréal.

Background: Nunavik comprises the northern part of Quebec, with 90% of its inhabitants being Inuit. The modern Inuit of Nunavik came from the Thule people which originated from coastal Alaska around 1000 AD and traveled eastward along the Arctic tree line, toward the Northwest Territories (NWT), Nunavut, Greenland, and reached Nunavik around 1500 AD as they became the ancestors of current Inuit residents. In our study, we chose to use a high throughput parallel sequencing approach to generate the specific genetic profile of the Inuit population for the first time. **Materials and Methods:** We performed exome sequencing on 101 Nunavik Inuit from 11 out of 15 villages in Nunavik, using Agilent SureSelect V4 capture kit and Illumina HiSeq platform. Standardized data processing is used to extract all exonic variants. **Results:** We performed preliminary analysis in order to identify Inuit specific novel or rare nonsynonymous variants. The analysis yielded 62 protein changing variants which have allele frequency over 0.5 in Inuit while less than 0.01 in other populations. Among these, a variant p.P479L in the *Carnitine Palmitoyltransferase 1A* (*CPT1A*) gene is the most prominent one. 88 individuals from our Inuit cohort carry the homozygote, and with the lack of homozygous wild-type, the p.P479L allele frequency in Nunavik reaches 0.94. This frequency appears to be the highest when compared to other Inuit populations: 0.43 in Alaska native newborns, 0.44 in NWT Inuit, 0.77 in Nunavaut and 0.73 in Greenland Inuit. Interestingly, the increasing frequency of this variant from Alaska to Greenland and the northeast of Canada seems to correlate with the migration pattern of Inuit over the course of a thousand years. **Discussion:** CPT1A is a key regulator of fatty acid metabolism, and the p.P479L allele is known to be only present in Inuit populations. This variant has been reported to be associated with impaired fasting tolerance in Alaska, and increased risk of sudden infant death (SIDS) in Nunavut. However, in Greenland Inuit the p.P479L allele is linked with higher levels of HDL-cholesterol and apoA-I, which are considered to be a protective allele against cardiovascular risk and infection. The interesting diverse phenotypic variation associated with this allele and its high frequency among older Nunavik adults (mean age 52) suggest the same variant may contribute to different physiological activities in Inuit during different life stages and from different environments.

2095F

Inference of population structure using sequence data. *G. Bhatia[1,2], A. Gusev[2,3,4], N. Zaitlen[5], B.J. Vilhjalmsson[2,3,4], D. Diogo[2,6,7], P.K. Gregersen[8], J. Worthington[9], L. Padyukov[10], S. Raychaudhuri[2,6,7,9], R.M. Plenge[2,6,7], B. Pasaniuc[11], A.L. Price[2,3,4].* 1) Harvard-MIT Division of Health Science and Technology, Boston, MA; 2) Medical and Population Genetics Program, Broad Institute, Cambridge, Massachusetts, United States of America; 3) Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America; 4) Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America; 5) Department of Medicine Lung Biology Center, University of California San Francisco, San Francisco, California, United States of America; 6) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; 7) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America; 8) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, United States of America; 9) Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK; 10) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 11) Department of Pathology and Laboratory Medicine, Geen School of Medicine at UCLA, Los Angeles, California, United States of America.

Application of standard principal components analysis (PCA) methods to sequence data does not perform well because of (1) pervasive linkage disequilibrium (LD) and (2) large numbers of rare variants. When applied to 0.78 Mb of targeted sequence data from 918 individuals of European ancestry (Nelson et al. 2012 Science), the top 5 PCs explained only a small proportion of the variance in true population labels ($r^2 = 0.023$). To improve the performance of ancestry inference, we examined the effectiveness of several methods to correct for LD and handle rare variants. To correct for LD, we considered LD pruning, LD shrinkage (reweighting by LD to surrounding SNPs) or LD regression (regressing out surrounding SNPs from each SNP). All of these methods improved results, with LD regression performing best. To handle rare variants, we considered exclusion of rare variants, exclusion of singletons only, or a rare variant reweighting approach (motivated by our theoretical derivations) that imposes an upper bound on the standard weight of 1/p(1-p). Rare variant reweighting and exclusion of singletons both performed well ($r^2 = 0.56$ when used in conjunction with LD regression), whereas the standard strategy of excluding rare variants was suboptimal ($r^2 = 0.44$). These results are consistent with results obtained from coalescent simulations, and suggest that rare variants can improve the performance of ancestry inference from sequence data. We also applied these techniques to ImmunoChip data from 186 loci and 2283 individuals of either Northern or Southern European ancestry (Eyre et al. 2012 Nat Genet). Running standard PCA on data with MAF < 0.01 variants removed, the top 5 PCs explained a moderate proportion of the variance in true population labels ($r^2 = 0.86$). Application of LD correction and inclusion of rare variants resulted in far greater accuracy ($r^2 = 0.99$), sufficient to correct for population stratification in association analysis. Overall, our results suggest that effective ancestry inference from targeted sequence data or targeted high-density genotyping data requires appropriate treatment of LD and rare variants. In particular, rare variants can improve ancestry inference, and should not be universally removed.

2096W

Population-specific variants in whole exome sequencing of American Indians: the Strong Heart Family Study. *S.A. Cole[1], P.E. Melton[2], K. Haack[1], C. Bizon[3], J.K. Sailsbery[3], K.C. Wilhelmsen[4], E.T. Lee[5], L.G. Best[6], L.A. Almasy[1], N. Franceschini[4].* 1) Dept Gen, Texas Biomedical Research Inst, San Antonio, TX; 2) Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley; 3) Renaissance Computing Institute, University of North Carolina, Chapel Hill, NC; 4) University of North Carolina, Chapel Hill, NC; 5) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 6) Missouri Breaks Research Industries, Inc, Timber Lake, SD.

Sequencing of the human exome has successfully identified novel variants in genes underlying monogenic disorders as well as complex diseases. Exome sequencing has shown that rare variants represent most of the genetic variation in human populations, and because they are of relatively recent origin, tend to be population-specific. Hence, disparities in disease prevalence are likely due to alleles that must be discovered though population-targeted sequencing. To characterize genetic variation in American Indians (AI) residing in the U.S., who are under-represented in 1000 Genome project (1000G) and Exome Sequencing Project (ESP) data, and who suffer disproportionately from diseases such as heart disease and type 2 diabetes, we performed whole exome sequencing of 94 AI samples selected from unrelated founders of large pedigrees of the Strong Heart Family Study Arizona and Oklahoma Centers. Sequence data were aligned to the Human Genome Reference version 37.1 (hg19). Variants were called and recalibrated, and QC metrics were calculated, using the Genome Analysis Toolkits (GATK Unified Genotyper version 2.1). Annotation was provided by SeattleSeq. For the 94 samples, an average of 97.2% of sequences mapped to the reference genome. A total of 250,320 autosomal, non-monomorphic variants passed QC, of which 226,887 were single nucleotide variants (SNVs) and 23,433 were indels. The mean call rate was 0.98, mean transition/transversion ratio (Ti/Tv) was 2.50 and mean variant read depth was 37x. Based on comparison with dbSNP, 1000G and ESP data, 45,602 variants were novel SNVs and 9,776 were novel indels. As expected, most of the novel SNVs were singletons (71%; n=32,350), but we observed several low frequency and common SNVs in exome and putative regulatory regions. Approximately 20% of novel SNVs have predicted functionality (n= 9,374, including frameshift, stop-gain/loss and missense/nonsense), and in 2,700 of those (28%) the minor allele was observed at least twice (minor allele count ≥2). The presence of population-specific exonic SNVs of functional potential may have significant biological implications as the novel variants may account for some of the increased disease burden present in American Indians.

2097T

Forward-in-time Simulation of Recombination, Quantitative Traits, and Selection. *D. Kessner[1], J. Novembre[2,3,1].* 1) Bioinformatics, UCLA, Los Angeles, CA; 2) Human Genetics, University of Chicago, Chicago, IL; 3) Ecology and Evolutionary Biology, UCLA, Los Angeles, CA.

Simulations have a long history in population genetics, both for verifying analytical results and for exploring population models that are mathematically intractable. In a typical forward-in-time simulation, individual fitness is parameterized by fixed per-genotype selection coefficients for a subset of loci in the simulation. While this approach is adequate for simulation scenarios involving simple traits where genetic loci have a constant effect on fitness, it is not flexible enough to handle more complicated scenarios involving polygenic quantitative traits. In addition, while individual haplotype sequences are typically output at the end of a simulation, information about the recombination events that occurred during the simulation has been lost. Recombination information is particularly useful when exploring the haplotype patterns that emerge in scenarios where substantial evolutionary change has taken place in a small number of generations due to recombination and/or selection -- for instance, during recent admixture between isolated populations, or artificial selection experiments. Motivated by such examples, we present a new forward-in-time simulation approach that, instead of tracking single-site variants, tracks individual haplotype chunks as they recombine over multiple generations. This design allows for fast simulation of chromosomal-scale regions, with very efficient memory usage, and also retains recombination information. The simulator uses a modular architecture to allow the user to specify recombination maps, mutation rates, demographic models, quantitative traits and fitness functions. With this modular approach, complicated scenarios are feasible to simulate, including as an example, selection for optimal values of a polygenic quantitative trait in multiple connected populations, where the optimal value may change depending on the population and generation. The simulation program (forqs) is implemented in C++ and is freely available.

2098F

A low frequency of copy number polymorphism for the *NPEPPS*, *POLR2J4*, and *PCDHB13* genes in a sample of admixed Brazilian population. *D. Jimenez[1], T. Lins[2], P. Taveira[1], R. Pereira[1,2].* 1) Universidade Catolica de Brasilia, Brasilia, DF, Brazil; 2) Patologia Molecular, Universidade de Brasília, Brasília, DF, Brasil.

Copy number variants (CNV) represent an important source of variation in the human genome, with some embedded genes differently distributed among human population groups. Therefore, it is necessary to understand the distribution of CNV within and between populations, especially those with admixed ancestry, such as the Brazilians. The aim of the study was to investigate the variability of a set of CNV-embedded genes in a sample of the Brazilian population. The CNV-embedded genes were chosen based on previous published data showing that they have differential copy variation distribution between population samples from Africa and Europe. Three genes (*NPEPPS*, *POLR2J4*, and *PCDHB13*) were investigated by real-time qPCR using TaqMan® assays (Applied Biosystems, Foster city, CA) in a sample of 96 Brazilian individuals previously classified by genetic ancestry informative markers. A low variability in the selected genes was identified, showing 2 copies in frequency of 0.927, 0.990 and 0.979, respectively. Mean distribution ± standard deviation of ancestry according to European, African and Native American proportion was, respectively, 0.702 ± 0.232, 0.189 ± 0.158 and 0.109 ± 0.147, for overall sample. Due to low variability, the genomic ancestry was correlated to copy number (diploid = 2 copies versus non-diploid) only in *NPEPPS* gene, but no correlation was observed (p= 0.257, 0.188 and 0.498, respectively for ancestry estimates). In conclusion, the results provided an overview of the corresponding frequency of gene copy number variation in a sample of the Brazilian population, serving as reference for further population studies, which may correlate these polymorphisms with other phenotypic differences. Nevertheless there is a need for upcoming research to revise these data by using more assays nearby the genes to search for specific breakpoints or use other specific methods, such as high-resolution technologies and second generation sequencing, to confirm and validate the copy number polymorphisms.

2099W

Amylase gene copy number polymorphism in ethnically admixed sample from Brazil. *T. Lins[1], D. Jimenez[2], P. Taveira[2], R. Pereira[1,2].* 1) Patologia Molecular, Universidade de Brasilia, Brasilia, DF, Brazil; 2) Ciencias Genomicas e Biotecnologia, Universidade Catolica de Brasilia, Brasilia, DF, Brazil.

Salivary amylase gene (*AMY1*) copy number variation (CNV) represent a significant variation among individuals and evolved independently in different human population groups worldwide. The *AMY1* genetic variation seems to be related with eating habits and dietary consumption of starch as the *AMY1* gene copy number, the concentration and activity of salivary amylase enzyme are positively correlated to the starch content diet (high or low starch diet). Therefore, it is necessary to understand the distribution of CNV within and between populations, especially those with admixed ancestry, such as the Brazilians. The aim of the study was to investigate the variability of the *AMY1* gene in a sample of 96 Brazilian individuals previously classified by genetic ancestry informative markers. The polymorphism was investigated by real-time qPCR using Taqman® assay (Applied Biosystems, Foster city, CA). The mean *AMY1* copy number was 2.8 with a range of 1 to 8 copies. The major copy number was identified for 2 copies (47.9%), followed by 3 copies (20.8%) and 4 copies (18.8%). Mean distribution ± standard deviation of ancestry according to European, African and Native American proportion was, respectively, 0.702 ± 0.232, 0.189 ± 0.158 and 0.109 ± 0.147, for overall sample. A T-test was performed to compare estimated genomic ancestry proportion among 2 copies and gain of copies (≥3 copies), but no statistical correlation was observed (p = 0.449; 0.494; 0.412, respectively). Despite the low copy number compared to other worldwide populations, the correlation of the polymorphisms with other phenotypic differences must be investigated, such as salivary amylase enzyme concentration and activity, as well as the nutritional status and starch content diet. Since the *AMY1* is highly variable, an increased number of individuals in the Brazilian population should be further evaluated to screen a wider range of variation. The results provided an overview of the corresponding frequency of *AMY1* gene copy number variation in a sample of the Brazilian population serving as reference for further population and nutrigenomic studies.

2100T

Mitochondrial Genome Database for Saudi Community. *I. Alabdulkareem[1], M. Albalwi[1,2], M. Alharbi[1], A. Alghamdi[1], B. Alhamad[1], M. Aljumah[1].* 1) KAIMRC, National Guard Health Affairs, Riyadh, Riyadh, Saudi Arabia; 2) King Abdulaziz Medical City for National Guard, Pathology Department, Riyadh, Saudi Arabia.

This study is aiming to create an electronic database for all Non Mendelian inheritance among Saudi populations that is will be available for all researchers and health providers at the country. Phase one of this investigations is focusing on the establishment of the mitochondria genome sequencing in healthy (mtDNA). mtDNA as extracellular components that do not subdue to Mendelian inheritance mode that is genetically considered as complete different in terms of evolutionary origin since many of the evaluation scientists thought that mtDNA may be derivative from the circular genomes of the bacteria that were engulfed by the early ancestors (endosymbiotic hypothesis). mtDNA was sequenced using the solid 5500™ and validate by Ion Personal Genome Machine® [PGM™] System. Studies showed that the Arab assembled mitochondrial genome contains 16,570 bp as compared to GenBank reference genome with 16,571 bp. Further analysis in mitochondrial genome detected 46 differences, with almost single point mutations and only a deletion of length one. Homology studies on 3,132 known mitochondrial sequences from NCBI with sequence length between 16,000 and 17,000 bp revealed more than 90% matches sequence to accession number EF184636 sequence. Reported single nucleotide polymorphisms (SNPs) are rs3927813 rs3928312 rs1599988 rs3021088 rs3929989, rs1978028 rs121434446, rs1981459, rs3902407, rs3915952, rs3926883, rs3899498 rs2835780, rs3135031, rs34799580. More subjects from healthy and diseased will be included in this investigation that will facilitate more understanding of many unexplained chronic diseases among the populations.

2101F

Evolutionary insights into genetic and environmental factors for Crohn's disease in Ryuku and Northern Kyushu. *S. Nakagome[1], H. Chinen[2], H. Said[3], W. Suda[3], A. Iraha[2], A. Hokama[4], F. Kinjo[4], J. Fujita[4], Y. Takeyama[5], S. Sakisaka[5], T. Matsui[6], J. Kidd[7], K. Kidd[7], S. Kawamura[2], T. Hanihara[8], R. Kimura[4], H. Ishida[4], H. Morita[9], M. Hattori[3], S. Mano[1], H. Oota[8].* 1) The Institute of Statistical Mathematics, Tokyo, Japan; 2) University Hospital, Faculty of Medicine, University of the Ryukyus, Okinawa, Japan; 3) Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan; 4) Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan; 5) Fukuoka University Faculty of Medicine, Fukuoka, Japan; 6) Fukuoka University Chikushi Hospital, Fukuoka, Japan; 7) Yale University School of Medicine, New Haven, CT; 8) Kitasato University School of Medicine, Kanagawa, Japan; 9) Azabu University, Kanagawa, Japan.

Crohn's disease (CD) involves a chronic inflammation in the intestinal tracts, which are caused by complex interaction between multiple genetic and environmental factors. More than seventy loci have been reported as susceptibility genes for CD in Europeans. However, these susceptibilities are different between Japanese and Europeans. Here, we hypothesize that there is the population-specific susceptibility even among Japanese populations. Then, we focus on the TNFSF15 locus whose association with CD has been shown in the Honshu Japanese, and investigate the susceptibility in the Ryukyu and the Northern Kyushu Japanese. The allelic association of TNFSF15 with CD was shown in both of the populations, while the genotype relative risk was different between the Ryukyu (recessive) and the Northern Kyushu (dominant) Japanese. Further, we found that the frequencies of the risk alleles were significantly higher in the Ryukyu Japanese than in the Northern Kyushu and the Honshu Japanese, as well as in the Korean population. These results suggest that the geographic distribution of risk alleles in the TNFSF15 locus can be explained by the demographic effects under the dual-structure model of Japanese populations, and the susceptibility to CD is determined by environmental factors.

2102W

The demographic and genetic dynamics of the human sex ratio from conception to birth. *S. Orzack.* Fresh Pond Research Institute, Cambridge, MA.

We describe the demographic and genetic trajectory of the human sex ratio from conception to birth by analyzing data from 1) three-day-old embryos, 2) induced abortions, 3) chorionic-villus sampling, 4) amniocentesis, and 5) fetal deaths and live births. Our data set is the most comprehensive and largest ever assembled to estimate the sex ratio at conception and the sex ratio trajectory and is the first to include all of these types of data. Our estimate of the sex ratio at conception is 0.5 (proportion male), which contradicts the common belief that the sex ratio at conception is male-biased. The sex ratio among abnormal embryos is male-biased and the sex ratio among normal embryos is female-biased. These biases are associated with the abnormal/normal state of the sex chromosomes, chromosome 15, and chromosome 17. The sex ratio of a cohort of conceptions increases for at least 10 to 15 weeks after conception. The sex ratio levels off after approximately 20 weeks and then declines slowly from 28 weeks to 35 weeks, when it declines markedly and becomes female-biased. This trajectory indicates that there is an early excess of female mortality during pregnancy followed by a later excess of male mortality, with net female mortality during pregnancy exceeding male mortality; this is a fundamental new insight into early human development.

2103T

Critical illness from pandemic influenza A/H1N1 and Streptococcus pneumonia co-infection in Nuevo León, Mexico. *B. Silva[1], A. Padrón-Rocha[1,3], D. Reséndez-Pérez[4], M. Bermúdez-de León[2].* 1) Departamento Inmunogenética, CIBIN-IMSS, Monterrey, Nuevo León, Mexico; 2) Departamento Biología Molecular, CIBIN-IMSS, Monterrey, Nuevo León; 3) FCB-UANL; 4) Departamento de Virología e Inmunología, FCB-UANL.

The influenza A/H1N1 virus pandemic of 2009 started in Mexico and then spread worldwide. Bacterial infection associated with influenza is a common cause of death on seasonal and pandemic influenza. Streptococcus pneumoniae is the major bacterial pathogen associated with the influenza in the 1918 Spanish pandemic. The contribution of bacterial coinfection to critical illness associated with influenza A/H1N1 virus infection remains uncertain. The objective of this study was to determine the association of the influenza pandemic of 2009 with S pneumoniae and its contribution to critical illness. Design: Retrospective cohort study Patients: We examined nasopharyngeal swab samples (NPS) of three hundred five patients who were confirmed for all samples with the WHO approved Real Time PCR influenza A/H1N1 virus pandemic of 2009 essays. Autolysin, encoded by (lytA) gene is required for S. pneumoniae pathogenesis and well-characterized virulence marker. Real-time PCR with sequence- specific primers and fluorescent TaqMan probes were selected to direct detection from NPS. Demographic data, comorbid conditions, illness progression and clinical outcomes were collected. Results: S pneumoniae Co-infection was identified in 10.49% of patients with severe influenza A/H1N1 virus infection. Patients with co-infection were young (median 24.5 [range, 0-76] years) .Co-infection was associated with increased ICU admission (75% patients). Hospital mortality was 16.23% in bacterial co-infection. The presence of S pneumonia was correlated with severe disease. Complex viral, bacterial and host factors contribute to pathogenesis of co- infection. Reductions in morbidity and mortality are dependent on prevention with available vaccines as well as early diagnosis and treatment.

2104F

Quantitative genetic analysis of reactivity to three distinct behavioral challenges in an infant primate model of susceptibility to anxiety disorders and depression. *G.L. Fawcett[1], A.M. Dettmer[2], D. Kay[3], M. Raveendran[1], J.D. Higley[4], N.D. Ryan[5], J.L. Cameron[5,6], J. Rogers[1].* 1) Department of Molecular & Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 3) Department of Clinical & Health Psychology, University of Florida, Gainesville, FL; 4) Department of Psychology, Brigham Young University, Provo, UT; 5) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; 6) Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR.

Prior data clearly show that susceptibility to anxiety disorders and depression is strongly influenced by genetic differences between individuals. Studies of human children in response to novel stressful environments have revealed moderate additive genetic heritability ($h^2 \approx 0.3$) of anxious temperament, a risk factor for later adult mood disorders. Rhesus macaques (*Macaca mulatta*) are the most commonly used non-human primate in biomedical research, including use as models for studies of alcoholism and drug addiction, response to stress and basic aspects of neurodevelopment. Given the significance of developmental experience as well as genetics in forming temperament later in life, we conducted a behavioral genetic analysis of a large (n=428) set of pedigreed infants raised in large social groups that mimic species-typical social organization and interaction. In order to reveal individual differences in underlying temperament, we challenged infant monkeys (aged 40-158 days) with a series of standardized behavioral tests. Each of these conditions varied the environment so as to alter the mildly stressful challenge conferred to the infant: (1) novel environment without threatening stimulus with the mother sedated but present (Freeplay test), (2) novel environment without threatening stimulus without the mother present (Human Intruder Test (HIT)-Alone), and (3) novel environment with exposure to a mildly threatening stimulus (i.e., human staring at monkey) without the mother present (HIT-Stare). We observed significant heritability for willingness to move away from the sedated mother to examine a novel environment ($h^2=0.25\pm0.13$; $p=0.003$). Infants showed heritable variation in a range of behaviors in response to separation stress or threat when the mother was not present ($h^2=0.29$-0.32, $p=1.5\times10^{-3}$-2.1×10^{-4}). By examining these different testing paradigms, we are able to identify genetic influences on reactivity to different environmental challenges. Future analyses will explore genetic correlations among these different reactions and test for genetic associations between different aspects of behavioral variation and a series of specific candidate genetic pathways.

2105W

Genetic and environmental variables contribute to genome-wide methylation variation in geographically diverse Africans. *S. Soi[1,2], L.B. Scheinfeldt[1], D. Diep[3,4], W. Beggs[1], N. Plongthongkum[3], S.A. Tishkoff[1,5], K. Zhang[3].* 1) Dept. Genetics, Univ. Pennsylvania, Philadelphia, PA; 2) Genomics and Computational Biology Graduate Program, Univ. Pennsylvania, Philadelphia, PA; 3) Dept. Bioengineering, Univ. California-San Diego, San Diego, CA; 4) Bioinformatics and Systems Biology Graduate Program, Univ. California-San Diego, San Diego, CA; 5) Dept. of Biology, Univ. of Pennsylvania, Philadelphia, PA.

DNA methylation varies temporally across cell types in an individual, regulating gene expression crucial to biological processes. While methylation is reprogrammed during early development, it also has a heritable component and recent work has shown that this trans-generational inheritance is influenced both by genetic variation as well as environmental factors such as diet. Here we have employed a targeted padlock probe bisulfite sequencing method to characterize DNA methylation in white blood cell samples obtained from a culturally diverse set of sub-Saharan Africa individuals practicing agricultural, pastoralism, agro-pastorialism and hunting-gathering. Using a dimension reduction approach, we demonstrate that the methylation profile of the West African Pygmy population samples is distinct from other Africans even after controlling for covariates such as age and sex. We find that the methylation sites contributing to this signal are enriched near genes that play a role in growth and immunity. In addition, after adjusting methylation frequencies for variation in nearby SNPs, or methylation quantitative trait loci (methQTL), we find that this Pygmy-specific signal still persists, suggesting an environmental component to differences in genome-wide methylation patterns of Pygmy individuals.

2106W

The myocardial infarction-associated gene PHACTR1 is controlled by MEF2. *M. Beaudoin[1], S. Langlois[1], KS. Lo[1], A. N'Diaye[1], JC. Tardif[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montreal, Quebec, Canada; 2) Université de Montreal, Montreal, Quebec, Canada.

Myocardial infarction (MI) is a leading cause of death and disability worldwide. Genetic factors play an important role in the pathogenesis of MI and coronary heart disease (CAD). To characterize the genetic causes of MI in French Canadians, we genotyped SNPs previously associated with MI or CAD in 1176 MI cases and 1996 controls selected from the Montreal Heart Institute (MHI) Biobank. Of the 45 SNPs that we genotyped successfully, 35 had an odds ratio (OR) consistent with the literature (binomial $P=3\times10^{-5}$ and 10 of these 35 SNPs were nominally significant (one-tailed $P<0.05$, binomial $P=4\times10^{-4}$). The strongest genetic association with MI risk that we observed was with rs12526453, a SNP located in the third intron of PHACTR1 (G-allele OR=0.79, $P=8.4\times10^{-4}$). To further characterize the genetic association between the PHACTR1 locus and MI risk, we performed fine-mapping using exon resequencing and imputation. The strongest association with MI was with a genotyped SNP, rs9349379, located ~24 kb downstream of rs12526453 but still in intron 3 of PHACTR1. This SNP was previously reported to associate with CAD, coronary artery calcification and hemodynamic indexes. When we conditioned on genotypes at rs9349379, no additional DNA markers were significantly associated with MI at the PHACTR1 locus ($P>0.10$). Finally, we tested association between genotypes at PHACTR1 rs9349379 and several risk factors for MI in the MHI Biobank samples: hypertension, blood pressure, type 2 diabetes, LDL- and HDL-cholesterol levels and smoking. After accounting for the number of phenotypes tested, none of the associations were significant. This suggests that genetic variation at the PHACTR1 locus influences MI risk through a less appreciated risk factor. We observed a strong correlation between genotypes at rs9349379 and PHACTR1 expression levels in human right coronary arteries ($r^2=0.52$, $P=0.008$). This result indicates that rs9349379 is an expression quantitative trait locus (eQTL) for PHACTR1 in human coronary arteries, providing a strong argument in favor of PHACTR1 being at least one of the genes at the locus implicated in MI. Finally, using electromobility shift assay (EMSA), we demonstrated that this SNP abolish a MEF2 (myocyte enhancer factor-2) consensus binding site. Those results offer mechanistic insights: since the G-allele at rs9349379 is associated with MI risk and low PHACTR1 expression, it suggests that low PHACTR1 expression level is a risk factor for MI.

2107T

Association of Variants in Inflammatory Genes with Disease Severity in Familial Cerebral Cavernous Malformations Type 1. *H. Choquet[1], L. Pawlikowska[1], J. Nelson[1], C.E. McCulloch[1], A. Akers[2], B. Baca[3], B. Hart[4], L. Morrison[3], H. Kim[1], Brain Vascular Malformation Consortium (BVMC).* 1) Center for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, University of California, San Francisco, CA; 2) Angioma Alliance, Durham, NC; 3) Department of Neurology and Pediatrics, University of New Mexico, Albuquerque, NM; 4) Department of Radiology, University of New Mexico, Albuquerque, NM.

Objective: Inflammation may exacerbate disease severity in familial Cerebral Cavernous Malformations (CCM), an autosomal dominant disease characterized by multiple lesions consisting of thin leaky capillaries. The purpose of this study was to investigate whether common variants in inflammatory genes are associated with increased disease severity, as manifested by greater lesion count in familial CCM type 1 (CCM1). **Methods**: 178 Hispanic CCM1 subjects who all carry the same founder Common Hispanic mutation (Q455X in *KRIT1* gene, named CHM) were recruited as part of the Brain Vascular Malformation Consortium (BVMC) study. Lesions were counted by a neuroradiologist on magnetic resonance imaging including susceptibility-weighted imaging performed at enrollment. Samples were genotyped on the Affymetrix Axiom Genome-Wide LAT1 Human Array. Ten candidate genes (*CD14*, *IL1A*, *IL1B*, *IL6*, *IL6R*, *IL10*, *NFKB1*, *TGFB1*, *TNF* and *TLR4*) involved in inflammatory pathways and implicated in vascular diseases or angiogenesis were selected, including 71 variants. Disease severity was quantified as residuals of log-transformed lesion count adjusting for age at enrollment and gender. Linear regression analysis was conducted for each variant assuming an additive genetic model; empirical *p*-values were generated by permutation to account for family structure. **Results**: Lesion count (range: 0-713; mean ± SD: 60.5 ± 116.3) was highly variable and positively correlated with age ($R^2=0.45$, $P<0.001$). The minor alleles of variants in *CD14* (rs2563298, $P=0.005$) and *IL6R* (rs114660934, $P=0.007$ and rs114879247, $P=0.044$) were associated with increased lesion count, although these associations did not pass Bonferroni correction for multiple testing. Variants in *IL1A*, *IL1B*, *IL6*, *IL10*, *NFKB1*, *TGFB*, *TNF* and *TLR4* genes were not associated with lesion count. **Conclusions**: Common variants in inflammatory genes *CD14* and *IL6R* explain additional variability in lesion count independent of age and gender in familial CCM1-CHM cases; however, larger studies are needed to confirm these findings. CD14 and IL6R may have clinical importance as biomarkers of CCM disease severity and progression.

2108F

Are loci associated with low density lipoprotein cholesterol (LDL-C) in non-diabetics similarly associated with LDL-C in type 1 diabetes (T1D)? *K.M. Eny[1], L. Sun[2,3], A.J. Canty[4], S.B. Bull[2,5], A.P. Boright[6], S.M. Hosseini[1], P.A. Cleary[7], J.M. Lachin[7], A.D. Paterson[1,2], DCCT/EDIC Research Group.* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Canada; 4) Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada; 5) Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Prosserman Centre for Health Research, Toronto, Canada; 6) Department of Medicine, University of Toronto, Toronto, Canada; 7) The Biostatistics Center, The George Washington University, Rockville, Maryland.

Elevated LDL-C is a risk factor for cardiovascular disease, a leading cause of death in T1D. Randomization to intensive therapy, aimed at achieving glycemic control in the non-diabetic range, was associated with lower LDL-C in comparison to conventionally treated subjects in the Diabetes Control and Complications Trial (DCCT). The largest genome-wide association study of lipids to date identified 37 loci associated with LDL-C using over 95,000 primarily non-diabetic subjects (Teslovich et al., 2010). Given higher glycemia in T1D, it is unknown whether these loci are similarly associated with LDL-C in T1D. Therefore, to test this hypothesis, we used longitudinal measures of LDL-C from 1304 white participants with T1D from the DCCT. Association results from linear mixed models adjusted for time, randomized treatment group and other covariates showed that the directions for the SNP effect at 35 of the 37 loci were consistent with those reported by Teslovich et al., far exceeding the null expectation of 18.5 (p<10-7). Individually, 17 loci showed evidence of association at p<0.05, but after correcting for multiple testing only 5 (LDLR, APOE-C1-C2, SORT1, ABCG5/8, APOB) remained statistically significant (p≤0.001), suggesting a larger sample is needed to confirm the other locus-specific associations. To test the hypothesis that these 37 loci jointly contribute to susceptibility of higher LDL-C in an additive fashion, we then considered the polygenic risk score method. Risk allele was defined based on the Teslovich et al sample, and risk scores were calculated for each person by multiplying the number of risk alleles by the beta estimate from Teslovich et al for each SNP and summing the product across all 37 loci. Results of this weighted polygenic risk score analysis showed a significant additive genetic effect (p<10-38). To ensure that the result was not driven by the top 5 loci in DCCT, we repeated the analysis focusing on the remaining 32 loci. The strength of the association was attenuated as expected but remained significant (p<10-15). We also tested for a SNP by TREATMENT interaction at each locus but observed only suggestive evidence for interaction at ABO (p=0.01) and ST3GAL4 (p= 0.03). In conclusion, loci ascertained in non-diabetics are similarly associated with lipids in T1D and appear to follow a polygenic additive model. Identifying such loci in T1D can ultimately help target who may benefit from more aggressive lipid management approaches.

2109W

Sequencing candidate genes associated with conotruncal heart defects. *K. Osoegawa[1], C. Parodi[1], N. Mohammed[1], K. Schultz[1], D. Noonan[1], M. Ladner[1], G. Shaw[2], E. Trachtenberg[1], E. Lammer[1].* 1) CHORI, Oakland, CA; 2) Stanford University School of Medicine, Stanford, CA.

Congenital heart defects are the most common malformations, and foremost causes of mortality in the first year of life. Among congenital heart defects, conotruncal defects represent about 20% and are severe malformations with significant morbidity. Conotruncal defects may arise from faulty development of the second heart field cells or from abnormal contributions of cranial neural crest cells. From a population base of 974,579 births during 1999-2004, we identified 389 California infants with tetralogy of Fallot or d-transposition of the great arteries (d-TGA). We isolated DNA from newborn blood spots of these 389 children and used this DNA for sequencing of candidate genes. We chose 15 candidate genes that are either expressed in secondary heart field, for which mutations have been reported in infants with conotruncal defects, or identified via array comparative genomic hybridization (array-CGH). We amplified exons of these genes using the Access Array system (Fluidigm) and generated sequences using the GS FLX+ instrument (Roche 454). DNA sequences were aligned against reference human DNA sequences, and potential novel variations or single nucleotide polymorphisms (SNPs) were identified within protein coding sequences using SeqNext software (JSI Medical Systems). Variations were confirmed by Sanger DNA sequencing. We found 10 novel variants that were not found in the 1000 Genome project and NHLBI Exome Sequencing Project (ESP) databases, in addition to 5 extremely rare SNPs that may be associated with conotruncal heart defects. We assessed the functional consequences using PolyPhen-2, SNAP, PROVEAN and SIFT programs. All four programs predicted that L394P and C392Y substitutions in GATA5 have deleterious effects. Three programs calculated that A459T and P555A in GATA6 and G164R in ZFPM2 are damaging properties.

2110T

Identification of sequence variants of hepatic lipase (LIPC) gene in individuals with extreme HDL-C/TG levels. *D. Pirim[1], F.Y. Demirci[1], X. Wang[1], J.E. Hokanson[2], R.F. Hamman[2], C.H. Bunker[3], M.M. Barmada[1], M.I. Kamboh[1].* 1) Human Genetics, University of Pittsburgh, GSPH, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, Univ Colorado Denver, Aurora, CO; 3) Epidemiology, GSPH, Univ Pittsburgh, Pittsburgh, PA.

The hepatic lipase gene (LIPC) encodes hepatic lipase (HL), which is one of the key lipolytic enzymes in HDL metabolism that regulate plasma triglyceride (TG) and HDL-C levels. Common genetic variation in lipid genes is associated with plasma lipid variation and the risk of coronary heart disease (CHD). We hypothesized that in addition to LIPC common polymorphisms influencing plasma lipid levels, undiscovered uncommon (rare and low frequency) LIPC variants would also contribute to the lipid phenotype. In order to discover rare and low frequency variants, the entire LIPC gene (except for the very large intron 1 which is ~30 kb) was sequenced in individuals with extreme HDL-C/TG levels (95 Whites, 95 Blacks) selected from a non-Hispanic White (NHW) population from the U.S. and a Black population from Nigeria. A total of 411 variants were identified; 128 variants were shared in both populations, 53 were unique to NHWs and 230 were unique to Blacks. One hundred and eighty nine of the observed variants were novel; of which 152 were present only in Blacks and 37 were seen only in NHWs. Two novel variants were located in exons and both were synonymous; others were either intronic or located in flanking regions. In NHWs, 88 of 181 variants were common (MAF≥0.05) and 93 were low frequency or rare variants (MAF<0.05). We found 79 rare variants that are unique to either high HDL-C/low TG or low HDL-C/high TG group. In Blacks, 68 of 358 variants were common and 290 were low frequency or rare; of which 97 were present either in high HDL-C/low TG group or low HDL-C/high TG group. Several variants showed allele frequency differences between the two extreme HDL-C/TG groups in both populations, supporting the previous associations of the LIPC polymorphisms with TG and HDL-C levels. In summary, sequencing LIPC in a relatively small number of individuals with extreme HDL-C/TG levels identified a large number of novel low and rare frequency variants, with the number in blacks approximately 5X that in NHWs. Selected uncommon variants and common tag SNPs are currently being genotyped in our entire sample sets of NHWs (n=623) and Blacks (n=788) for validation and genotype-phenotype association analyses.

2111F

Replication of Hypertension Risk Score in Two Hispanic Populations. *G. Beecham[1], A. Beecham[1], N. Vasudeva[1], L. Wang[1], Z. Liu[1], T. Rundek[2], P. Goldschmidt[3], M. Pericak-Vance[1], D. Seo[2], R.L. Sacco[1,2], S. Blanton[1,2].* 1) Dr. John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst.for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, University of Miami, Miami, FL; 3) Division of Cardiology, University of Miami, Miami, FL.

Hypertension is an important risk factor for coronary artery disease and stroke. Although elevated blood pressure (BP) is associated with environmental factors such as diet and exercise, it is also under genetic influence. Several genome-wide association studies (GWAS) of hypertension have been conducted. In a recent meta-analysis of GWAS data from individuals of European descent the International Consortium for Blood Pressure (ICBP) has identified 29 SNPs that are genome-wide significant. A genetic risk score computed using the 29 SNPs has been replicated in Asian and African populations. Given the sparse data in Hispanics, we sought to evaluate the 29 SNPs and the risk score in Hispanics. We utilized two datasets: one with subjects ascertained through cardiac catheterization labs(the Miami Cardiovascular Registry: MCR); the other with subjects drawn from the community-based Northern Manhattan Study (NOMAS). Both sets are predominantly Hispanic in ethnicity. Only Hispanic samples were used in this study (MCR: N=1063, predominantly Cuban; NOMAS: N=931, 68% Dominican). Genotyping was performed in both sets using the Affymetrix 6.0 array; extensive quality control tests were performed to ensure data integrity. We added 10mmHg and 5mmHg for systolic blood pressure (SBP) and diastolic blood pressure (DBP), respectively, when the subject took antihypertensive medication. Linear regression was performed using the PLINK software, assuming an additive genetic model with age, age^2, and gender included as covariates. Principal components from Eigenstrat were used to control for population substructure. The risk score was computed as described by the ICBP. While confirmation at individual SNPs was marginal, the composite effect of multiple loci was replicated (SBP p=7E-4; DBP p=2E-03), supporting the importance of these genetic variants in influencing BP in Hispanics. Of note, the effect size of the risk score in the MCR was much larger than that in the NOMAS (β=1.27 vs 0.65), with the MCR effect approaching that in European population (β=1.65) and the NOMAS effect approaching that in the African American (β=0.41). The effect size difference in the two Hispanic samples is consistent with the population substructure analysis demonstrating more European ancestry in the MCR. This suggests that, while overall the risk score replicates, the underlying population substructure is relevant to the utility of the risk score.

2112W

Role of the rs1704 HLA-G and rs2227631 PAI-1 polymorphisms in acute coronary syndrome. *I.J. Garcia-Gonzalez[1], Y. Valle[2], F. Rivas[3], LE. Figuera-Villanueva[4], JF. Muñoz-Valle[5], F. Ortega-Gutiérrez[6], BE. Gutiérrez-Amavizca[1], NO. Dávalos-Rodríguez[7], JR. Padilla-Gutiérrez[2].* 1) Doctorado en Genética Humana, CUCS, UdeG, Guadalajara, Mexico; 2) Departamento de Clínicas Médicas, CUCS UdeG; 3) Secretaria de Salud Jalisco; 4) CIBO, IMSS; Centro de investigación Biomédica de Occidente; 5) Departamento de Biología Molecular, CUCS UdeG; 6) CMNO, IMSS; Centro Médico Nacional de Occidente; 7) Instituto de Investigación en Genética Humana, CUCS UdeG.

Introduction: Acute coronary syndrome (ACS) is an important cause of morbidity and mortality. The pathogenesis is linked to immunologic and thrombotic process[1], in which the HLA-G[2] and PAI-1[3] could be involved. The rs1704 is a del/ins of 14 bp at +2960 position in exon 8 of HLA-G, that has been associated with an alternative splicing with changes in the stability of the mRNA[2]. The rs2227631 is a G>A variation in the promoter sequence -844 in PAI-1 gene, the A allele has been related with gene expression changes[3]. We evaluate the participation of rs1704 and rs2227631 polymorphisms in patients with ACS. **Methods**: Were recruited 86 patients with ACS according to American College of Cardiology and 150 healthy subjects age-matched. The study was made in accordance to the Declaration of Helsinki, with acceptance to participate and with an informed written consent. Ethical approval was obtained by the CUCS, UdeG (C.I. 069-2012). The rs1704 polymorphism was identified by conventional PCR and the rs2227631 polymorphism by PCR-RFLP. Fragments were separated in polyacrylamide gel electrophoresis. The genotype and allele differences were estimated by Fisher's exact test. The association measure was evaluated by OR and 95% of confidence intervals. The significance level was p<0.05. **Results**: Both polymorphisms were on Hardy-Weinberg equilibrium expectations. The allele and genotype distributions showed no significant differences between groups. The most common ACS diagnostic and related risk was STEMI and high blood pressure, respectively. The risk factors by genotype were similar among patients. **Conclusion**: The rs1704 HLA-G and rs2227631 PAI-1 polymorphisms are not genetic susceptibility markers to Acute Coronary Syndrome in Western Mexico population. This study was supported by Grant No. 2012-07-190437 to JRPG from the Fondo Mixto CONACYT-Gobierno del Estado de Jalisco. **References**: 1. Galkina E, Ley K. Immune and inflammatory mechanisms of atherosclerosis. Annu Rev Immunol 2009;27:165-97; 2. Boiocchi C, Bozzini S, Zorzetto M, Pelissero G, Cuccia M, Falcone C. Association between two polymorphisms in the HLA-G gene and angiographic coronary artery disease. Mol Med Report. 2012;5:1141-5; 3. Padilla-Gutiérrez JR, Palafox-Sánchez CA, Valle Y, Orozco-Barocio G, Oregón-Romero E, Vázquez-Del Mercado M, Rangel-Villalobos H, Llamas-Covarrubias MA, Muñoz-Valle JF. Plasminogen activator inhibitor-1 polymorphisms (-844 G/A and HindIII C/G) in systemic lupus.

2113T

FLNA mutations found in patients with aortic aneurysm/dissections.
H. Morisaki[1], I. Yamanaka[1], A. Yoshida[1], R. Sultana[1], K. Minatoya[2], I. Shiraishi[3], H. Ichikawa[4], T. Kosho[5], H. Sonoda[6], T. Morisaki[1,7]. 1) Dept Bioscience & Genetics, NCVC Res Inst, Suita, Osaka, Japan; 2) Dept Cardiovasc Surgery, NCVC, Suita, Osaka, Japan; 3) Dept Pediatric Cardiol, NCVC, Suita, Osaka, Japan; 4) Dept Pediatric Cardiovasc Surgery, NCVC, Suita, Osaka, Japan; 5) Dept Med Genetics, Shinshu U Grad Sch Med, Matsumoto, Nagano, Japan; 6) Dept Cardiovasc Surgery, Kyushu U Grad Sch Med Sci, Fukuoka, Fukuoka, Japan; 7) Dept Mol Pathophysiol, Osaka U Grad Sch Pharm Sci, Suita, Osaka, Japan.

Thoracic aortic aneurysm /dissections (TAAD) often result from genetic predisposition, since about 10 to 20 % of these patients have a first degree relative with TAAD. To find disease-causing gene mutations for TAAD, we generated exome sequence from 213 individuals with young-onset or familial TAAD and/or related connective tissue disorders, who were confirmed not to have any pathogenic mutation in *FBN1*, *TGFBR1*, *TGFBR2*, *ACTA2* or *SMAD3*. Sequence data was interrogated for sequence changes, compared with the data base of 1000 Genomes. While pathogenic mutations in several genes potentially related to aortopathy emerged from this screening, *FLNA* mutations were recurrently identified in 4 probands. Three females (20-54 years old) were heterozygotes for nonsense or frame-shift mutations and 1 male (22 years old) was a hemizygote for missense mutation in the X-linked *FLNA* gene, encoding filamin A. Family study revealed a mother of the male patient carried the same *FLNA* mutation as her son, and she also had dilated cardiomyopathy with severe mitral valve insufficiency. Brain MRI study exhibited periventricular nodular heterotopias (PNH) in all 5 patients. Further studies on the clinical features showed several cardiovascular events including aortic surgery in 2, severe mitral valve insufficiency in 3, aortic valve insufficiency in 3, severe orthostatic hypotension in 2 and patent ductus arteriosus in 1. Also, several features related to connective tissue disorders, including joint hypermobility in 4, inguinal hernia in 3, thoracic deformity in 3 and venous thrombosis in 2, were observed. Regarding neurological features, all had normal intelligence although two patients were diagnosed as adult-onset epilepsy. *FLNA* was shown as a responsible gene for X-linked PNH. Also, several distinct genetic disorders, including otopalatodigital syndrome, frontometaphyseal dysplasia, Melnick-Needles syndrome and Ehlers-Danlos syndrome, were caused by *FLNA* mutation. The major clinical features of our cases were aortic disorder with some connective tissue involvement. Based on these results and literature reviews, we suggest that it is advisable to screen PNH in patients with young-onset or familial TAAD of unknown etiology, even if there are no neurological findings, since individuals with *FLNA* mutation can exhibit cardiovascular phenotype dominantly. Conversely, we propose to screen cardiovascular manifestations in individuals with X-linked PNH.

2114F

Arrhythmogenic Right Ventricular Cardiomyopathy-Molecular Analysis.
P. Nallari. Dept. of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India.

Arrhythmogenic right ventricular cardiomyopathy (ARVC) is an autosomal dominant inherited disease with incomplete penetrance and variable expression. It is a leading cause of sudden cardiac death in the ≤35-year age group. The aim of this study was to highlight the genetic basis of Arrhythmogenic right ventricular cardiomyopathy in Indian origin. Blood sample from 100 control individuals and 40 probands with clinical data was obtained. The ARVC diagnosis was established based on Task Force Criteria (TFC) set by the ESC. Mutational screening of two desmosomal genes Plakophilin-2 (PKP-2), Desmoplakin (DSP) and two non-desmosomal genes Ryanodine receptor-2 (RyR-2) and Transforming growth factor β -3 (TGFβ -3) was carried out. Variant band pattern were confirmed on commercial sequencing and insilico analysis was carried out. The study revealed 23.529% of patients had mutations in desmosomal genes where as 14.705% of patients had mutations in non-desmosomal genes. 44.1% familial cases clearly establishing the genetic basis of ARVC. The total percentage of mutations reported is 33.3%. Three different mutations in PKP-2 gene CT deletion, P244L and G>A transition in intronic region of exon 4 and an insertion in intronic region of exon 28 of RYR-2 and an amino acid substitution in DSP (I305F) gene/s were observed . Novel SNPs were also reported in RYR-2(A>C, A>G and A>C) and DSP (deletionG) gene/s. The potentials of pathogenecity and functionality of the above results are carried out by In-silico analysis.

2115W

A Variant in the 5-Hydroxytryptamine Receptor 1 A (HTR1A) Gene Is Associated with Platelet Reactivity and Incident Cardiovascular Events.
S.H. Shah[1,2], M.A. Babyak[3], E.R. Hauser[2], D. Craig[2], M. Chryst-Ladd[2], C. Haynes[2], B. Brumett[3], W.E. Kraus[1,2], R. Becker[1], R.B. Williams[3]. 1) Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, NC; 2) Duke Institute for Molecular Physiology, Durham, NC; 3) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC.

Background. Variation in the 5-hydroxytryptamine receptor 1 A (HTR1A) gene has been associated with psychiatric phenotypes. Given that depression has been linked to cardiovascular disease (CVD) events, we hypothesized that HTR1A SNP rs1364043 is associated with platelet reactivity (an intermediate phenotype for CVD events), and is associated with CVD event endpoints. Methods. Rs1364043 in HTR1A was genotyped using an ABI Taqman assay in 8697 participants in the CATHGEN biorepository of patients referred to Duke University for cardiac catheterization for evaluation of CVD. Incident events were defined as all-cause death and/or myocardial infarction (MI) after enrollment. Race-stratified Cox proportional hazards modeling was used to test for association with time-to-event in two models (death and death or MI), adjusted for age, sex, hypertension, diabetes, ejection fraction, CAD, dyslipidemia, smoking, and BMI. Expression quantitative trait loci (eQTL) analyses were used to test for cis and trans effects of SNP on RNA expression using additive multivariable linear regression. Association between rs1364043 and platelet reactivity was tested in N= 410 healthy individuals, defined as a binary trait of hyperreactivity (>60% aggregation to epinephrine [2 μM] and serotonin [10 μM]), using generalized linear models adjusted for age, sex and race. Results. In whites, the C allele of rs1364043 was associated with time to death in additive (hazard ratio [HR] 1.09, 95% CI [1.0-2-1.12], p=0.04) and dominant models (HR 1.11 [1.0001-1.24], p=0.05), and with time to death/MI (HR 1.10 [1.02-1.19], p= 0.02 for additive and HR 1.13 [1.03-1.25], p=0.02 for dominant model). eQTL analyses identified no cis effects. The top trans transcripts associated with rs1364043 including several genes involved in cell cycle processes, protein synthesis and stability including CDKN1a, RPL32, ASNS, DDHD2, MED27 and STOML2 (p=0.002-0.0003). Platelet reactivity studies showed a trend for higher proportion of individuals with high platelet reactivity in rs1364043 C allele homozygotes as compared with heterozygotes or wildtype (73.7% vs. 58.0% vs. 52.0%, p=0.04, dominant model). Conclusions. Genetic variants associated with intermediate traits may themselves serve as disease risk variants. We have shown that a variant in HTR1A, previously associated with intermediate psychosocial traits, is also associated with platelet hyperreactivity and predicts CVD events, with an up to 13% increased risk.

2116T

Exome sequencing identifies BAG3 gene mutation in dilated cardiomyopathy.
M. Taylor[1], R.L. Begay[1], V.D. Myers[2], S.L. Graw[1], D. Slavov[1], P. Boyer[1], L. Mestroni[1], A. Feldman[2]. 1) Adult Med Gen Prog, University of Colorado - CU Cardiovascular Institute, Aurora, CO; 2) Temple University School of Medicine, Philadelphia, PA.

Background: Recently BAG3 has been identified as a cause of genetic forms of dilated cardiomyopathy. The consequences of BAG3 mutations on cardiac pathology and the role of BAG3 in heart failure are not known. Methods: Whole exome DNA analysis was performed on a multi-generational family affected with dilated cardiomyopathy. Prior studies including linkage analysis, candidate gene analysis, and SNP array analysis had failed to locate the pathogenic mutation. Exome capture was achieved using the Agilent SureSelect Exome capture system and sequencing was completed on an Illumina HiSeq2000. Analysis of explanted human heart tissue from one mutation carrier was performed and BAG3 levels in control and failing hearts were assayed by Western blot analysis. Results and Conclusions: A 10-nucleotide deletion in exon 4 of the BAG3 gene was identified through exome sequencing from distantly affected family members. The mutation was present in seven affected family members and absent in eight healthy relatives. Standard histology of a BAG3 mutation carrier's explanted heart revealed mild to moderate myocyte hypertrophy and patchy mild to moderate interstitial fibrosis; changes more prominent in the left ventricle and septum than the right ventricle. Also, electron microscopy suggested increased glycogen deposition, possibly related to the role of BAG3 protein in autophagy and cellular trafficking. BAG3 levels were markedly depressed without evidence of truncated BAG3 protein in the explanted heart specimen, consistent with haploinsufficiency. Conclusions: Exome sequencing is increasingly useful in identifying pathogenic mutations in cardiomyopathy families where distant relatives are available for segregation analysis. The microscopic evaluation identified notable features of BAG3 heart biopsy due to apparent haploinsufficiency of BAG3, suggesting that modulation of BAG3 levels could represent a novel avenue for therapeutic exploration.

2117F

Haploinsufficiency of *RERE* contributes to cardiovascular defects associated with 1p36 deletions. *H. Zaveri[1], B.J. Kim[1], A. Hernandez-Garcia[1], T.F. Beck[1], O.A. Shchelochkov[2], M. Justice[1], B. Lee[1], S.R. Lalani[1], D.A. Scott[1].* 1) Molec & Human Gen, Baylor College Med, Houston, TX; 2) Dept of Pediatrics, The University of Iowa, Iowa City, IA.

Terminal deletions of chromosome 1p36 affect 1 in 5000 newborns and cause a clinically recognizable syndrome characterized by cognitive impairment, eye/vision problems, hearing loss, facial clefting, cardiovascular malformations and cardiomyopathy. Although cardiovascular anomalies and/or cardiomyopathy affect 71% of individuals with 1p36 deletions, and are the most common cause of morbidity and mortality in the newborn stage,the genes that contribute to these phenotypes have yet to be identified. Using a clinical and molecular cytogenetic data from individuals with terminal and interstitial deletions, we have created a deletion/phenotype map of the 1p36 region. This map includes five non-overlapping critical regions for cardiovascular defects and two non-overlapping critical regions for cardiomyopathy. The majority of these regions contain positional candidate genes which have been shown to cause cardiovascular malformation and/or cardiomyopathy in mice. These genes include *DVL1*, *SKI*, *PRDM16*, *UBE4B*, *PDPN*, *SPEN*, *HSPG2*, *CDC42*, and *LUZP1*. The arginine-glutamic acid dipeptide repeats gene (*RERE*) encodes a nuclear receptor coregulator and is located in a critical region for both cardiovascular malformations and cardiomyopathy on 1p36 and is required for normal retinoic acid signaling. Since perturbations in retinoic acid signaling can cause cardiovascular malformations, and retinoic acid signaling suppresses myocardial hypertrophy and cardiac fibrosis, we hypothesized that RERE deficiency contributes to the cardiac phenotypes seen in individuals with terminal and interstitial 1p36 deletions that include *RERE*. To test this hypothesis, we generated an allelic series of RERE-deficient mice using an *Rere* null-allele and a hypomorphic *Rere* allele (*eyes3*) identified in our laboratory. A portion of *Rere*$^{+/-}$ mice had aortic arch anomalies. *Rere*$^{-/eyes3}$ mice had more serious cardiac defects which included aortic arch, conotruncal, and septal defects. In absence of structural cardiac defects, these mice spontaneously developed cardiac fibrosis and cardiomegaly. *Rere* null mice (*Rere*$^{-/-}$) had unlooped hearts and died of cardiac failure at E9.5. Cardiac defects were also documented in mice in which *Rere* was conditionally ablated in the first and second heart fields. We conclude that RERE functions in a cell autonomous manner to direct cardiac development and that deletion of *RERE* contributes to the cardiac malformations and cardiomyopathy associated with 1p36 deletions.

2118W

Analysis of Common and Coding Variants with Cardiovascular Disease in the Diabetes Heart Study. *J.N. Adams[1,2,3], L.M. Raffield[1,2,3], B.I. Freeman[4], C.D. Langefeld[5], M.C.Y. NG[2,3], J.J. Carr[6], A.J. Cox[2,3,7], D.W. Bowden[2,3,7].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC; 5) Division of Public Health Sciences, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 6) Department of Radiologic Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 7) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC.

Type 2 diabetes (T2D) affects ≈8% of the US population. T2D is a major risk factor for cardiovascular disease (CVD). Up to 65% of T2D deaths are CVD-related. Identification of genetic risk factors for CVD is important to understand individual disease risk. Recent GWAS meta-analyses in the CHARGE consortium identified loci associated with coronary artery calcification (CAC), myocardial infarction (MI), carotid intima media thickness (IMT), and/or carotid plaque. CAC and IMT are clinically relevant measures of subclinical CVD and strongly predict mortality. We examined if prior associations from the population-based CHARGE study were applicable in the Diabetes Heart Study (DHS), a family-based cohort (n=1220) with >80% T2D affected subjects. SNPs (n=36) from CHARGE were evaluated in GWAS data from the DHS. Phenotypes including vascular calcification and conventional CVD risk factors were tested. No SNPs were significantly associated after correction for multiple comparisons ($p<0.0014$), although multiple SNPs showed nominal significance: rs599839 in *PSRC1* (β=-0.315, p=0.0084) and rs646776 in *CELSR2* (β=-0.380, p=0.0099) were associated with CAC; rs17398575 in *PIK3G* was associated with aortic calcification (AACP) (β=11.0, p=0.0054). Additional SNPs at *COL4A2* and *CXCL12* were nominally associated with all-cause or CVD-mortality. Genes implicated by CHARGE were investigated further by examining exonic variants. 209 coding variants were investigated using Illumina HumanExome BeadChip genotype data from the DHS; rs61730407 in *OR2G3* was significantly associated with AACP (β=85.4, $p=7.1\times 10^{-5}$). Additional SNPs were nominally associated with multiple lipid measures and subclinical CVD (CAC, AACP, and IMT) ($0.00026<p<0.0097$). Genetic risk scores (GRS) were calculated to evaluate multiple variants from the CHARGE study. We constructed a GRS containing CHARGE SNPs (n=12) associated with CAC. A second GRS was calculated containing SNPs associated with both CAC and MI (n=8). GRS associations with mortality, CVD events and vascular calcified plaque were examined. The GRS containing CAC SNPs was associated with history of CVD events (p=0.033; OR=1.09). The CAC/MI GRS was associated with history of MI (p=0.026; OR=1.15). We found that genetic risk factors for subclinical CVD in the general population (CHARGE) were, in part, associated with T2D-related risk factors and outcomes.

2119T

***VKORC1* and *CALU* eQTLs and susceptibility to venous thromboembolism among African Americans.**
W. Hernandez[1], E.R. Gamazon[1], A. Konkashbaev[1], K. Aquino-Michaels[1], T.J. O'Brien[2], A.F. Harralson[2], R.A. Kittles[3], A. Barbour[4], M. Tuck[5], S.D. McIntosh[4,5], J.N. Douglas[5], L.H. Cavallari[6], M.A. Perera[1]. 1) Department of Medicine, Section of Genetic Medicine, University of Chicago, Chicago, IL; 2) The George Washington University, Department of Pharmacology and Physiology Washington, DC; 3) University of Illinois, Department of Medicine, Institute of Human Genetics, Chicago, IL; 4) The George Washington University, Department of Medicine, Washington, DC; 5) Uniformed Services University of the Health Sciences, Department of Veterans Affairs, Washington, DC; 6) University of Illinois, Chicago, Department of Pharmacy, Chicago, IL.

Venous thromboembolism (VTE) is a complex disease encompassing deep vein thrombosis (DVT), pulmonary embolism (PE), or a combination of both. In the US, each year over 900,000 individuals are hospitalized due to VTE and over 300,000 patients die. African Americans (AAs) have the highest incidence and mortality rates of DVT and PE. Warfarin is used to treat and prevent DVT and/or PE. However, warfarin dose requirement has been shown to be higher among AAs as well as DVT/PE patients. In this study, we aimed to investigate the role between *VKORC1* and *CALU* expression quantitative trait loci (eQTLs) and DVT/PE susceptibility. As subtle differences in plasma levels of proteins involved in the coagulation pathway may be risk factors, eQTLs may yield novel insights into the genetic risk of DVT/PE. Using publicly available liver eQTL data, we identified both cis and trans eQTLs for *VKORC1* and *CALU* genes known to be associated with warfarin dose. 72 SNP eQTLs were genotyped in a study population of 462 AAs on stable warfarin dose; of which 256 individuals were treated with warfarin due to DVT/PE (cases) and the remaining due to a variety of other conditions including atrial fibrillation (controls). We found a significant decrease in risk of DVT/PE for carriers of the minor allele of two *VKORC1* eQTLs and one *CALU* eQTL (rs9925964, OR=0.53, P=0.01; rs12597511, OR=0.48. P=0.02; and rs11054879, OR=0.61, P=0.03 respectively). We also found our DVT/PE patients had a higher percentage of West African ancestry and were younger compared to controls (t=-1.991, p=0.04 and t= 2.720, p=0.007 respectively). Interestingly, the frequency of these protective alleles were approximately 10% in our study population and significantly higher in the HapMap European Americans (CEU) at approximately 40% but lower among Yorubans (YRI) at 5%. The lower frequency of these protective alleles among populations of African descent compared to European Americans (EAs) is particularly interesting as the risk of DVT/PE is lower among EAs than AAs. In this study, we also observed *VKORC1* -1639 is highly associated with warfarin dose requirement (β=-0.42 and P=6.79E-12) but not to DVT/PE susceptibility (OR=0.79 and P=0.32). By investigating eQTLs for genes known to contribute to warfarin dose requirement we have uncovered novel disease loci involved in the coagulation cascade and risk of DVT/PE. These findings may help explain the increased risk of DVT/PE seen in the African American population.

2120F

Multi-allelic Haplotype Association Identifies a new Protective Haplotype in the KCNN3 Gene for Post-Operative Atrial Fibrillation.
M. Heydarpour[1], C. Collard[2], A. Fox[1], J. Muehlschlegel[1], M. Sigurdsson[1], S. Shernan[1], S. Body[1]. 1) Department of Anesthesiology, Preoperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA; 2) Division of Cardiovascular Anesthesia, Texas Heart Institute, Houston, TX.

Atrial fibrillation (AF) is the most common arrhythmia after cardiac surgery. Two regions have been associated with AF in ambulatory populations (KCNN3, ZFHX3), but not yet associated with AF in cardiac surgical populations. Candidate single nucleotide polymorphisms (SNPs) that describe risk haplotypes for AF in ambulatory populations were genotyped in this study. The objective was to replicate findings of AF-associated loci while accounting for previous association of 4q25 with postoperative AF (poAF). Using a case-control association study of 325 Caucasian patients with poAF and 727 Caucasian controls, we examined 488 SNPs in three regions (1q21.3, 4q25 and 16q22.3) for SNP and haplotype association analysis using logistic regression analysis while accounting for age, gender, prior AF and other clinical variables, and an additive genetic model. Maximum likelihood hierarchical clustering methods were used for haplotype clustering analysis. Permutation testing was used to assess statistical significance. Significant association was identified between SNPs rs6683557 (1q21.3, OR=0.70, P= 0.00051), rs3853445 (4q25, OR=0.68, P=0.0016), and rs67402452 (16q22.3, OR=1.45, P=0.0033) and poAF after adjusting for prior association of SNPs in 4q25 and for clinical variables. We identified a 1q21.3 haplotype, AAAAG which conferred significant protection against poAF (OR=0.564, P= 0.0011). All five SNPs in this haplotype were intronic to KCNN3, which encodes a potassium channel protein involved in atrial repolarization. The rs3853445 in 4q25 region as a protective variant against AF has been reported by Lubitz et al., 2010, (OR=0.71, P=4.1×10-5) which is confirmed by this study. Two other susceptible markers (rs2200733 & rs13143308) are identified in 4q25 for risk of poAF which previously reported by Body et al. 2009 to predict poAF after coronary artery bypass graft (CABG). These results suggest that a common haplotype AAAAG of KCNN3 gene confers a significant protective effect on the development of poAF. Our results confirmed the role of these two regions variants for AF risk in previous studies (Body et al. 2009, Benjamin et al. 2009, Ellinor et al. 2012) for European-descent population. Consideration of multiple susceptibility signals at these regions identifies individuals in high risk of AF or protective against AF and they may localize regulatory elements at the locus with biological relevance in the pathogenesis of AF.

2121W

Elucidating the role of genes encoding sarcomeric structural proteins in modulating cardiac hypertrophy in Hypertrophic Cardiomyopathy.
C.J. Kinnear[1], L. Bloem[1], L. van der Merwe[2], M. Revera[3], M. Heradien[4], A. Goosen[4], PA. Brink[4], JC. Moolman-Smook[1]. 1) University of Stellenbosch/ Medical Research Council (US/MRC) Centre for Molecular and Cellular Biology, Department of Biomedical Sciences, Stellenbosch University, South Africa; 2) Biostatistics Unit, Medical Research Council of South Africa; 3) Department of Cardiology, IRCCS San Matteo Hospital, Pavia, Italy; 4) Department of Medicine, Faculty of Health Sciences, University of Stellenbosch, South Africa.

Hypertrophic cardiomyopathy (HCM) is an autosomal dominantly inherited cardiac muscle disease clinically defined by unexplained left ventricular hypertrophy (LVH). The degree and location of the LVH in HCM patients is extremely variable, even in patients carrying the same HCM-causing mutations. Phenotypic expression of HCM is therefore dependant on the disease-causing mutation as well as additional determinants including environmental factors and modifier genes. Many investigations aimed at identifying modifier genes in HCM have focused on genes that are relatively far removed from the primary disease-causing genes. We hypothesise that since most HCM-causing mutations are found in genes encoding proteins of the cardiac sarcomere, factors that closely interact with defective sarcomeric proteins, will have a greater capacity to modulate them and thus modulate the LVH phenotype. Therefore genes encoding sarcomeric structural proteins as well as those encoding enzymes that function in sarcomere-based energetics may be considered plausible modifier genes of LVH in HCM. In the present study we focused on genes encoding sarcomeric structural proteins as candidate modifiers of LVH. A total of 24 single nucleotide polymorphisms (SNPs) in 9 genes (ACTC1, TNNT2, TNNI3, TPM1, ACTN3, MYL3, MYH7, MYL2 and MyBPC3) were investigated for association with cardiac hypertrophy traits in a familial HCM-cohort. A total of 388 individuals from 27 HCM South African HCM families in which one of three HCM founder mutations segregates (R92WTNNT2, R403WMYH7 and A797MYH7) were genotyped by validated Taqman® SNP genotyping assays. All participants provided written informed consent and was conducted in accordance with the ethical guidelines as set out in the 'Declaration of Helsinki', 2008. We identified a modifying role in the development of hypertrophy in HCM for each of the candidate genes investigated in the present study. More specifically, single variant association analyses identified a modifying role for variants within MYH7, TPM1 and MYL2, while Haplotype-based association analyses identified combined modifying effects for variants within ACTC, TPM1, MYL2, MYL3 and MYBPC3. The data presented here indicates that variations sarcomeric structural genes may act as modifiers of cardiac hypertrophy in HCM. These findings further enhance our understanding of genotype/ phenotype correlations in HCM and could potentially improve patient risk stratification and management.

2122T

Analysis of Coding Variants in C1q/TNF Superfamily Genes in the Diabetes Heart Study.
L.M. Raffield[1,2,3], A.J. Cox[2,3,4], C.D. Langefeld[5], M.C.Y. Ng[2,3], J.J. Carr[6], B.I. Freedman[7], D.W. Bowden[2,3,4]. 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC, USA; 2) Center for Human Genomics, Wake Forest University Health Sciences, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA; 5) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 6) Department of Radiologic Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 7) Department of Internal Medicine - Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, USA.

Members of the C1q/tumor necrosis factor (TNF) superfamily have a variety of roles in inflammation and metabolism and have been associated with cardiovascular disease (CVD) and type 2 diabetes (T2D). Previously, we have shown that coding variants in the collagen domain of the C1q family member adiponectin dramatically lower serum adiponectin levels and impede multimerization. We hypothesized that coding variants influencing multimerization or other protein-protein interactions in C1q/TNF superfamily members and their binding partners/ receptors could impact CVD- and T2D-related phenotypes. This analysis was performed in 1190 European American individuals from 468 families in the Diabetes Heart Study (DHS). In the DHS cohort, 83.7% of participants were affected by T2D with average disease duration of 10.5 ± 7.2 years. Genotyping was completed using the Illumina Infinium HumanExome BeadChip array. 464 coding variants in 97 C1q/TNF related genes were analyzed for associations with CVD- and T2D-related phenotypes, including measures of vascular calcified plaque, glucose control, dyslipidemia, inflammation, renal function, bone density, and adiposity. Single variant association analyses were performed using a variance components based approach in SOLAR, while analyses of all variants in a given genetic locus were performed using the rare variant Sequence Kernel Association Test program for family data (famSKAT). These analyses were adjusted for age, sex, and T2D status. While significant associations with mutations in domains important for multimerization were rare, several associations were observed in binding partners/receptors, including a variant in cadherin 13 (Asn39Ser, MAF= 0.006) associated with abdominal aortic calcification (p= 6.23×10^{-6}) and a variant in calreticulin (Glu381Ala, MAF= 0.001) associated with CRP levels (p= 1.80×10^{-5}). Adiponectin can bind both cadherin 13 and calreticulin, and these loci have been previously associated with CVD-related phenotypes and prevention of systemic inflammation, respectively. The 8 variants in the cadherin 13 locus were also cumulatively associated with abdominal aortic calcification (p= 2.52×10^{-5}) in analysis using famSKAT. Replication of these uncommon coding variants in additional cohorts will be necessary, but these results suggest important roles for C1q/TNF superfamily members in CVD- and T2D-related phenotypes which can be revealed by analysis of exonic variants.

2123F

Cross-sectional and Longitudinal Replication Analyses of Genome-wide Association Loci of Type 2 Diabetes in Han Chinese Populations.
Q. Zhao[1], X. Kong[2], J. Hong[2], X. Zhang[2], J. He[1], W. Yang[2]. 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Endocrinology, China-Japan Friendship Hospital, Beijing, China.

Most of the genetic loci for type 2 diabetes (T2D) identified by genome-wide association studies (GWAS) in populations of European ancestry are still controversial about their associations with T2D of East Asians. The purpose of our study is to examine recently GWAS-identified loci for their associations with T2D and quantitative glycemic traits and their effects on longitudinal changes in fasting plasma glucose (FPG) and T2D development in Han Chinese populations. Single nucleotide polymorphisms (SNP) from 26 loci were genotyped in 10,001 Chinese Hans (5,338 T2D cases and 4,663 controls). Eight SNPs in or near *WFS1*, *CDKAL1*, *CDKN2A/2B*, *CDC123*, *HHEX*, *TCF7L2*, *KCNQ1*, and *MTNR1B* showed significant associations with T2D ($P < 0.05$). The most significant SNP, rs10811661 near *CDKN2A/2B* ($P = 1.11 \times 10^{-8}$), were also associated with glucose level of 2-h post an oral glucose tolerance test ($P = 9.11 \times 10^{-3}$) and insulinogenic index ($P = 2.71 \times 10^{-2}$). A total of 16 SNPs showed consistent association directions with the reported T2D GWAS meta-analysis results of the Asian Genetic Epidemiology Network (AGEN) (6,952 cases and 11,865 controls). Twelve SNPs of the 16 aforementioned SNPs reached the significance of 0.05 in a combined analysis of AGEN and the current study (including 28,818 subjects), and most of them had similar effect sizes to those seen in European populations. In a cohort of 1,881 Chinese Hans without T2D at baseline, individuals carrying more risk alleles of the replicated SNPs in the combined analysis had greater FPG increase and T2D incidence over a 7.5-year follow-up period, with each quartile increase in the number of risk alleles being associated with a 0.06 mmol/l greater increase in FPG ($P = 0.03$) and 19.0% higher odds for developing T2D ($P = 0.058$). Our results indicate that some of GWAS-identified loci for T2D in Europeans are associated with T2D and glycemic traits and may also predict risk for T2D development in the Chinese population.

2124W

Sex-specific genetic variants on 11p15 influence high density-lipoprotein cholesterol levels in long-lived subjects: The Long Life Study Family (LLFS). *M.F. Feitosa[1], M.K. Wojczynski[1], C.M. Kammerer[2], W. Daw[1], R.J. Straka[3], J.H. Lee[4], K. Christensen[5], A.B. Newman[6], M.A. Province[1], I.B. Borecki[1].* 1) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 2) Department of Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 3) Department of Experimental and Clinical Pharmacology, College of Pharmacy, University of Minnesota, Minneapolis, MN; 4) Sergievsky Center and Taub Institute, College of Physicians and Surgeons, Columbia University, New York, NY; 5) Institute of Public Health at the University of Southern Denmark, Denmark; 6) Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.

Studies have shown that high levels of high-density lipoprotein cholesterol (HDL-C) have protective effects on atherosclerosis and cardiovascular disease, and also have been associated with longevity. Some of the common variants, which have been identified by genome-wide association (GWA) studies, appear to have sex-specific effects. However, while GWA is suited for interrogation of common variation in the genome, rare variants will likely be missed and may account for some degree of the missing heritability. In order to identify regions of the genome that are likely to harbor collections of rare variants influencing HDL-C, we returned to classical linkage analysis, using data from 2,256 women and in 1,858 men from 534 healthy-eldest families participating in the LLFS. A linkage marker panel was created by estimating haplotypes across the genome comprised of sets of up to 5 SNPs within a 0.5 cM interval from the GWA panel, resulting in markers with high polymorphic information content. The proportion of alleles shared Identical-by-descent over all such loci was estimated for all relative pairs using LOKI, and linkage analysis was carried out using SOLAR. For subjects taking lowering-lipid medications, a constant was added to their measured HDL-C levels specific to the reported class of medication. HDL-C was then adjusted for age effects and principal components for stratification using stepwise regression, within sex. Significant linkage was found on 11p15 for HDL-C in women (LOD=4.2) at 19 cM, but not in men (LOD=0.5) or in sex-combined data (LOD=2.2). To test whether common variants from GWA explained this linkage peak, we further adjusted HDL-C for the effects of our GWA SNPs with p <1e-4 and repeated the linkage analysis. Despite the decrease in LOD from 4.2 to 3.0, evidence of linkage remained, suggesting that other variants in the region influence HDL-C, and these may be rare variants. Because it is likely that different variants are segregating in different families, we aggregated the 277 families (1,065 women) supporting linkage to this region with Heterogeneity LOD=20.9. The strong evidence of genetic effects in these families were within 5-11 Mb region. Chromosome regions thus detected can be prioritized for sequencing to identify the specific rare variants influencing HDL-C in these healthy-longevous families.

2125T

Influence of genetic determinants and lifestyle factors on blood lipid traits: 10-year follow-up of the GLACIER Study. *T.V. Varga[1], D. Shungin[1,2,3], R.W. Koivula[1], G. Hallmans[2], I. Johansson[3], I. Barroso[4,5,6], F. Renström[1], P.W. Franks[1,2,7].* 1) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Skåne University Hospital Malmö, SE-205 02, Malmö, Sweden; 2) Department of Public Health & Clinical Medicine, Umeå University Hospital, Umeå, Sweden; 3) Department of Odontology, Umeå University, Umeå, Sweden; 4) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 5) NIHR Cambridge Biomedical Research Centre, Institute of Metabolic Science, Addenbrooke's Hospital Cambridge, UK; 6) University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital Cambridge, UK; 7) Department of Nutrition, Harvard School of Public Health, Boston, USA.

Background - Recently, a genome-wide meta-analysis identified 95 loci associated with lipid traits (Teslovich *et al.*, Nature, 2010), here we attempted to test whether these loci associate (singly and in trait-specific genetic risk scores (GRS)) with longitudinal changes in total cholesterol (TC) and triglyceride (TG) levels in a population-based prospective cohort from Northern Sweden, the GLACIER Study. We also investigated the extent to which reducing extraneous model variance by adjusting for lifestyle factors (e.g. diet, physical activity, socio-economic position) affects the ability to detect associations. **Methods and Results** - A total of 6,064 participants of the GLACIER Study were genotyped with the MetaboChip array. Up to 3,495 had longitudinal (10-yr follow-up) data available. The TC- and TG-specific GRSs were strongly associated with change in lipid levels over the 10-yr period (β=0.02 mmol/l per effect allele, P=1.98/10^{-11} for TC; β=0.02 mmol/l per effect allele, P=0.0005 for TG). One TC locus, rs4420638 in *APOE* (β=0.12 mmol/l per effect allele, P=2/10^{-5}) and one TG locus, rs2954029 in *TRIB1* (β=0.09 mmol/l per effect allele, P=8.5/10^{-4}) remained significantly associated with longitudinal changes in the respective traits after correction for multiple testing, results which we are currently replicating. Additionally adjusting for lifestyle factors in the analyses did not materially change these results. **Conclusion** - In conclusion, in both cross-sectional and longitudinal analysis, the trait-specific GRSs were significantly associated with TC and TG levels. Furthermore, two SNPs were individually associated with changes in lipid levels over the 10-year follow-up period.

2126F

HAPLOTYPES OF CYTOKINE GENE POLYMORPHISMS IN DILATED CARDIOMYOPATHY. *M.L. Satyanarayana[1], V. Viswamitra[1], N. Swapna[2], N. Pratibha[1].* 1) Department of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India; 2) Department of Cardiology, CARE Hospitals, Hyderabad, Andhra Pradesh, India.

Dilated cardiomyopathy is characterized by left ventricular dilatation, impaired systolic function, reduced myocardial contractility and inefficient pumping of the heart. The symptoms include tachycardia, palpitations, syncope and a risk for sudden cardiac death, culminating into heart failure or premature deaths in children and adults. Myocarditis, systemic diseases and myocardial toxins have been identified as causative factors for DCM. Elevated levels of inflammatory cytokines have been reported in patients with DCM. However, inflammation is an important component and one of the possible mechanisms in cardiomyopathy. Hence present study is designed to study the role of polymorphisms associated with cytokines like IL-1β (-31C/T, -511C/T) and TNF-α (-308G/A). The study includes 200 healthy individuals and 105 DCM cases. IL-1β (-31C/T, -511C/T) genotyping was done by PCR based RFLP method and TNF-α by allele-specific PCR. Allelic and genotypic frequencies were estimated and control groups and appropriate statistical tests were employed. A significant association of CT genotype (C-31T) with an odds ratio of 1.69, (95% CI 1.03 - 2.77, p= 0.04) was observed in DCM. However, no significant association was observed with respect to other two polymorphisms studied. The loci combinations C-31T:C-511T and C-511T:G-308A were in perfect LD (D' = 0.99) highlighting the synergistic role of these polymorphisms in the pathology of the disease. On haplotype analysis TTA haplotype was found to exhibit 1.5 folds increased risk for DCM compared to the controls (OR 1.50, 95% CI 1.02 - 2.25, p=0.05), whereas the haplotypes CTG and CTA were found to exhibit protective effect (OR 0.47, 95% CI 0.27 - 0.83, P=0.009; OR 0.28, 95% CI 0.13 -0.60, P=0.001). The T allele of C-31T polymorphism could be the risk allele for DCM and TTA haplotype could be the risk haplotype for DCM, whereas CTG and CTA haplotypes could be protective haplotypes.

2127W

Multi-ethnic case-control studies of Moyamoya disease by high-depth exome sequencing of unrelated affected individuals and matched controls. *M. Clark[1], L. Shoemaker[2,3], A. Patwardhan[1], R. Chen[1], G. Chandratillake[1], S. Garcia[1], N. Leng[1], S. Chervitz[1], M. Pratt[1], H. Lam[1], D. Newburger[1], S. Kirk[1], C. Haudenschild[1], J. West[1], R. Chen[1], G. Steinberg[2,3].* 1) Personalis, Inc., Menlo Park, CA; 2) Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA; 3) Stanford Institute for Neuro-Innovation & Translational Neurosciences, Stanford University, Stanford, CA.

Moyamoya disease (MMD) is a rare cerebrovascular disorder characterized by stenosis/occlusion of the internal carotid artery and proliferation of microvessels. Symptoms include stroke, seizures, aphasia and cognitive impairments. It exhibits low penetrance autosomal dominant inheritance. To determine genetic associations with MMD, 125 ethnically diverse unrelated patients were matched based on gender and broad ethnic categories to 125 controls obtained from the 1000 Genomes Project. All case and control samples were exome sequenced to high depth and analyzed together using the Personalis Pipeline. Rather than using public 1000 Genomes data, case and control exomes were resequenced on the same platform and analyzed together to reduce experimental biases. Case-control analyses were performed within broad ethnic subsets (70 Asian, 136 Caucasian) to identify enriched variants. In order to investigate associations with MMD among cases in which the known founder RNF213 mutation was absent, an additional matched case-control analysis using only samples without the mutation was conducted. Variants were collapsed into genes to identify genes enriched for mutations in these MMD cases. The most highly enriched variant in Asian cases, R4810K (p=6.01×10-5), confirmed a known founder mutation in the gene RNF213 in the Asian population. Notably, this mutation was highly enriched in East Asians, but not Southeast Asian or Pacific Islander samples (p=9.52×10-4). A second variant (R4810G) was identified at the same locus as the East Asian founder mutation in one self-identified Filipino sample. Frameshift, nonsense, and splice site mutations were not observed in RNF213 in any of the case and control samples, suggesting that severe loss of function mutations in RNF213 may not be tolerated. In the Caucasian subset, the most highly enriched variant was P562L in the ZXDC gene (p=7.93×10-4), but no strong founder mutation was identified. Among cases without the RNF213 founder mutation, the collapsing method ranked OBSCN as the gene most enriched for variants (p=5.31×10-5). Our analyses leveraged high-depth exome sequencing, an internally re-sequenced multi-ethnic control set, and an analysis pipeline that identified disease-associations despite limited samples sizes. We independently confirmed the RNF213 founder mutation for MMD in the East Asian population, found it absent in other ethnicities, and identified new candidate genes via collapsing methods.

2128T

Digenic inheritance of mutations in ITGA7 and MYH7B results in congenital myopathy with left ventricular non compact cardiomyopathy.

T. Esposito[1], S. Sampaolo[2], G. Limongelli[3], A. Varone[4], D. Formicola[1,2], D. Diodato[2], O. Farina[2], F. Napolitano[1], G. Pacileo[3], F. Gianfrancesco[1], G. Di Iorio[2]. 1) Inst Gen & Biophysics, Italian Natl Res Council, Naples, Italy; 2) Department of Medical Sciences, Surgery, Neurological, Metabolic and Aging, Second University of Naples, Italy; 3) Department of Cardiological Sciences, Second University of Naples, Italy; 4) Department of Neuro-sciences, 'Santobono-Pausilipon' Hospital, Naples, Italy.

We describe an unusual association between congenital fiber type disproportion (CFTD) and left ventricular non compaction cardiomyopathy (LVNC) in the proband of an Italian family. The CFTD myopathy is a genetically heterogeneous disorder characterized by relative hypotrophy of type 1 muscle fibers compared to type 2 fibers on skeletal muscle biopsy. The hallmark features of LVNC include prominent trabeculations and deep endocardial recesses associated with arrhythmias, thromboembolic events, and heart failure. Clinical and instrumental analysis of the family members identified the LVNC phenotype also in the mother, the sister and the first-degree cousin of the proband suggesting a dominant mode of inheritance with high phenotype heterogeneity and reduced penetrance. Recessive mode of inheritance was assumed for the CFTD phenotype which is characteristic of the proband. The study was focused at the identification of the responsible gene/s through whole exome sequencing approach. Two homozygous missense mutations in two genes, the myosin heavy chain 7B (MYH7B) and the integrin alpha 7 (ITGA7) were identified. Both genes are expressed in heart and muscle tissues and both mutations were predicted as deleterious and were not found in healthy population. The R890C mutation in the MYH7B gene segregates with the LVNC phenotype in the examined family, and was also found in one not related patient affected by LVNC. The arginine at position 890 of the MYH7B gene is highly conserved in all species and this region is also conserved in MYH7 gene in which when mutated causes LVNC. The E882K mutation in ITGA7 gene, a key component of the basal lamina of muscle fibers, was found only in the proband. The glutamic acid at position 882 is highly conserved in all species and bioinformatics tools predict that when the 882E is substituted with the lysine the secondary structure of the protein changes. This study identifies two novel disease genes. Mutation in MYH7B cause a classical LVNC cardiomyopathy and further support the concept that sarcomere genes are associated with LVNC, whereas, mutation in ITGA7 gene cause CFTD. Both phenotypes represent alteration of skeletal and cardiac muscle maturation and are usually not severe. The severe phenotype of the proband is probably due to a synergic effect of these two mutations. This study provides new insights into the genetics underlying Mendelian traits showing a role for digenic inheritance in complex phenotypes.

2129F

Genetic variants associated with C-reactive protein in African Americans: a MetaboChip analysis in the Population Architecture using Genomics and Epidemiology (PAGE) study.

J.M. Kocarnik[1], C.L. Carty[1], A. Reiner[1], C.L. Avery[2], C.M. Ballantyne[3], J. Haessler[1], A. LaCroix[1], A. Young[1], T.L. Assimes[4], M. Barbalic[5], W. Tang[6], L.G. Best[7], J.L. Ambite[8], I. Cheng[9], L.A. Hindorff[10], G. Heiss[2], C.A. Haiman[11], C.L. Kooperberg[1], L. Le Marchand[12], U. Peters[1]. 1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 5) Division of Epidemiology, Human Genetics & Environmental Sciences, The University of Texas, Houston, TX; 6) Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN; 7) Missouri Breaks Industries Research Inc., Timber Lake, SD; 8) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 9) Cancer Prevention Institute of California, Fremont, CA; 10) Division of Genomic Medicine, NHGRI, NIH, Bethesda, MD; 11) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 12) University of Hawaii Cancer Center, Honolulu, HI.

Introduction: Inflammation is an important health outcome related to many common complex diseases, several of which have differential disease burden by ethnicity. C-reactive protein (CRP) is a circulating biomarker indicative of systemic inflammation. Genome-wide association studies (GWAS) have successfully identified susceptibility loci important to inflammation and inflammation-related diseases, though discovered primarily in populations of European ancestry. To test if genetic variants associated with inflammation generalize to populations of African ancestry, we evaluated the association between serum CRP levels and the large number of single nucleotide polymorphisms (SNPs) on the MetaboChip array. Methods: We analyzed 9,648 African American participants from three studies~Atherosclerosis Risk in Communities (ARIC), Multiethnic Cohort (MEC), and Women's Health Initiative (WHI)~participating in the Population Architecture using Genomics and Epidemiology (PAGE) study or the SNP Health Association Resource (SHARe) project. Participants had serum high-sensitivity CRP measurements and genotype information for the 196,725 SNPs on the MetaboChip, a customized Illumina iSelect array targeting 257 regions identified by GWAS for metabolic, atherosclerotic and cardiovascular endpoints. Fixed-effect meta-analyses combined study-specific linear regression estimates to evaluate the association between each SNP and log-transformed CRP. Additive genetic models adjusted for age, sex, top 4 principal components of genetic ancestry, and study-specific factors. Results: 28 SNPs in 5 regions reached a Bonferroni-corrected p-value cutoff of $p<3.3e-7$: CRP (10 SNPs below the correction threshold; rs3091244 had the lowest p-value at $p=4.2e-76$), TOMM40-APOE-APOC1 (8 SNPs; rs446037 $p=2.9e-16$), NOS1AP-OLFML2B (5 SNPs; rs6676438 $p=3.7e-10$), HNF1A (4 SNPs; rs7979473 $p=4.8e-10$), and LEPR (rs4655779 $p=2.6e-7$). Discussion: We identified five genetic regions that were associated with CRP levels in an African American population. One of these regions, NOS1AP-OLFML2B, has not previously been associated with CRP and may represent a novel risk locus for CRP levels. SNPs in this region were included on the MetaboChip because of previous associations with QT interval. Future analyses will further evaluate these associations in other race/ethnicity groups, compare LD patterns in these regions between these groups, and use conditional analyses to identify independent loci.

2130W

Predictive Profile for the early detection of metabolic syndrome in a pediatric Turkish cohort. *M.J. White[1,2], M. Agirbasli[3], F. Eren[4], D. Agirbasli[5], S.M. Williams[1,2].* 1) Center for Human Genetics Research, Vanderbilt Univ, Nashville, TN; 2) Department of Genetics, Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH; 3) Department of Cardiology, Marmara University School of Medicine, Istanbul, Turkey; 4) Department of Medical Biology, Marmara University School of Medicine, Istanbul, Turkey; 5) Department of Medical Biology, Acibadem University School of Medicine, Istanbul, Turkey.

Metabolic syndrome (MetS) is characterized by the presence of three or more cardiovascular disease (CVD) risk factors, including abnormalities in lipid profiles. The presence of MetS increases CVD risk by approximately two-fold and is also associated with increased mortality in adults. Recent evidence of childhood MetS underscores the impact of the global childhood obesity burden and its potential impact on long-term health. MetS prevalence rates are as high as 60 percent in some overweight pediatric/adolescent populations and autopsy studies in children and adolescents have revealed that the presence of MetS is related to the early stages of atherosclerosis. These data indicate that CVD morbidity and mortality may be increased through childhood MetS. We investigated the impact of genetic variation (single nucleotide polymorphisms - SNPs) in six candidate genes known to influence lipid profiles and circulating sex hormone levels in adults, as well as known CVD risk factors, on susceptibility to MetS in a pediatric Turkish cohort (n=360). Turkish children and adolescents have lipid profiles characterized by lower high-density lipoprotein cholesterol, elevated triglyceride levels (two hallmarks of MetS), and lower prevalence of obesity than American pediatric populations. This unique lipid profile makes Turkish children and adolescents an at risk population for MetS. Logistic regression analysis revealed that a SNP in *SHBG* (rs1799941) was significantly associated with MetS (p=0.010) after adjustment for known CVD risk factors, suggesting an independent genetic effect. Exploratory single and multi-locus analysis was performed using multi-factor dimensionality reduction (MDR) to determine the genetic and metabolic profile most predictive of MetS. MDR identified rs1799941, high triglycerides, BMI, and insulin resistance as the most predictive model (Testing balance accuracy=0.7965, p<0.001). In conclusion, our study identifies a positive association between rs1799941 and MetS, and presents a predictive model for the early detection of MetS in an at risk pediatric population. This result has the potential for future clinical use.

2131T

Investigation of functional variants of eight SNPs in lipid level modifier genes in healthy Roma and Hungarian population. *B. Melegh[1,2], K. Sumegi[1,2], L. Jaromi[1,2], L. Magyari[1,2], E. Kovesdi[1,2], B. Duga[1], R. Szalai[1], P. Matyas[1], Zs. Banfai[1], A. Szabo[1], J. Bene[1].* 1) Department of Medical Genetics, University of Pecs, Pecs, Hungary; 2) Szentagothai Research Centre, University of Pecs, Pecs, Hungary.

Purpose: Investigation of the role of triglyceride metabolism in various diseases, such as cardiovascular or cerebrovascular diseases is still in the focus of numerous studies. Genome-wide association studies reported a number of polymorphisms associated with plasma lipid level changes. The aim of our study was to investigate the distribution of eight variants: rs12130333 at the ANGPTL3, rs16996148 at the CILP2, rs17321515 at the TRIB1, rs17145738 and rs3812316 of the MLXIPL, rs4846914 at GALNT2, rs1260326 and rs780094 residing at the GCKR loci.

Methods: A total of 399 Roma (Gypsy) and 404 Hungarian population samples were genotyped using PCR-RFLP method.

Results: Significant differences were found between Roma and Hungarian population samples in the allele frequencies of GALNT2 variant (G allele frequency of rs4846914: 46.6% in Romas vs. 54.5% in Hungarians, p<0.05), of ANGPTL3 variant (T allele frequency of rs1213033: 12.2% in Romas vs. 18.5% in Hungarians, p<0.05) and of both MLXIPL variants (C allele frequency of rs17145738: 94.1% vs. 85.6%, C allele frequency of rs3812316: 94.2% in Romas vs. 86.8% in Hungarians, p<0.05), while no differences could be verified in the remaining SNPs and the known minor alleles showed no correlation with triglyceride levels in any population samples.

Conclusions: Our results may be presented as risk factors for metabolic or cardio-cerebrovascular diseases in different population samples. The current study revealed fundamental differences of known triglyceride modifying SNPs in Roma populations, however no association with triglyceride levels could be established.

2132F

Arterial Tortuosity in Patients with Vascular Abnormalities due to Mutation in Filamin A. *E. Reinstein[1], S. Morris[3], D. Rimoin[2], R. Lacro[4].* 1) Medical Genetics Inst, Rambam Health Care, Haifa, Israel; 2) Medical Genetics Inst, Cedars Sinai Medical Center, LA, CA, USA; 3) Texas Children's Hospital / Baylor College of Medicine TX, USA; 4) Boston Children's Hospital, MA, USA.

Introduction: Arterial tortuosity of the head and neck vessels, as measured by the Vertebral Artery Tortuosity Index (VTI), was recently demonstrated to be elevated in patients with Loeys-Dietz and Marfan syndromes, and correlated with adverse cardiovascular outcomes in those patients. Mutations in the Filamin A (FLNA) locus give rise to several disorders including the otopalatodigital syndrome (OPD; a bone dysplasia), X-Linked Periventricular Heterotopia (XL-PH; a disorder of abnormal neuronal migration), and Cardiac Valvular Dystrophy (CVD). Previous reports also linked mutations in FLNA to a connective tissue disorder characterized by joint, skin and vascular abnormalities, primarily aneurysms of the thoracic aorta. Methods: In this study, we evaluated a mother-daughter pair because of vascular aneurysms due to mutation in the FLNA gene. Calculation of arterial tortuosity following magnetic resonance angiography analysis was performed. Results and Discussion: Significant expansion of aortic aneurysm was seen in the mother despite normal VTI, culminating in an early aortic replacement surgery. In the daughter, a slight expansion in the size of a subclavian artery aneurysm has been observed and surgery is pending. These findings suggest that vascular surveillance is warranted in patients with arterial aneurysms due to FLNA mutations despite the presence of normal vascular turtuosity.

2133W

Phenotype, Genotype and Natural History of Arrhythmogenic Dilated Cardiomyopathy. *A. Spezzacatene[1,2], G. Sinagra[2], M. Merlo[2], G. Barbati[2], D. Slavov[1], A. Di Lenarda[2], X. Zhu[1], S. Graw[1], E. Salcedo[1], M. Taylor[1], L. Mestroni[1].* 1) Cardiovascular Institute, University of Colorado, Colorado; 2) Cardiovascular Department, Hospital and University of Trieste, Italy.

BACKGROUND Arrhythmogenic dilated cardiomyopathy (aDCM) is a form of dilated cardiomyopathy (DCM) frequently encountered in clinical practice and characterized by severe ventricular arrhythmias. The purpose of this study was to determine the prevalence, genotype-phenotype correlations and long-term outcome predictors of aDCM, to implement current criteria for risk-stratification. **METHODS** From February 1979 to November 2012, we studied 461 patients (364 families) with DCM enrolled in the Familial Cardiomyopathy Registry. Criteria for aDCM were DCM with VT, SD, AICD shock, syncope, cardiac arrest, PVC>1000/24h. A subgroup of patients was tested for DCM genes (MYH6, MYH7, MYBPC3, TNNT2, TTN, LMNA, LAP2, SCN5A, DES). For survival analysis, endpoints were: 1) death or heart transplant, 2) death, heart transplant or malignant ventricular arrhythmias (MVA), 3) MVA. **RESULTS** Over a follow-up of 96±82 months (up to 20 years), we identified 211 patients (45.8%) with aDCM: among them 72 (15.6% of the total DCM population) experienced MVA during follow-up. When compared to DCM patients, aDCM patients had worse survival (endpoint 1: P=0.006; endpoint 2: P<0.001). Independent risk factors for endpoint 1 in the aDCM population were LVEF<34% and the presence of QRS>110 ms in V1-V3. There was a cumulative risk for death or heart transplant per additional risk factor. Predictors of MVA in the aDCM population were the presence of QRS>110 ms in V1-V3 and family history of MVA. aDCM was most frequent among LMNA (70.6%), TTN (54.5%) and SCN5A variant carriers (50%). Furthermore, nuclear envelope (LMNA and LAP2) variant carriers had worse NYHA (P=0.030) in spite of smaller LVEDD at baseline (P=0.003), and experienced more heart transplants (P=0.010). TTN carriers had worse event-free survival for endpoint 2 (P=0.037). Finally, SCN5A mutation carriers were younger at enrollment (P=0.004) compared to noncarriers, had prevalence of males (P=0.029) and epsilon waves (P=0.030). **CONCLUSIONS** In a large and extensively studied DCM cohort, we define a novel subpopulation characterized by prominent ventricular arrhythmia and different prognosis. Our results suggest that aDCM may benefit from more aggressive therapeutic interventions including modified ICD criteria.

2134T

Advancing Genomic Research on Congenital Cardiac Malformations in Africa. *P. Kruszka[1], M. Muenke[1], B. Solomon[1], P. Lwabi[2], L. Harris[3], A. Beaton[4], C. Sable[4], B. Lanpher[3].* 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Uganda Heart Institute, Kampala, Uganda; 3) Department of Genetics and Metabolism, Children's National Medical Center, Washington, D.C; 4) Department of Cardiology, Children's National Medical Center, Washington, D.C.

Background: Congenital heart disease (CHD) is the most common birth defect and is the leading cause of mortality among congenital malformations. Many types of CHD are associated with syndromes; some isolated genes have been identified in non-African populations. Previously, we have identified mutations in several CHD-related genes: CFC1, GDF1, and FOXH1. However, many causes of CHD remain unknown. Our study will harness emerging genomic technologies to better understand the etiology of CHD in African and other populations. Also, genotype-phenotype analyses will characterize syndromes with known genetic etiologies that have yet to be characterized in African populations. Methods: Using comprehensive medical genetic/family history, dysmorphology exam, and echocardiogram, CHD patients were phenotyped and assigned clinical diagnoses when possible. In the future, we will conduct trio-based high-throughput genomic investigation (including exome/genome sequencing and copy number analyses). Results: From 2012-2013 we clinically evaluated 211 Ugandan patients with CHD. Forty-eight (22.7%) were clinically diagnosed with a known syndrome. Of these 48, 20 (41.7%) were suspected to have 22q11 deletion syndrome, 15 (31.3%) had convincing signs of aneuploidy, and 9 (18.8%) were diagnosed with Noonan syndrome or related Rasopathies. Other syndromes recognized included Williams, Holt-Oram, and Kabuki. Notably, truncus arterious was more frequent (2.8%) than expected from known prevalence in other populations. Conclusions: This study is anticipated to advance our understanding of CHD, both by characterizing known syndromes in this understudied population as well as revealing novel disease loci in the African cohort. Through this international collaboration, we will pilot the practice of clinically-oriented genomic medicine in a resource-limited setting, the ultimate goal of which is to assist local clinicians and researchers in their management of this complex cohort of patients.

2135F

Advantages of Next Generation Sequencing in the diagnosis of diseases associated with sudden cardiac death. *D. Cantalapiedra[1], A. Romera[1], L. Pérez-Cabornero[1], V. Felipe[1], D. Valero[1], G. Hernández[1], C. Buades[1], A. Arilla[1], C. Rodríguez[1], C. Collado[1], V. Fernández-Pedrosa[1], O. Rodríguez[1], J. Durbán[1], JC. Triviño[1], S. Zúñiga[1], A. Ballester[1], R. Miñambres[1], M. Gil[1], A. Forteza-Gil[2], M. Martínez-Atienza[3], MF. Márquez-Murillo[4], S. Santillán[1].* 1) Medical Genetics Unit, Siatemas Genómicos, SL, Paterna, Valencia, Spain; 2) Marfan Unit, Hospital Universitario 12 de Octubre, Madrid, Spain; 3) Servicio de Genética, Hospital Virgen de las Nieves, Granada, Spain; 4) Instituto Nacional de Cardiología "Ignacio Chávez", Mexico DF, Mexico.

OBJECTIVE: Genetic characterization of patients with hereditary diseases associated with sudden cardiac death, in a quick, comprehensive and cost-effective manner, using NGS resequencing panels. METHODS: We studied 172 patients (72 cases of cardiomyopathy, 21 of arrhythmogenic right ventricular dysplasia/cardiomyopathy, 27 cases of heart rhythm disorders, 37 of isolated and syndromic thoracic aortic aneurysm, 7 of idiopathic sudden death, and 8 with a family history of sudden death. Up to 90 genes associated with sudden cardiac death, were resequenced (exons, splicing junctions, 5' and 3' UTR). These regions were captured (SureSelect) and sequenced on a high-capacity NGS platform (SOLiD 5500xl). The results were confirmed by Sanger sequencing. RESULTS: We identified a total of 42 mutations (24.4% cases), which can be considered as pathogenic. The percentage of characterized cases by pathology, varies between 0% and 100% depending on the genetic heterogeneity of the condition, the degree of phenotypic overlap among pathologies, and in some cases, chance (if the number of samples is small). In 8 cases (4.6%) we identified mutations that could be considered as pathogenic, associated to a different disease than the one being stated in the clinical diagnosis of the patient. CONCLUSIONS: NGS sequencing: 1) Confirms the diagnosis. Contributes to the knowledge of new genetic mechanisms, by the identification of mutations in genes that would not have been studied by conventional techniques. 2) Identifies modifier genes that explain the intra- and interfamilial variable expressivity. 3) Helps to identify patients who require preventive measures against the risk of sudden death. 4) Helps to identify family members at risk, so that monitoring and prevention measures can be established, as well as those family members who are not at risk of sudden death. 5) Optimizes health resources, as preventive measures are offered to the actual population at risk. 6) It allows an adequate familial genetic counseling, with specific reproductive choices.

2136W

Dilated Cardiomyopathy - An updated Gene Database. *S.Justin Carlus, KM. Al Harbi.* Centre for Genetics and Inherited Diseases (CGID), Taibah University, Madinah, Saudi Arabia.

Cardiomyopathy is a rare and severe disease of the heart muscle. Dilated cardiomyopathy (DCM) is the most common form of cardiomyopathy and accounts for more than half of all cardiac transplantations in children. Mutations in 41 genes have been found to be associated with DCM and accounting for only 5- 25% genetic causes; etiologies of a larger percentage of DCM still remain unknown. To date most of the research data show only a limited number of gene tests like TTN, LMNA, MYH7, TNNT2, SCN5A, DES, MYBPC3, TNNI3, TPMI, ACTC, PLN, LDB3 and TAZ are routinely used for DCM. This is because of the lack of a comprehensive gene database available for DCM. Here, for the first time, the present gene database provides an account of 104 currently known genes, which are either directly or indirectly implicated in DCM. This updated information would help the cardiologists and the researchers to analyze more genes to screen the causative mutations in the DCM cases for early detection. The updated gene database for dilated cardiomyopathy would be available at the time of presentation.

2137T

Clinical consequences of CNVs detection in Congenital Heart Disease. *L. Kulikowski[1,2], E. Zanardo[1,2], F. Piazzon[1,2], R. Dutra[1,2], A. Dias[1,2], M. Montenegro[1,2], G. Novo-Filho[1,2], M. Basso[1,2], T. Costa[1,2], A. Nascimento[1,2], M. Grassi[2], M. Carneiro-Sampaio[2], C. Kim[2].* 1) Pathology Division- Cytogemonics Lab - LIM 03, University de Sao Paulo, São Paulo, SP,Brazil; 2) Departament of Pediatrics, The Children Institute, University de Sao Paulo, SP,Brazil.

Despite recent medical and surgical advances, congenital heart disease (CHD) remains mostly deadly birth defect, affecting nearly 1% of all newborns. And although new cytogenomics methods allowed the molecular investigation of the cardiac malformations, a limited number of studies have established the relevance of copy number variations (CNVs) in the etiology of CHD. We prospectively studied a cohort of 107 patients with CHD and concomitant no-cardiac anomalies such as facila dysmorphisms and/or developmental delay using distinct MLPA (Holland) kits and arrays (HumanCytoSNP-12, Illumina). Unexpected our analysis identified duplications (8,4%) as well as deletions (17,8%) associated with 22q11.2 region, and smaller atypical CNVs (19,6%) that implicated *IL17RA, CECR1, PEX26, PRODH, CLDN5, SHANK3, ACR* deletions. Additional genomic changes were located at 1p36, 4p16.3, 4q35.2, 5q35.3, 7p21.3, 7q36.3, 8p23.3, 9q34, 10p12.3, 11q13, 11q24, 15q11.2, 16p13.3, 17p11, 19p13.3 and 20q13.3 regions. Some of these alterations comprise genes (e.g. *CREBBP, RAI1, NKX2.5,NOTCH1,GATA4*) known for causing syndromic or isolated CHD suggesting that these CNVs are pathogenic and consistent with a model where imbalance of multiple genes associated with contributes to congenital heart disease. Also our findings showed that at least 10% of cases can be attributed to novel CNVs and that further investigation of these loci can be of great prognostic value. Application of new technologies contributes to better understanding of the role of CNVs and their effect on congenital heart phenotype. Support CNPq401910/2010-5 and FAPESP 53105/9.

2138F

Sequence variants in the mitochondrial D-loop region are associated with ventricular arrhythmias and appropriate ICD-therapy. *H. Tao, S. Rosenberg, J. wingrove.* CardioDx, Palo Alto, CA94043, CA.

Background: Sudden death from cardiac ventricular arrhythmias is a leading cause of mortality in the industrialized world. Mitochondrial dysfunction has been implicated in both heart failure and arrhythmias. Mitochondrial sequence was evaluated in a primary prevention heart failure (HF) population with implantable cardioverter-defibrillators (ICDs).

Materials and Methods: Mitochondria DNA (mtDNA) was analyzed in subjects enrolled in DISCERN, a multi-center clinical trial (www.clinicaltrials.-gov; NCT00500708) designed to identify markers associated with ventricular tachycardia or fibrillation (VT or VF). Cases were defined as subjects with adjudicated ICD therapy; controls were subjects with no documented ICD therapy for ≥ 2 years. MtDNA was initially sequenced in pooled case/control samples (Set 1; 19 subjects/pool) using the Ion Torrent PGM; the mitochondrial D-loop region was subsequently re-sequenced in individual Set 1 subjects plus an additional 40 subjects (Set 2). Melting temperatures (Tm) in candidate regions was assessed by SYBR Green RT-PCR across Sets 1, 2 and a larger set of 260 subjects (Set 3).

Results: A region within the D-loop, spanning 160-280 bp of the mitochondrial genome, showed significant sequence variant enrichment in the control pool (p=0.007). The Tm of this region was significantly elevated in individual Set 1 control samples (p=0.016); this increase was validated in two additional independent sets of subjects (Set 2: n=40, p=0.007; Set 3: n=260; p=0.0025). Sequencing of the entire D-loop region in individual Set 1 and 2 subjects showed increased sequence variant enrichment in control samples (n=78, p=0.0025); this increase was also significantly associated with lower 7s DNA copy number (p=0.004).

Conclusion: The number of variants in the mitochondrial D-loop is associated with ventricular arrhythmias and appropriate ICD therapy, and may affect D-loop formation and mitochondrial function.

2139W

Relative contributions of gene expression, genetics, and clinical factors for diagnosis of obstructive coronary artery disease. *J. Wingrove, A. Johnson, H. Tao, S. Rosenberg.* CardioDx, Palo Alto, CA.

Background - Several genetic models for predicting coronary artery disease (CAD) risk have recently been published. We previously validated a whole-blood gene expression score (GES) for obstructive CAD in patients from PREDICT, a population clinically referred for angiography. The performance of the genetic models in the PREDICT population were evaluated independently and in the context of clinical risk factors and GES.

Methods - 20 SNPs derived from 3 published models (Elosua, Kathiresan, Wells) were genotyped across 955 subjects from PREDICT (37% cases; case = ≥ 50% occlusion in ≥1 major artery by QCA). Genetic risk scores were calculated using published methods; expression levels of the 23 genes in the GES were measured and score calculated. A clinical risk score was constructed by logistic regression using age, sex, hypertension, dyslipidemia, and smoking as covariates. A model including all 20 SNPs was constructed and included in logistic regression models combining all pairs of variables as well as combining 3 variables into a single model. Genetic models, GES, and clinical risk scores were compared by AUC to assess performance. Interactions between the 20 SNPs and gene-expression levels were examined.

Results - The AUC for the published genetic models ranged from 0.55-0.59, consistent with previous data. The AUC for the GES and clinical models were similar, 0.750 (95%; CI 0.723-0.776) and 0.747 (95%; CI 0.717-0.768), respectively. The AUC increased in all models combining data types, with the highest AUC observed in the 20 SNP/GES/clinical model (AUC 0.798, 95%; CI 0.766-0.833). This model had a significantly increased AUC compared to any of the baseline models of SNPs, GES or risk factors alone (p < 0.001, each comparison). Multiple trans-acting interactions were observed between SNPs and gene expression levels.

Conclusions - The performance of 3 published genetic models for CAD risk was consistent with that previously reported. The addition of genetics to clinical factors and gene expression significantly increased the ability to detect obstructive CAD by AUC analysis.

2140T

Functional Assessment of Potential Splice Site Variants in Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy. *D. Dooijes[1], J.A. Groeneweg[2, 3], A.S. Ummels[1], M.J. Mulder[1], H. Bikker[4], J.J. van der Smagt[1], J. Post[1], J.F. van der Heijden[2], A.C. Houweling[5], J.D.H. Jongbloed[6], A.A.M. Wilde[7], J.P. van Tintelen[6], R.N. Hauer[2, 3].* 1) Dept. Medical Genetics, University Medical Center Utrecht, Utrecht, Utrecht, The Netherlands; 2) Dept. Cardiology, University Medical Center Utrecht, Utrecht, Utrecht, The Netherlands; 3) Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands; 4) Dept. Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands; 5) Dept. Clinical Genetics, VU Medical Center, Amsterdam, The Netherlands; 6) Dept. Clinical Genetics, University Medical Center Groningen, Groningen, The Netherlands; 7) Dept. Cardiology, Academic Medical Center, Amsterdam, The Netherlands.

Introduction: Interpretation of genetic screening results in Arrhythmogenic Right Ventricular Dysplasia/Cardiomyopathy (ARVD/C) is often difficult. Pathogenicity of variants with uncertain significance may be predicted by software algorithms. However, functional assessment can unambiguously demonstrate the effect of such variants. **Aim**: Functional analysis of putative splice site variants identified in ARVD/C patients. **Methods**: Eleven variants in desmosomal (PKP2, JUP, DSG2, DSC2) and non-desmosomal (TMEM43) genes with potential RNA splicing effect (predicted by SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, and Human Splicing Finder) were analyzed. The variants were found in patients with ARVD/C (fulfilment of 2010 Task Force Criteria) or suspected ARVD/C (3 patients: DSC2 c.1350A>G, JUP, TMEM43 variant). Total RNA was isolated from fresh blood samples and subjected to rtPCR. Obtained cDNA products were amplified by PCR with exonic primers specific for fragments of interest. All cDNA fragments were separated according to size using gel electrophoresis. PCR fragments were subjected to direct sequence analysis. **Results**: Of the 11 variants, 6 were intronic and 5 exonic. Eight variants, including 2 missense variants, had a functionally deleterious effect on mRNA splicing by causing exon skipping, generating new splice sites, or activating cryptic sites. All 6 intronic variants tested affected mRNA splicing. Two of 5 exonic variants severely impaired pre-mRNA processing. The remaining 3 exonic variants had no detectable effect on splicing, heterozygous presence in mRNA confirmed biallelic expression in these cases. **Conclusion**: Eight variants of uncertain significance had a functional effect on mRNA splicing, indicative of being ARVD/C related pathogenic splice site mutations. This highlights the importance of functional assessment of potential splice site variants to enhance patient care and facilitate cascade screening.

2141F

Functional analysis of collagen VI variants and their contribution to atrioventricular septal defect in Down syndrome. *C. Ackerman[1], P. Holden[2], J. Fitzgerald[2], C. Maslen[1].* 1) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR; 2) Department of Orthopedics & Rehabilitation, School of Medicine, Oregon Health & Science University, Portland, OR.

Atrioventricular septal defect (AVSD) is a common heart defect frequently associated with Down syndrome (DS). Children with DS have a 2000-fold increased risk of the defect, yet many have a normal heart indicating that trisomy 21 alone is insufficient to cause AVSD. We conducted a candidate gene study to test if rare variants in genes involved in heart development contribute to the risk of AVSD in DS. We discovered a significant excess of potentially damaging rare variants in children with DS-associated AVSD compared to children with DS and no heart defect, including mutations in the genes encoding the α1 and α2 chains of collagen VI. We hypothesized that AVSD-associated variants in *COL6A1* and *COL6A2* impact collagen VI biosynthesis disrupting microfibril formation in the extracellular matrix in the developing heart that leads to AVSD. To determine if rare variants play a pathogenic role in AVSD we analyzed the biosynthesis and trafficking of proteins expressed by three missense variants in *COL6A1* (V117A, Q786H, and R872W) and two variants in *COL6A2* (p.E106K and p.R853Q). Immunoblot analysis showed that the α1 protein expressed by the *COL6A1* variant p.R872W was undetectable in the media, and α2 proteins for both *COL6A2* variants, p.E106K and p.R853Q, were observed in reduced amounts in the media compared to wildtype (WT). In addition, the α2 protein expressed by the R853Q variant was retained in the cell, compared to WT. To test if the α1 chains produced by missense variants were triggering cellular processes activated by endoplasmic reticulum (ER) stress, we analyzed expression of HERPUD, which directs misfolded proteins to the proteosome for degradation, and CHOP, a pro-apoptotic transcription factor activated by ER stress. We found that the cells expressing α1 p.V117A, p.Q768H, and p.R872W variants expressed significantly increased levels of HERPUD than WT α1. Cells expressing α1 p.Q768H and p.R872W variants also expressed significantly higher levels of CHOP compared to WT. Our results indicate that rare variants in *COL6A1* and *COL6A2* associated with AVSD cause defective protein secretion and intracellular retention of protein chains that form collagen VI microfibrils. In addition, these variants are coupled with increased expression of proteosomal degradation and pro-apoptotic genes as a result of ER stress. Further analysis of these variants will provide mechanistic insight into how collagen VI contributes to heart defects during development.

2142W

Epigenomic regulation in the promoter of miR-210 gene affects HIFa1 binding and vascular diseases. *S.H Juo[1], K.C. Chen[1], Y.S. Wang[1], Y.C. Liao[2,3].* 1) Genome Medicine, Kaohsiung Medical Univ, Kaohsiung City, Taiwan; 2) Taichung Veterans General Hospital, Taichung, Taiwan;; 3) Department of Neurology, National Yang-Ming University School of Medicine, Taipei, Taiwan;.

microRNAs are short non-coding RNAs that can regulate gene expression by inhibiting target mRNAs. Therefore, microRNAs are considered to play a regulatory role for gene expression. However, the regulation of microRNA expression is mainly unknown. In this study, we presented the data to indicate DNA methylation can control microRNA expression. microRNA-210 (miR-210) has been reported to be involved in hypoxia-related events. Several studies suggested that miR-210 expression can be regulated by hypoxia-inducible transcription factors-1αα (HIF-1αα) during hypoxia. Furthermore, by using the bioinformatics analysis, the promoter of the miR-210 gene contains several CpG-rich regions. However, there are no reports about epigenetic regulation of miR-210 gene. We investigated the role and mechanism of miR-210 in the cardiovascular system. We first found miR-210 gene was upregulated in vascular smooth muscle cells (VSMC) and human umbilical venous endothelial cells (HUVEC) when they were exposed to oxidized low-density lipoprotein (oxLDL). By combining with 5-Aza-2'-deoxycytidine treatment or DNA methyltransferase 3b knockdown, we found that miR-210 gene expression could be epigenetically regulated by oxLDL. By using the bisulfite sequencing assay, the decreased methylation levels of miR-210 gene was identified in vitro and in vivo. This methylation change could affect the HIF-1αα-regulated miR-210 gene expression. Furthermore, over-expression or knockdown of miR-210 influenced the ability of oxLDL-mediated VSMCs and HUVECs migration. According to the bioinformatic prediction, SPRED2 (an inhibition regulator in ERK activation) could be a major miR-210's target gene involved in cell migration. Our reporter assays using wild-type and mutant 3' UTR of SPRED2 gene confirmed that SPRED2 was a novel direct target of miR-210. Over-expression or knockdown of SPRED2 affected oxLDL-mediated ERK/c-Fos/MMPs pathways in cell migration. Altogether, the data suggest that epigenetic regulation of miR-210 gene and its effect on cell migration are involved in the cardiovascular diseases formation.

2143T

Distribution of hepatic lipase C-514T and G-250A polymorphisms and their association with plasma lipid profile in young Jordanians. *O. Khabour, M. Alomari, K. Alzoubi, M. Gharaibeh, F. Alhashimi.* Medical Laboratory Sciences, Jordan Univercity of Science and Technology, Irbid, Jordan.

Hepatic lipase is an enzyme involved in the metabolism and regulation of the plasma lipoproteins. Variants in the HL gene (LIPC) may influence plasma lipoproteins levels. In this study, distribution of LIPC C-514T and G-250A polymorphisms and their associations with plasma lipid profile in young Jordanians was investigated. Genotyping of the polymorphisms in 348 young Jordanian adults was performed using polymerase chain reaction-restriction fragment length polymorphism technique. The G-250 and C-514 alleles are abundant in the Jordanian population with frequencies of 79% and 80%, respectively. However, no significant difference in lipid-lipoprotein profile (total cholesterol, triglycerides, LDL and HDL) between the different genotype groups of either C-514T or G-250A polymorphisms even when males and females were examined separately ($P > 0.05$). Thus, in young Jordanian adults, the examined LIPC polymorphisms seem to play a limited role in determination of plasma lipoprotein levels.

2144F

Novel mutation in PRKAG2 gene highlights the allosteric site of AMPK. *C. Phornphutkul[1], A. Gray[2], K. Rotondo[1], J. Padbury[3], Y. Tseng[3], D.G. Hardie[2].* 1) Pediatrics, Rhode Island Hosp, Providence, RI; 2) College of Life Sciences, University of Dundee, Dundee, Scotland, UK; 3) Pediatrics, Women and Infants' Hospital, Providence, RI.

Mutations in PRKAG2 have been implicated in hypertrophy cardiomyopathy (HCM). PRKAG2 encodes a (γ2)isoform of the nucleotide-binding regulatory subunit of AMP-activated protein kinase (AMPK), a heterotrimeric enzyme with major roles in regulation of energy metabolism in response to cellular stress. Twelve different heterozygous point mutations within the nucleotide-binding domains have been reported. De novo mutations (R531Q and R384T) were associated with severe disease and death during infancy. Case study: An abnormal 27-week prenatal ultrasound, consistent with HCM, was noted. At birth, the cardiac echo confirmed HCM. Parents had normal cardiac evaluation. Molecular testing for HCM panel was conducted and a de novo novel mutation in the PRKAG2 gene [Lys475Glu (K475E) likely disease causing] was identified in the child. At three years of age, the child has modest HCM but clinically is doing well on supportive medications. Unlike the previous cases with R531Q and R384T mutations, this case appears to have a better clinical course. To investigate the significance of the K475E mutation on AMPK complex, functional studies were performed. Method: HEK-293 stably expressing wild type (WT) or K475E FLAG-tagged γ2 from a tetracycline-inducible promoter were established. AMPK complexes containing the WT or K475E mutant γ2 were immunoprecipitated using anti-FLAG antibodies and their activities determined at varying AMP concentrations. We also compared the effects of metabolic stress induced using phenformin. Results: Our molecular studies show that the K475E mutation has three effects on the regulation of the AMPK complex: (i) markedly increasing the basal phosphorylation of Thr-172 and associated kinase activity; (ii) reducing the sensitivity to AMP in allosteric activation; (iii) preventing the increased Thr-172 phosphorylation and activity observed in response to the metabolic stressor, phenformin, in intact cells. Conclusions: K475 is conserved in γ1 and γ2 isoforms of all species, and examination of crystal structures shows that its side chain could form electrostatic interactions with the phosphate groups of AMP, ADP or ATP bound in site 1. The mutation almost abolishes allosteric activation by AMP, but also increases the basal Thr-172 phosphorylation and activity and prevents further activation by metabolic stress. Ongoing studies of the effect of these changes on downstream pathways are underway in patient fibroblasts.

2145W

Systematic cell-based functional screening for novel cardiovascular risk genes. *H. Runz[1,3,4], C. Schuberth[1,3], P. Blattmann[2,3], G. Domschke[1], A. Thormaehlen[1,3], S. Kathiresan[4], R. Pepperkok[2,3].* 1) Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; 2) European Molecular Biological Laboratories (EMBL), Heidelberg, Germany; 3) Molecular Medicine Partnership Unit (MMPU), Heidelberg, Germany; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, USA.

Elevated levels of plasma low-density lipoprotein cholesterol (LDL-C) are a causal, heritable risk factor for myocardial infarction (MI), and strategies to reduce LDL-C efficiently lower MI-risk. Sequencing and genotyping of large population cohorts are uncovering numerous genes and genetic variants with putative relevance for blood LDL-C levels and predisposition to disease. However, discovery of novel disease genes and improvement of therapies is challenged by the fact that 1) most candidate genes harbor vast numbers of potentially relevant, yet rare and often neutral variants that reduce the power of association testing; and 2) a modest mechanistic understanding if and how genes and variants therein impact on relevant phenotypes. To address these limitations, we have established technology that complements human genetics by systematic, quantitative phenotypic analyses of gene and variant function in cultured human cell models. Our unbiased, scalable functional genomics strategy is based on reducing (by RNAi-interference), increasing (by cDNA-overexpression) or complementing cellular levels of candidate genes, high-throughput functional assays, automated microscopy and multiparametric image analysis. We will present data on how we are applying this technology to systematically characterize >100 candidate genes suggested as likely associated with blood lipid levels, coronary artery disease and/or myocardial infarction for functions in regulating LDL-C uptake and turnover in cells. We will provide mechanistic insight how several genes without previously known lipid-regulatory roles may affect levels and function of the LDL-receptor and provide hypotheses how variation in these genes could predispose to disease. Our study proposes that the combination of genetics with systematically-acquired functional data in cells can pinpoint new genes responsible for altered LDL-C and MI-risk.

2146T

Using Genetic Information to Analyse and Define the Complex Aetiology of Blood Pressure. *K.H. Wade, N.J. Timpson, G. Davey Smith.* School of Social and Community Medicine, University of Bristol, Bristol, Bristol, United Kingdom.

Blood pressure (BP) is an important phenotype that has had its study hampered by complexity. The discovery of 29 SNPs, robustly associated with BP, has provided an opportunity to examine the aetiology of BP through SNP-based associations to further understand the mechanisms that maintain and alter BP at different stages of development. We aimed to assess the contribution of 29 SNPs (individually and in a weighted score) reported in the International Blood Pressure Consortium GWAS on BP in children (mean age 3-18 years). The Avon Longitudinal Study of Parents and their Children is a geographically-based birth cohort investigating factors influencing health and development of children. After data cleaning, 8,361 unrelated individuals had genome-wide information and at least one systolic and diastolic BP (SBP and DBP, respectively) measured at 11 times (mean ages: 3- 5, 8, 10-15, and 18). We examined the associations between SNPs and BP at different ages in the same samples cross-sectionally and longitudinally, and assessed whether the overall contribution of SNPs changed with age. Cross-sectionally, the score was associated with SBP at ages 8-15 years and with DBP at ages 8-18 years. For example, the score was associated with a 12.67mmHg increase (95%CI: 4.88, 20.46; p=0.001) in SBP at age 11 and an 18.87mmHg (95%CI: 8.18, 29.56; p=0.001) increase in SBP at age 15. Individually, few SNPs showed associations with BP cross-sectionally. At age 11, rs7129220 was associated with a 0.68mmHg (95%CI: 0.12, 1.25; p=0.018) increase in SBP and rs1799945 was associated with a -0.51mmHg change (95%CI: -0.99, -0.04; p=0.035). Longitudinally, age was strongly associated with a 0.0377mmHg (95%CI: 0.0372, 0.0382; p<0.0001) increase in SBP and a 0.0138mmHg (95%CI: 0.0134, 0.0142; p<0.0001) increase in DBP, with no evidence of interaction between score (separated into centiles) and age (p=0.890 for SBP, p=0.223 for DBP). However, there was evidence of interaction between 6 SNPs and age (p=0.031 for rs419076; 0.048 for rs805303 and DBP; 0.011/0.04 for rs932764 and SBP/DBP, respectively; 0.036 for rs1327235 and SBP; 0.016 for rs3774372 and SBP; 0.032 for rs17249754 and SBP; and 0.02 for rs12940887 and SBP). We have replicated the association between 29 SNPs and BP in a large sample of children aged 3-18 years. We aim to further assess the patterns of association and whether these conform to biologically relevant groups, where the complexity of BP aetiology may be explained.

2147F

Genome-wide meta-analysis of homocysteine and methionine metabolism identifies five one carbon metabolism loci and an association of *ALDH1L*1 with ischemic stroke. *S.R. Williams[1,2], Q. Yang[3,4], F. Chen[1], X. Liu[3], K.L. Keene[1,5,6,7], P. Jacques[8], W.M. Chen[1], G. Weinstein[9], F.C. Hsu[10], A. Beiser[3,9], L. Wang[11], E. Bookman[12], K.F. Doheny[13], P.A. Wolf[9], M. Zilka[13], J. Selhub[8], S. Nelson[14], B.B. Worrall[5,15,16], S. Seshadri[9], M.M. Sale[1,16,17].* 1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Cardiovascular Research Center, University of Virginia, Charlottesville, VA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) The Framingham Heart Study, Framingham, MA; 5) Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 6) Department of Biology, East Carolina University, Greenville, NC; 7) Center for Health Disparities Research, East Carolina University, Greenville, NC; 8) Jean Mayer USDA Human Nutrition Research Center on Aging and Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA; 9) Department of Neurology, Boston University School of Medicine, Boston, MA; 10) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 11) 11Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic College of Medicine, Rochester, MN; 12) National Human Genome Research Institute, 5635 Fishers Lane, Bethesda, MD; 13) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 14) Department of Biostatistics, University of Washington, Seattle, WA; 15) Department of Neurology University of Virginia, Charlottesville, VA,; 16) Department of Medicine, University of Virginia, Charlottesville, VA; 17) Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA.

Circulating homocysteine levels (Hcy), a product of the folate one carbon metabolism pathway (FOCM), are heritable and have been associated with an increased risk of stroke and dementia. However, the genetic determinants of elevated Hcy, hyperhomocysteinemia, are poorly understood. Variants located in genes involved in the FOCM have been consistently linked to stroke risk but the underlying biological processes remain unclear. We conducted independent genome wide association studies and a meta-analysis of methionine metabolism, characterize by post-methionine load test Hcy levels(ΔPOST), in 2710 participants from the Framingham Heart Study (FHS) and 2100 participants from the Vitamin Intervention for Stroke Prevention (VISP) clinical trial, and then examined the association of the identified loci with incident stroke. Five genes in the FOCM pathway (*GNMT* [1.60×10^{-63}], *CBS* [3.15×10^{-26}], *CPS1* [9.10×10^{-13}], *ALDH1L1* [7.3×10^{-13}] and *PSPH* [1.17×10^{-16}) were associated with the difference between pre- and post-methionine load test tHcy levels (ΔPOST). *GNMT*, the Glycein N-Methyltransferase gene, is responsible for the demethylation of of S-adenosyl-L-methionine to give rise to S-adenosyl-L-homocysteine. This step affects the availability of all methyl groups to the cell and may have a major impact on the epigenetic state of the whole genome. Additionally, one variant in the *ALDH1L1* locus, rs2364368, was also associated with incident ischemic stroke in FHS (p = 0.02). Analysis indicates that circulating homocysteine levels do not correlate well with POST or ΔPOST, demonstrating the need for investigation of the true measure each test explains. Haplotype differences indicate that the 5' *GNMT* regulatory region may play a major role in differential methionine metabolism. Followup promoter analyses and pyrosequencing of the *GNMT* gene reveal genetic and epigenetic mechanisms that explain *GNMT* transcription differences and likely impact methionine metabolism. Finally, a risk-score consisting of the 5 loci, explains 13% of the variance of ΔPOST in FHS and 6% of the variance in VISP. Association between variants in FOCM genes with ΔPOST suggest a novel mechanism, altered transcriptional regulation, by which variation in methionine metabolism occurs affecting parallel pathways and stroke risk. These data emphasize the importance of a concerted effort to understand regulators of one carbon metabolism in future studies and their potential as therapeutic targets.

2148W

Analysis of vWA and TPOX loci reveals a highly associated relationship with thrombosis disease in Mexican Mestizo population. *R. Camacho[1], A. Majluf[2], G. Noris[3], C. Santana[3], M.A. Meraz[4], J. Hernández[2], R. Gómez[1].* 1) Departamento de Toxicologia, Cinvestav-IPN, México D.F; 2) Unidad de Investigación Médica en Trombosis y Aterogénesis, Instituto Mexicano del Seguro Social, México D.F; 3) Laboratorio BIMODI (Biología Molecular Diagnóstica), Querétaro, Qro; 4) Departamento de Biomedicina Molecular, Cinvestav-IPN, México D.F.

Thrombosis disease is an important target for biomedical research due to its high morbidity and mortality around the world. Thrombosis is a complex disease and several risk factors are evolved (age, trauma, hormonal alterations, immobility, obesity, diabetes type 2, and hypertension) as well as environmental factors. In addition, genetic factors are strongly associated with this disease. Recent studies suggest that ethnicity is considered an important risk factor in thromboembolism incidence, showing a variable frequency of the disease among different ancestries. In this study we identify truly genetic biomarkers, type short tandem repeat (STRs), associated with thrombosis in a population with multiethnic backgrounds from Mexico. Genotypes were obtained using a validated method of genotyping from patients (n=177) with venous thrombosis. The allelic frequencies were compared with non-related healthy individuals (n=531). Bayesian methods were used to correct population stratification in order to avoid spurious associations due to admixture. We found that vWA-18 (OR=1.5), TPOX-9 (OR=17), and TPOX-12 (OR=2.3) alleles were significantly associated with thrombosis disease, showing accurate confident intervals. This association remains after admixture correction, confirming a strong genetic association (P<0.0001). In addition, the combination of allele 18 (vWA) with allele 12 (TPOX), shows a significant difference (P<0.05) with high accurate CI95% (1.02 - 3.64), suggesting that this combination could be associated with an increase of two times of the thrombosis risk. In summary, our data propose that vWA and TPOX are good biomarkers, which may improve in the diagnostic methods reducing significantly treatment costs by providing preventive and personalized medicine.

2149T

GALNT2, a HDL cholesterol and triglyceride candidate gene, regulates lipoprotein metabolism in vivo. *S. Khetarpal[1], A. Edmondson[1], A. Raghavan[1], S. Kathiresan[2], D. Rader[1].* 1) Dept Med, Univ Pennsylvania, Philadelphia, PA; 2) Broad Institute and Massachusetts General Hospital, Boston, MA.

Several genome-wide association studies (GWAS) for blood lipids have uncovered novel genomic loci influencing these traits. One such locus associated with high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG) is that harboring GALNT2 on chromosome 1q42. GALNT2 encodes GalNAc-T2, an enzyme catalyzing the initiation of O-glycosylation of specific protein targets. We previously demonstrated an in vivo role for GALNT2 in lipoprotein metabolism through GALNT2 overexpression and knockdown in mice using adeno-associated virus vectors. To further study the mechanism through which GALNT2 regulates plasma lipids, we have generated whole-body and liver-specific GALNT2 knockout (KO) mice. Whole-body GALNT2 deficiency resulted in elevated VLDL-TG levels compared with wild-type and a moderate reduction in HDL-C (20%, P < 0.05). GALNT2 KO mice did not exhibit differences in post-prandial TG clearance or VLDL-TG secretion. Liver-specific deletion of GALNT2 reduced HDL-C (26% decrease relative to WT, P < 0.01), and reduced numbers of HDL particles (19% decrease relative to WT, P < 0.01) with minimal changes in plasma TG or dietary TG clearance. Our data further implicate GALNT2 as an important regulator of plasma lipoprotein metabolism with tissue-specific regulatory roles in vivo.

2150F

Disease-relevant pathways modulate a cis-regulatory element at the TCF21 coronary heart disease locus. *C. Miller[1], D. Anderson[1], R. Kundu[1], A. Raiesdana[1], S. Nürnberg[1], R. Diaz[1], N. Leeper[1], E. Schadt[3], C. Hsiung[2], T. Assimes[1], T. Quertermous[1].* 1) Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 2) Division of Biostatistics and Bioinformatics, National Health Research Institutes, Zhunan, Taiwan; 3) Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, NY.

Coronary heart disease (CHD) is the leading cause of mortality in both developed and developing countries worldwide. Genome-wide association studies (GWAS) have now identified 46 independent susceptibility loci for CHD, however the biological and disease-relevant mechanisms for these associations remain elusive. The large-scale meta-analysis of GWAS recently identified in Caucasians a CHD-associated locus at chromosome 6q23.2, a region containing the transcription factor TCF21 gene. TCF21 (Capsulin/Pod1/Epicardin) is a member of the basic-helix-loop-helix (bHLH) transcription factor family, and regulates cell fate decisions and differentiation in the developing coronary vasculature. Herein, we characterize a cis-regulatory mechanism by which the lead polymorphism rs12190287 disrupts an atypical activator protein 1 (AP-1) element, as demonstrated by allele-specific transcriptional regulation, transcription factor binding, and chromatin organization. Further, this element is shown to mediate signaling through platelet-derived growth factor receptor beta (PDGFR-β) and Wilms tumor 1 (WT1) pathways. A second disease allele identified in East Asians also appears to disrupt an AP-1-like element. Thus, both disease-related growth factor and embryonic signaling pathways may regulate CHD risk through two independent alleles at TCF21.

2151W

A functional polymorphism that affects the *APOA5* gene expression determines plasma triglyceride levels conferring coronary atherosclerosis risk in Han Chinese Population. *W. Shou[1], F. Xie[1], Y. Wang[1], B. Wang[1], Z. Wang[1], J. Shi[1], W. Huang[1,2].* 1) Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai, China; 2) Rui Jin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China.

Elevated plasma triglyceride (TG) level has been established as an independent risk factor for multiple cardiovascular diseases. Thus, it is clinically important to identify genetic determinants of plasma TG levels. *APOA5* in 11q23.3-q23 gene cluster, the top hit of the TG-associated loci, encodes an apolipoprotein APOA5, which promotes lipoprotein lipase mediated plasma TG hydrolysis. A polymorphism rs2266788 in 3' untranslated region of the *APOA5* gene was identified significantly associated TG levels, but was underrepresented compared to the lead variants in genome-wide association studies. So, we investigated the association between rs2266788 and plasma TG regulation, and between rs2266788 and coronary artery disease (CAD) incidence in Han Chinese population. rs2266788 was genotyped in 3222 unrelated subjects consisting of 2,062 CAD cases and 1,160 matched controls. *APOA5* mRNA expression level and allelic expression imbalance (AEI) was measured in 89 normal liver specimens. Luciferase assays were conducted in BEL-7402, BEL-7405, Huh7 and HepG2 cell lines to compare transcription activity of two alleles. Through multiple linear regression analysis, a significant association was observed between rs2266788 and plasma TG levels in the total cohort ($P = 1.02 \times 10^{-13}$), and the association remained strong ($P = 5.17 \times 10^{-13}$) even after adjustment for age, sex and CAD status. Plasma TG levels were higher in samples with C allele than T allele. In stratified analyses, the association was all significant in controls, in cases, in females and in males. In logistic regression analyses adjusted for age and sex, rs2266788 was significantly associated with CAD in total samples (OR = 1.151, 95% CI =1.012-1.309, $P = 0.033$), and in females (OR = 1.397, 95% CI =1.139-1.714, $P = 0.001$). The C allele conferred an increased risk for CAD development. Analyses of the mRNA expression levels among different genotypes and AEI of *APOA5* in the liver samples showed C allele had a significantly lower transcription activity than T allele. Concordantly, the mean transcription activity of C allele in different cell lines was 70% that of T allele in luciferase assays. We provided strong evidence that functional rs2266788 was significantly associated with plasma TG levels conferring CAD risk in Han Chinese population due to its *cis*-acting effect to the gene expression.

2152T

Use of allele-specific FAIRE for identification of functional variants at cardiometabolic loci. *A.J.P. Smith[1], J. Palmen[1], E. Romeo[1], F. Drenos[1], P. Howard[1], A.D. Hingorani[2], P.J. Talmud[1], S.E. Humphries[1].* 1) Cardiovascular Genetics, University College London, London, United Kingdom; 2) Genetic Epidemiology, University College London, London, United Kingdom.

GWAS for cardiometabolic diseases have identified many novel loci for further analyses, although identification of functional, non-coding variants at the majority of these loci remains limited. Using FAIRE followed by allele-specific analysis with the Cardiometabochip array, we have examined GWAS loci for effects on chromatin accessibility as a marker of regulatory potential, to differentiate functional variant/s from non-functional variants in strong LD. Using this methodology, we have localised functional SNPs at several loci. We confirm the functionality of two variants at the 9p21 locus, previously implicated in coronary artery disease and type 2 diabetes susceptibility, and describe their effects on protein-DNA interactions, reporter gene expression and biomarkers.

2153F

Identification and characterization of patients with autosomal dominant hypercholesterolemia caused by gain-of-function mutations in proprotein convertase subtilisin/kexin type 9 and comparison with patients with Familial Hypercholesterolemia (FH) and Familial Defective apolipoprotein B (FDB). *G.D. Swergold[1], S.W. Fouchier[2], B. Sjouke[2], J. Mendoza[1], S. Mellis[1], S. Hamon[3], J.C. Defesche[2] on behalf of the PCSK9 Natural History Study (NHS) Group.* 1) Regeneron Pharmaceuticals, Inc., Tarrytown, NY, USA; 2) Department of Vascular Medicine of the Academic Medical Center at the University of Amsterdam, The Netherlands; 3) Statistical and Genetic Consulting, LLC, Darien, CT, USA.

Autosomal Dominant Hypercholesterolemia (ADH) is a common disorder of lipid metabolism characterized by high levels of serum LDL cholesterol (LDLC), and early onset cardiovascular disease. The most frequent mutations causing ADH are found in the LDL receptor (LDLR; FH) or its ligand apolipoprotein B (ApoB; FDB). Gain-of-function mutations (GoFm) in proprotein convertase subtilisin/kexin type 9 (PCSK9), a potent modulator of LDLR on hepatocytes, appear in ~2% of patients with ADH. To date, relatively few patients with PCSK9 GoFm have been described, and their clinical syndrome incompletely characterized. To better understand the geographic and familial distribution of PCSK9 GoFm, their clinical manifestations, and their comparison to FH and FDB we conducted a retrospective, cross-sectional parallel-group observational cohort study. Patients with PCSK9 GoFm confirmed by molecular testing were matched for age and sex with patients with molecularly proven FH and FDB. Data collected included baseline and on-treatment lipid profiles, the presence of xanthoma, xanthelasma, and corneal arcus, and the occurrence and age of onset of CVD. We initiated 200 site contacts and collected data on 164 PCSK9 GoFm patients (83 men, 81 women) from 12 sites in France, Japan, Norway, Portugal, South Africa, The Netherlands, the UK, and the USA. We matched these patients with 2126 patients with FH and 470 with FDB. We characterized the LDLR mutations as "defective" (missense, small in-frame indel, synonymous with added splice site) or "deficient" (large or frame-shifting indel, nonsense, splice site, promoter snp) and compared their lipid profiles. In general, individual PCSK9 GoFm appeared to have restricted geographic distributions and to be concentrated within a small number of pedigrees. Examples include 22 patients with R215H found only in 2 pedigrees in Norway, and 12 patients with V4I and 30 patients with E32K found only in Japan. These data support a similar (recent origin) history for PCSK9 GoFm and LDLR FH mutations that is distinct from ApoB FDB mutations. Mean baseline LDLC was highest in patients with PCSK9 GoFm and lowest in those with FDB. Of the patients with PCSK9 for whom data were available, 53% had evidence of xanthoma, and 33% had a history of CAD. Index patients with either FH or FDB had higher baseline LDLC levels than family members and patients with deficient LDLR mutations had higher baseline LDLC than those with defective mutations.

2154W

***GATA4* is implicated in the pathogenesis of neonatal and childhood-onset diabetes.** *E. De Franco[1], C. Shaw-Smith[1], H. Lango Allen[1], S. Flanagan[1], M. Borowiec[2], W. Mlynarski[2], M. Batlle[3], J. Ferrer[3], A. Hattersley[1], S. Ellard[1].* 1) University of Exeter Medical School, Exeter, United Kingdom; 2) Department of Paediatrics, Oncology, Haematology and Diabetology, Medical University of Lodz, Lodz 91-738, Poland; 3) Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), Barcelona, Spain.

Recent studies have shown that GATA6 haploinsufficiency is the most common cause of pancreatic agenesis and can also result in diabetes diagnosed outside the neonatal period. GATA4 is a transcription factor closely related to GATA6 and is known to be involved in the first stages of pancreatic development in mouse models: conditional knock-out of Gata4 in mouse embryos causes agenesis of the ventral pancreas. In humans GATA4 has been long regarded as a candidate regulator of pancreatic development but the evidence reported so far has been inconclusive. A single patient with a congenital heart defect and pancreatic agenesis harboring a GATA4 missense mutation has been reported, but causality was not proven. Similarly to GATA6, GATA4 mutations and deletions are a known cause of congenital heart defects.

We report four patients with diabetes, congenital heart malformations, and deletions including the GATA4 gene locus. In three cases diabetes presented in the neonatal period (age at diagnosis: 1-7 days) and the fourth case was diagnosed at 13 years. In addition we identified a fifth patient with a de novo GATA4 missense mutation (p.Asn273Lys) with complete absence of the pancreas confirmed post mortem. This mutation affects a highly conserved residue located in the second zinc finger domain of the GATA4 protein. Functional in vitro studies showed reduced DNA binding and transactivational activity of the mutant protein.

These results suggest that GATA4 haploinsufficiency is a cause of neonatal or childhood-onset diabetes with or without pancreatic exocrine insufficiency. GATA4 mutation testing should therefore be considered in patients with neonatal/childhood onset diabetes and congenital heart defects in whom no GATA6 mutation has been identified.

2155T

SMAD3 haploinsufficiency has a causative role in development of TAAD. *L. Gong[1], X. Duan[1], P. Yang[2], X.F Wang[2], O.A Moffitt[1], D.M Milewicz[1].* 1) Department of Internal Medicine, University of Texas Health Science Center at Houston, 6431 Fannin, MSB 6.100, Houston, TX 77030, USA; 2) Department of Pharmacology and Cancer Biology, Duke University, Durham, North Carolina, USA.

Thoracic aortic aneurysms leading to acute aortic dissections (TAAD) can be inherited in families in an autosomal dominant manner with reduced penetrance and variable clinical expression. Mutations in genes in the TGF-beta signaling pathway are believed as the causes of the disease in these families. The majority of these mutations are predicted to decrease TGF-beta signaling but there is evidence of increased signaling based on phosphorylated Smad2 (pSmad2) in aortic tissue of patients with these mutations. SMCs were explanted from a patient with SMAD3 R279K and Smad3-/- mice to investigate TGF-beta signaling. After exposure to TGFbeta1, the patient's SMCs showed decreased phosphorylation of Smad3 but increased pSmad2 compared with wild-type (WT) SMCs and the Smad3-/- SMCs also increased and sustained pSmad2 signaling with exposure to TGF-beta1. When treated with SB-431542, a small molecule inhibitor of the type I TGF-beta receptor, the pSmad2 signaling on Smad3-/- SMCs was reduced. Both Smad3-/- and SMAD3R279K SMCs also had reduced expression of SMC contractile proteins compared with WT SMCs, along with decreased expression of other genes known to be dependent on Smad3 signaling, including Col1A1 and Col3A1. The expression of DCN in the patient's SMCs was increased. Based on transfection of mutant SMAD3R279K, WT with SMAD4, it was determined that the mutant Smad3 could not complex with Smad4 but smad3 WT could. In both the transfected and SMAD3 R279K SMCs, SMAD3 remained in the cytoplasm 1 hour after exposure to TGFbeta1, but with WT transected or WT SMC, SMAD 3 was in the nucleus. The increased tgfbr3, which increases tgfr1 signaling, may be responsible for the increased smad2 signaling. Therefore, SMAD3 mutations disrupt Smad3 signaling in SMCs thus disrupting expression of contractile proteins and collagens. At the same time, SMCs have increased signaling through smad2, including increased proteoglycan production. Further studies will determine if the aortic disease results from loss of Smad3 signaling or increased Smad2 signaling.

2156F

Tnfaip8 is Involved in Host Response to _Staphylococcus aureus_ Infection in Mice and Humans. *Q. Yan[1], S.H. Ahn[2], B.K. Sharma-Kuinkel[1], C.L. Nelson[3], W.K. Scott[4,5], A.S. Allen[3,6], T.H. Rude[1], V.G. Fowler Jr[1,3].* 1) Department of Medicine, Duke University Medical Center, Durham, NC; 2) Department of Biochemistry School of Dentistry, Chonnam National University, Bukgu, Gwangju, Korea; 3) Duke Clinical Research Institute, Duke university Medical Center, Durham, NC; 4) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 5) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 6) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC.

Background: We previously showed that *Tnfaip8* was associated with susceptibility to *S. aureus* infection in susceptible (A/J) but not resistant (C57BL/6J) mice (Ahn, PLOS Pathogens 2010). However, the role of *Tnfaip8* in host response to *S. aureus* is unknown. **Methods:** RNA and protein expression profiles of both variants of *Tnfaip8* in both *S. aureus*-susceptible (A/J) and resistant (C57BL/6J) mice were evaluated by real-time PCR and western blot. Expression profiles of *Tnfaip8* in patients with bloodstream infection (BSI) due to *S. aureus* (n=32) or *E. coli* (n= 19) or healthy subjects (n=44) were evaluated with whole genome expression data. **Results:** Levels of *Tnfaip8* RNA were respectively 34% (variant 1) and 13% (variant 2) lower in susceptible vs. resistant mice blood (n=6). By western blot lower levels of TNFAIP8 protein was confirmed in A/J mice, but the protein signal was predominated in immune-related tissues (e.g., kidney, spleen, lymph node) in both mouse strains. Both variants of *Tnfaip8* were highly expressed in both susceptible and resistant mice blood. However, when challenged with *S. aureus* or *E. coli*, whole blood expression of *Tnfaip8* variant 1 decreased to non-detectable level in susceptible mice but had more than 40% increase in resistant mice (n=6). Expression of *Tnfaip8* variant 2 decreased about 15% in susceptible A/J mice blood, but increased more than 11% in resistant C57BL/6J mice (n=6). Similarly, susceptible patients with *S. aureus* BSI had 15% decrease (p=0.0005) of expression, while *E. coli* BSI exhibited 18% ($p<0.001$) decrease of *Tnfaip8* as compared to healthy human subjects. **Conclusions:** These findings further support the importance of *Tnfaip8* in immune response to *S. aureus*, and suggest that both variants of *Tnfaip8* may participate in this function.

2157W

Association of eNOS 4 a/b polymorphism in Mexican patients with coronary artery disease. *M. Gallegos Jr[1], RP. Mariaud[2], LE. Figuera[3], AM. Puebla[4], GM. Zúñiga[5].* 1) Dept Med Molec, Guadalajara, CIBO, IMSS, Jalisco, Mexico; 2) Departamento de Clínicas Odontológicas Integrales, Centro Universitario de Ciencias de la Salud, Universidad Guadalajara; 3) Division de Genética, CIBO, IMSS; 4) Laboratorio de Inmunofarmacología, CUCEI, Universidad de Guadalajara; 5) Laboratorio de mutagénesis,-CIBO, IMSS.

Coronary artery disease (CAD) is a major cause of illness and death, affecting more than 7 million people in the world, their behavior is multifactorial with genetic and environmental predisposition possible. Several polymorphisms in the gene for endothelial nitric oxide synthase (eNOS) have been associated with the development of CAD, one of the most studied has been the 4a/ b, and however the results have been contradiction in different parts of the world. The aim of this study was to determine the association of eNOS 4a /b polymorphism in CAD patients with stenosis coronary positive. Were Included 114 patients with CAD and stenosed coronary and 116 controls from the general population, by the PCR was amplified the intron 4 of the eNOS gene. The allele identification was performance by polyacrylamide gel electrophoresis after staining with silver nitrate. The observed genotype frequencies for controls and CAD patients with stenosis coronary positive were 1% and 17% for a/a, 15% and 42% for a/b, 1% for b/c only control group, and finally 83% and 41% for b/b respectively. The frequencies of the eNOS 4 a/b polymorphism showed statistically significant differences when comparing the study groups ($p < .05$). We conclude that the genotypes a/a-a/b of the eNOS 4 a/b polymorphism contribute significantly to CAD patients with stenosis coronary positive susceptibility in the analyzed sample from the Mexican population.

2158T

KLF15 regulates circadian cardiac susceptibility to ischemia reperfusion injury. *L. Zhang[1], D. Prosdocimo[2,3], M. Jain[2,3].* 1) Dept Med/Human Gen, Case Western Univ SOM, Cleveland, OH; 2) CVRI, Case Western Univ SOM, Cleveland, OH; 3) HHVI, Univ Hosp Case Med Ctr, Cleveland, OH.

It is well recognized that heart attacks have a diurnal variation in their incidence and severity. However, molecular mechanism of the regulation of circadian susceptibility to ischemia/reperfusion (I/R) injury is largely unknown thus preventing effective intervention. Kruppel-like-factors (KLFs) are zinc finger transcriptional factors that serve key roles in cardiovascular biology, in particular, we have previously identified that in multiple cardiovascular diseases, including astherosclerosis and heart failure, *KLF15* expression is reduced. Interestingly, its expression also oscillates in a circadian fashion under the direct regulation of core clock gene, *BMAL1*. Reactive oxygen species (ROS) have long been recognized as a major insult in I/R injury, however, antioxidants have been disappointing in multiple clinical trials. We hypothesize that cellular activities that generates ROS, such as metabolism need to be meticulously coordinated with that dispose ROS in a temporal fashion, so that homeostasis can be maintained despite the dynamic activities of the organism in a 24 hour period. We demonstrate that Klf15 deficiency leads to increased cellular oxidative stress. In primary myocardium cell culture and in vivo I/R model, Klf15 deficient cardiomyocyte/myocardium has increased susceptibility to I/R injury. We also show that Klf15 deficiency affects the myocardium lipid-flux program and is essential for cardiac metabolic homeostasis during fasting (most prominent during sleep to wake transition). Additionally, we have identified and validated multiple ROS clearance enzymes involved in the I/R injury of the myocardium as direct targets of KLF15. Based on these observations, we propose that KLF15 is a hub coordinator of circadian regulation of metabolism and ROS clearance and a key determinant of the temporal susceptibility of myocardium to I/R injury in the heart. Further understanding of the precise molecular mechanism of this regulation will facilitate the discovery of novel preventative therapy against cardiac I/R injury.

2159F

Interactions among Depressive Symptoms and Genetic Influences on Cardiac Outcomes. *L. Frazier[1], J. Sanner[2], T.-Y. Yu[2], A. Morrison[3], E. Boerwinkle[3].* 1) College of Nursing, University of Arkansas for Medical Sciences, Little Rock, AR; 2) School of Nursing, University of Texas Health Science Center, Houston, TX; 3) School of Public Health, University of Texas Health Science Center, Houston, TX.

INTRODUCTION Depressive symptoms are associated with increased inflammatory protein levels but only in certain individuals. In a prospective study of patients with acute coronary syndrome (ACS), we tested a biobehavioral model in which inflammatory protein gene polymorphisms interact with depression, and we predicted greater increases in inflammatory protein levels will be observed compared to the levels caused by either gene polymorphisms or depression alone. The purpose of the proposal was to determine a well-defined, high-risk subgroup of ACS patients in which the interaction between depression and genetic polymorphisms increases the risk of subsequent major adverse coronary events (myocardial infarction, revascularization procedures, stroke, and death) more than either factor alone, in part because of their combined effect on increasing inflammatory protein levels. **HYPOTHESIS** We hypothesized that genes with a depression-by-gene interaction effect on inflammatory protein level will also have a depression-by-gene interaction effect on risk of subsequent major adverse coronary events. **METHODS** We enrolled ACS patients from two large tertiary care centers and obtained blood samples to measure inflammatory protein levels immediately after hospital admission. Inflammatory proteins and genes measured include Interleukin (IL) 6, C-reactive Protein (CRP), Tumor Necrosis Factor Alpha (TNFα), E-Selectin (SELE), and Monocyte Chemoattractant Protein-1 (MCP-1). Patients were screened for depression using Beck Depression Inventory-II scores. Demographic and clinical risk factors were collected. The hypothesis will be tested using logistic regression and survival analysis. **RESULTS** Study results are pending data analysis and are currently underway. **CONCLUSIONS** We will present study findings and discuss how the discovery of a relationship among depressive symptoms, genetics, and inflammatory protein levels in a subgroup of ACS patients provides a rationale for studying environmental triggers of depressive symptoms and the effects of depression interventions such as different medications, psychotherapies, treatment combinations, and self-management techniques like exercise on inflammatory protein levels. The results may lead to specific secondary intervention methods that consider the integrated response of the patient to inflammatory triggers and provide support for the role of genetics in modifying the relationship between depression and adverse coronary events.

2160W

A genome-wide interaction study identifies *PDE1C* genetic variants and renal function as jointly associated with coronary artery disease.
C. Ward-Caviness[1], M. Winn[1,2], C. Blach[1], C. Haynes[1], E. Dowdy[1], S. Gregory[1], S. Shah[1,2], W. Kraus[1,2], E. Hauser[1]. 1) Center for Human Genetics, Duke Univ Med Ctr, Durham, NC; 2) Department of Medicine, Duke Univ Med Ctr, Durham, NC.

Cardiovascular disease (CVD) and chronic kidney disease (CKD) are significant causes of mortality in developed nations. CAD and CKD often co-occur, and each has a well-established a genetic basis. However potential interactions between genetic variants (SNPs) and markers of CKD that may influence cardiovascular outcomes have not been explored. We used CATHGEN, a large cardiac catheterization cohort, to perform a genome-wide interaction study that estimated associations between cardiovascular disease, indexed by the number of diseased coronary vessels, and SNP by estimated glomerular filtration rate (GFR) interactions. We used race-stratified cohorts of 2202 whites (EA) and 663 blacks (AA) followed by meta-analysis. A total of 905,956 SNPs were analyzed using a cumulative link model with a logit link. Thus we modeled the log-odds for a change in the number of diseased coronary vessels (our dependent variable/outcome), given a particular SNP-GFR interaction (our independent variable of interest). We adjusted for age, sex, BMI, hypertension, hyperlipidemia, diabetes, smoking, and race-specific principal components to adjust for potential ethnic stratification within each race-stratified cohort. We identified the 10 strongest SNP-GFR interactions for each cohort and looked for those that validated ($P<0.05$ and consistent direction of association between cohorts), or were among the strongest meta-analysis results, ranked by p-value. Using these criteria we identified rs10951304, an intronic *PDE1C* variant ($P=1.09\times10^{-04}$ AA, $P=3.51\times10^{-03}$ EA, $P=9.77\times10^{-06}$ meta-analysis). Another intronic *PDE1C* variant, rs2058411 (AA $P=1.32\times10^{-05}$), was among the 10 most significant AA associations, however this SNP did not have $P<0.05$ in the EA. *PDE1C* is a phosphodiesterase that hydrolyzes cAMP and cGMP, two important cell signaling molecules. *PDE1C* is responsible for a significant portion of cAMP hydrolytic activity in the heart, particularly in smooth muscle cells cultured from atherosclerotic lesions. These analyses take a novel approach to understanding the co-occurrence of CVD and CKD. Using our genome-wide interaction approach we have identified genetic variants in a biologically relevant gene that interact with GFR and are associated with clinically significant atherosclerosis in coronary arteries.

2161T

A 62 variant type 2 diabetes genetic risk score is not associated with subclinical atherosclerosis in the Framingham Heart Study - SNP Health Association Resource (The FHS SHARe Study).
M. Dauriz[1,2,3], B. Porneala[1,2], J.L. Vassy[2,4,5], M.F. Hivert[1,2,6], J.M. Murabito[7,8], J.C. Florez[9,10], J. Dupuis[8,11], C.J. O'Donnell[8,12], J.B. Meigs[1,2]. 1) Massachusetts General Hospital, General, Medicine Division, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA, USA; 3) Division of Endocrinology and Metabolic Diseases, Department of Medicine, University of Verona Medical School and Hospital Trust of Verona, Verona, Italy; 4) Section of General Internal Medicine, VA Boston Healthcare System, Boston, MA, USA; 5) Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, USA; 6) Division of Endocrinology and Metabolism, Department of Medicine, Université de Sherbrooke, Sherbrooke, Québec, Canada; 7) Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; 8) National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA; 9) Diabetes Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 10) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 11) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 12) Cardiology Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

Aim Type 2 diabetes (T2D) and subclinical atherosclerosis (SCA) are associated in adults and may share genetic risk factors. The genetic risk for T2D could act as common ground for the development of both T2D and cardiovascular disease (CVD). *Hypothesis* We tested whether T2D genetic risk is associated with higher indices of SCA. *Methods* We conducted a cross-sectional analysis in up to 2,818 participants of the Framingham Offspring Study without a history of CVD. We tested the association of a genetic risk score (GRS) of 62 confirmed T2D single nucleotide polymorphisms (SNP) with measures of SCA: coronary artery (CACS) or abdominal aortic calcium score (AACS), internal (ICA-IMT) and common carotid intima-media thickness (CCA-IMT), and ankle-brachial index (ABI). ICA-IMT, CCA-IMT, AACS, CACS, insulin and Tg/HDL ratio were log-transformed to improve normality. We used linear mixed effect models to account for family relatedness, and tested a genetic-only model (GRS adjusted for sex) and a full risk factor model (GRS adjusted for sex, Tg/HDL ratio, waist circumference, BMI, fasting insulin, fasting glucose, parental T2D). We conducted subsidiary analyses of two distinct sub-GRS comprised of 20 SNPs associated with beta-cell function (B_GRS) or 10 associated with peripheral insulin resistance (IR_GRS) to test the hypothesis that genetic risk for IR in particular would be associated with SCA. We had 80% power to detect association of the 62 SNP GRS with at least 0.42% of the variance in any SCA trait; type 1 error rate set at p-value<0.01 (p=0.05 divided by 5, the number of traits analyzed). The B_GRS and IR_GRS each had 80% power to detect at least 0.48% of the variance in any trait at p-value<0.005 (0.05 divided by 5x2 models). *Results* We found no statistically significant association of the 62 SNP T2D GRS, the B_GRS or the IR_GRS with any SCA trait in any of the models, with the smallest p-value = 0.02 *Conclusions/interpretation* Despite the strong clinical association of T2D and CVD, and known association of the GRS with T2D risk, we did not find that an increasing burden of common variant risk for T2D to be associated with any of five SCA traits. Study strengths include a comprehensive T2D GRS, detailed phenotypic characterization and an adequately powered population sample size. Limitations include the use of tag SNPs from GWAS findings; as functional variants at the 62 loci become known, a genetic basis for shared T2D-CVD risk might become apparent.

2162F

Saturated fat intake modulates the association between an obesity genetic risk score and BMI. *P. Casas-Agustench[1,2], DK. Arnett[3], CE. Smith[1], C-Q. Lai[1], LD. Parnell[1], IB. Borecki[4], Y-C. Lee[1], JM. Ordovás[1,2,5].* 1) Nutrition and Genomics Laboratory, JM-USDA-HNRCA at Tufts University, Boston, MA, USA; 2) IMDEA Alimentación, CEI UAM+CSIC, Madrid, Spain; 3) Department of Epidemiology, School of Public Health, and Clinical Nutrition Research Center, University of Alabama at Birmingham, AL, USA; 4) Division of Statistical Genomics in the Center for Genome Sciences, Washington University School of Medicine, Saint Louis, MO, USA; 5) Department of Cardiovascular Epidemiology and Population Genetics, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain.

Background and objectives: Combining multiple genetic variants related to obesity into a global genetic risk score (GRS) might improve identification of individuals at risk of developing obesity. Moreover, characterizing gene-diet interactions is a research challenge aimed at investigating dietary recommendations for individuals with higher predisposition to obesity. The aim was to analyze the association between obesity GRS and body mass index (BMI) in a US population, with focus on gene-diet interactions with total and saturated fat intake. **Methods:** A cross-sectional study including 783 participants from the Genetics of Lipid Lowering Drugs and Diet Network study. A weighted GRS was calculated on the basis of 62 obesity-associated variants. Anthropometrical and biochemical measurements were taken using standard procedures. Dietary intakes were estimated with a validated questionnaire. **Results:** The obesity GRS was evaluated by tertiles for BMI, participants with a higher GRS had a higher BMI (P for trend <0.001). Participants in the highest GRS tertile and with BMI $<30kg/m^2$ showed lower total fat (2.3%) and SFA (1.0%) intake compared to participants in the same tertile and with BMI $\geq 30kg/m^2$ (P=0.018 and 0.022, respectively). An interaction was observed between total fat intake and the obesity GRS for BMI (P for interaction=0.014 for continuously evaluated dietary and GRS variables). We also identified interactions between SFA and MUFA intake and GRS evaluated continuously for BMI (P for interaction=0.010 and 0.027, respectively). Finally, we obtained similar results when the SFAs were evaluated categorically according to low and high values (P for interaction=0.014 for categorical SFA intake). Interactions were also observed between total fat, SFA and MUFA intake and the obesity GRS for waist circumference (P for interaction=0.004, 0.004 and 0.015, respectively, for continuously evaluated dietary and GRS variables). **Conclusions:** Total fat and especially SFA intake interact with an obesity GRS in determining BMI in a US population. Dietary recommendations to reduce BMI in population with high obesity GRS would be to reduce total fat intake mainly by limiting SFAs.

2163W

GxE GWAS and path analysis identify a cardiovascular and metabolic risk gene EBF1. *A. Singh[1, 2, 4], M.B. Babyak[1, 4], D.K. Nolan[2], B.H. Brummett[1, 4], R. Jiang[1, 4], I.C. Siegler[1, 4], S.H. Shah[2, 3], R.B. Williams[1, 4], E.R. Hauser[2, 3, 5].* 1) Behavioral Medicine Research Center, Duke University Medical Center, Durham, NC; 2) Center for Human Genetics, Duke University Medical Center, Durham, NC; 3) Department of Medicine, Duke University Medical Center, Durham, NC; 4) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC; 5) Durham Epidemiologic Research and Information Center, Durham Veterans Affairs Medical Center, Durham, NC.

Understanding the relationship between genetic susceptibility, psychosocial stress and metabolic factors acting in combination to increase the risk of CVD needs diverse methodologies including gene-by-stress interaction analysis and disease pathways analysis. In this study we performed gene-environment interaction genome-wide association analysis (GxE GWAS) to identify single nucleotide polymorphisms (SNPs) whose effects on metabolic traits are modified by chronic psychosocial stress in the Multi-Ethnic Study of Atherosclerosis (MESA). The GxE GWAS for central obesity trait hip circumference identified five SNPs within the Early B-cell Factor 1 (EBF1) gene in MESA Whites. All of the five SNPs were in strong linkage disequilibrium (LD). For these SNPs the gene-by-stress interaction term p-values were genome-wide significant (Ps = 7.14E-09 - 2.33E-08), while genetic main effect p-values were not significant at the conventional genome-wide significance (Ps = 0.015 - 0.031). Further analysis identified gene-by-stress interaction effects for waist circumference, body mass index (BMI), fasting glucose, type II diabetes status, and common carotid intimal-medial thickness (CCIMT), suggesting that gene-by-stress interaction may underlie the relationship between increased central obesity and increased blood glucose, diabetes, and finally increased CVD risk and atherosclerosis. Structural equation path analysis suggested that the path from chronic psychosocial stress to CCIMT via hip circumference and fasting glucose was larger (estimate = .26, p = .033, 95%; CI = .02, .49) in the EBF1 rs4704963 CT/CC genotypes group than the same path in the TT group (estimate = .004, p = .34, 95%; CI = -.004, .012). Published analyses of the EBF1 gene interaction network connects EBF1 to PPARγ expression and SLC2A4 (GLUT4) suggesting that EBF1 plays a role in metabolic disease pathways involved in CVD pathogenesis.

2164T

Influences of SNPs of *LIPC*, *LIPG*, *APO E*, *PLTP* and *PON1* genes on plasma high-density lipoprotein, cholesterol and paraoxonase-1 activity. *D.Z. Scherrer[1], V.H.S. Zago[1], R. Secolin[2], E.S. Parra[1], N.B. Panzoldo[1], F. Alexandre[1], I.V. Calanca[1], E. Nakandakare[3], E.C.R. Quintão[3], E.C. de Faria[1].* 1) Clinical Pathology, Universty of Campinas (UNICAMP), Campinas, São Paulo, Brazil; 2) Medical Genetics, Universty of Campinas (UNICAMP), Campinas, São Paulo, Brazil; 3) Lipid Laboratory, University of São Paulo Medical (USP) School, São Paulo, Brazil.

Introduction: The determinants of HDL-C levels are under strong control of genes encoding HDL components like lipases, lipid transfer proteins and HDL receptors. The aim of this study was to investigate the influences of several single nucleotide polymorphisms (SNPs) and components of HDL metabolism on plasma levels of HDL-C, cholesterol and paraoxonase-1 (PON1) activity. Material and Methods: Healthy normolipidemic volunteers (Females=159, Males=153; 19 to 75 years of age) were studied. The study was approved by the local Ethics Committee on Research, following the declaration of Helsinki. Genomic DNA was extracted from peripheral blood cells and the SNPs of *LIPC* (rs1800558, rs2070895), *LIPG* (rs3813082) *APO E* (rs429358, rs7412), *PLTP* (rs6065904) and *PON1* (rs662) detected in the OpenArray® Real Time PCR Platform (Applied Biosystems). Serum Cholesterol and HDL-C were determined by enzymatic methods in an automated system (Hitachi, Roche). The size (nm) and volume (nm3) of HDL were measured in the Nanotrac Particle Size Analyzer (Microtrac, USA). Plasma activities of CETP, PLTP, hepatic lipase and lipoprotein lipase were determined by radioisotopic assays, while PON1 activity by a chromogenic method and endothelial lipase, PLTP mass and plasma insulin were determined by ELISA assays. Lecithin cholesterol acyltransferase (LCAT)-mediated cholesterol esterification rate was determined using endogenous and exogenous substrates. Multivariate stepwise regression analyses were performed in the R software to determine the main influences on plasma cholesterol, HDL-C and PON1 activity. Results: Plasma HDL-C was predicted by plasma cholesterol (R2=0.40; p≤0.0001), PLTP mass (R2=0.12; p≤0.0001), endogenous LCAT activity (R2=0.065; p≤0.0001), *LIPC* rs1800588 (TT genotype: R2=0.0025; p≤ 0.01) and *LIPC* rs2070895 (GG genotype: R=0.0017; p=0.02). Plasma cholesterol was predicted only by plasma HDL-C (R=0.40; p≤0.0001) and *APO E* rs7412 (TT genotype: R= 0.0186; p ≤0.03). On the other hand, PON1 activity was predicted by *PON1* rs662 in the TT (R=0.38; p≤0.0001) and CT (R=0.19; p≤0.0001) genotypes. Conclusions: Influences of *LIPC* rs1800588 and rs2070895 on HDL-C levels were observed in healthy normolipidemic individuals. In addition, PON1 activity was strongly predicted by *PON1* rs662 TT and CT genotypes which induced increases in its antioxidant activity. Support: FAPESP (2006/60585-9), CNPq (159980/2012-7).

2165F

Analysis of allele specific expression and transcription factor activity identifies potential gene-by-environment mechanisms of coronary heart disease risk. *O.V. Sazonova[1,2], A. Raiesdana[1], S.T. Nurnberg[1], K.S. Smith[2], V. Anaya[2], T. Quertermous[1], S.B. Montgomery[2,3].* 1) Department of Cardiovascular Medicine, Stanford University, Stanford, CA; 2) Department of Pathology, Stanford University, Stanford, CA; 3) Department of Genetics, Stanford University, Stanford, CA.

Coronary heart disease (CHD) is a complex disorder caused by a combination of environmental and genetic risk factors, the interactions of which are poorly understood. Gene-by-environment effects on the regulation of coronary artery smooth muscle cell (CASMC) phenotype are of particular interest as de-differentiation of these cells is strongly implicated in plaque development. We hypothesized that stimulating de-differentiation with fetal bovine serum would lead to changes in gene expression (GE) and allele-specific expression (ASE) in human CASMCs. To discover serum-induced genetic regulation, we applied genomic methods to quantify and statistically validate differential ASE (dASE) in CASMCs cultured in the presence and absence of serum. For gene-by-environment changes discovered through dASE, we further mined transcription factor binding data to identify potentially causal variants. In detail, CASMCs from 7 individuals were cultured using 0% or 10% serum, genotyped using the Illumina HumanExome BeadChip, and subjected to RNA-Seq. Our method identified 72 dASE genes (FDR 17%) and 10 dASE SNPs (FDR 20%) that exhibit dASE in 2 or more individuals. Only one of 72 dASE genes also showed differential gene expression (dGE) (957 dGE genes in total discovered with DESeq, FDR 5%), suggesting dASE can recover distinct mechanisms for environment-induced regulation not captured by dGE alone. Current studies are underway to identify dASE and dGE pathways related to CHD and their potential regulators. To identify potential causal mechanisms of observed gene-by-environment effects, we examined via ChIP-Seq the binding profile of TCF21 near dASE and non-dASE genes. TCF21 was selected because it regulates coronary artery development, is within a significant CHD GWAS locus, and is upregulated in CASMCs with serum exposure. We hypothesized that altered TCF21 activity may be one mechanism underlying dASE, and found statistically significant enrichment of TCF21 motif-disrupting SNPs among dASE genes compared to non-dASE genes with similar TCF21 associations ($p<0.02$). This finding indicates that dASE genes are susceptible to allele-specific TCF21 binding. Collectively, our study demonstrates that serum exposure induces significant changes in patterns of ASE via mechanisms that may involve variants within the TCF21 binding motif. This analysis provides a powerful new avenue for identifying causal gene-by-environment variants involved in CHD and other complex diseases.

2166W

A genome-wide association study reveals novel susceptibility loci for coronary artery disease in Saudi Arabs. *N. Dzimiri[1], C. Nguyen[3], S.M. Wakil[1], N.P. Muiya[1], E. Andres[1], N. Mahzar[1], B. Baz[1], S. Hagos[1], M. Alshahid[2], G. Morahan[3], B.F. Meyer[1].* 1) Dept Genetics, King Faisal Spec Hosp & Research Centre, Riyadh, Saudi Arabia; 2) King Faisal Heart Institute, King Faisal Spec Hosp & Research Centre, Riyadh, Saudi Arabia; 3) Western Australian Institute for Medical Research, University of Western Australia, Australia.

Although multiple loci have been identified for coronary artery disease (CAD) to date, using genome-wide association (GWAS) approach, no GWAS of incident CAD has been reported for Saudi Arabs. In this study, we performed a GWAS for CAD incidence in 5,418 Saudis of Arab ancestry. We discovered 57 genome-wide significant SNPs susceptible for CAD disease that reached the threshold of genome-wide significance ($P < 1 \times 10\text{-}5$). These loci mapped in or near a number of genes including CLCKA and SMYD3 on chromosome (chr) 1, LOC388946, vomeronasal 1 receptor 18 pseudogene, MTND4P28, PLCL1, and CPS1 on chr 2, RPL36AP17 and LOC131054 on chr 3, GRIK2, MAP3K5 and TRNAV37P on chr 6, ZNF804B and MKLN1 on chr 7, GINS4 on chr 8, CDKN2B-AS1 and FAM120A on chr 9, MYEOV and NARS2 on chr 11, CRADD on chr 12, MIS18BP1, PAPOLA, FMN1, LIG3, RFFL on chr 17, RAD51D, PSMB3, BC068609 on chr 19, ARHGAP40 gene on chr20s. This is the first finding of incident CAD loci identified by GWAS in Saudi Arabs. We also replicated 19 previously discovered genetic associations (including CDKN2BAS, CNNM2, PHACTR1, PPAP2B, SLC12A9), providing support for our study design. These findings provide new insights into pathways contributing to the susceptibility for CAD in the Saudi Arab population.

2167T

A genome-wide association study reveals novel susceptibility loci for obesity in Saudi Arabs. *N.P. Muiya[1], C. Nguyen[3], S.M. Wakil[1], E. Andres[1], N. Mazhar[1], B. Baz[1], D. Gueco[1], M. Najai[1], O. Alboudary[1], S. Hawari[1], S. Hagos[1], M. Alshahid[2], G. Morahan[3], B.F. Meyer[1], N. Dzimiri[1].* 1) King Faisal Spec. Hospital & Research Centre, Riyadh, Central, Saudi Arabia; 2) King Faisal Heart Institute, King Faisal Spec. Hospital & Research Centre, Riyadh, Central, Saudi Arabia; 3) Western Australian Institute for Medical Research, University of Western Australia, Australia.

To date, several genetic loci and gene networks have been detected for obesity using genome-wide association studies (GWAS) methodology. However, no GWAS study on obesity has been reported for the Saudi population yet. In this study we performed a GWAS for the prevalence of obesity in 5,418 Saudis of Arab ancestry. We discovered 94 genome-wide significant SNPs that reached the threshold of genome-wide significance ($p<1\times10\text{-}5$) and may contribute to susceptibility to acquiring obesity. These loci mapped onto or near a number of genes, including the PRE-b-Cell leukemia homeobox 1 (PBX1) and proprotein convertase subtilisin/kexin type 9 (PCSK9) on chromosome (chr) 1, NCK-associated protein 5 on chr 2, catenin (cadherin-associated protein), alpha 1 (CTNNA1) and zinc finger RNA binding protein (ZFR) on chr 5, Cub and Sushi domain-containing protein 1 (CSMD1) and zinc fingers and homeoboxes 2 (ZHX2) on chr 8. Other genes mapping on these loci were the V-Rel reticuloendotheliosis viral oncogene homolog A (avian) (RELA), the olfactory receptor, family 52, subfamily E, members 1 and 2 (OR52E1/2), signal-induced proliferation-associated 1 (SIPA1) on chr 11, bicaudal D homolog 1 (drosophila) (BICD1) on chr 12, AKAP6 on chr 14, ribosomal protein L15 pseudogene 21 (RPL15P21) on chr 17, chromosome 20 open reading frame 26 (C20orf26) and cadherin 4, type 1, R-cadherin (retinal) (CDH4) on chr 20. This is the first finding of incident obesity loci identified by GWAS in Saudi Arabs. One other previously reported genetic association, the fat mass and obesity associated (FTO) gene, was replicated, providing support for our study design. These results show a significant correlation of several genomic loci with obesity and provide new insights into pathways contributing to susceptibility for obesity.

2168F

Investigation of genome-wide DNA methylation marks associated with FV Leiden mutation in patients with venous thrombosis. *D. Aïssi[1], M. Ladouceur[2], J. Dennis[2], F. Gagnon[2], P.E. Morange[3], D.A. Trégouët[1].* 1) INSERM UMR_S 937, Paris, France; 2) Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 3) INSERM UMR_S 1062, AP-HM, Marseille, France.

Several lines of evidence are emerging to suggest that DNA methylation marks are involved in the susceptibility of various complex human diseases including thrombosis-related disorders. As part of a general project aimed at identifying DNA methylation-sensitive regulatory mechanisms associated with biomarkers of the fibrinolysis/coagulation cascade, we measured genome-wide DNA methylation levels in peripheral blood samples of 349 patients with venous thrombosis (VT) using the dedicated Illumina Human-Methylation450 bead array. In the current study, we undertook a genome-wide analysis of 485,577 CpG sites to assess whether patients carrying the FV Leiden mutation might be associated with specific DNA methylation profiles. We identified a locus on chromosome 1 characterized by two CpG sites whose DNA methylation levels differed significantly (p = 1.32 10-11 and 3.16 10-11) between carriers (N = 98) and non-carriers (N = 251). Of note, the smallest p-value observed at the F5 gene, also on chromosome 1, was p = 0.144. The two associated CpG sites replicated (p = 8.52 10-11 and 3.57 10-7, respectively) in an independent sample of 214 individuals from 5 large French-Canadian families selected from an index case with venous thrombosis among which 53 were carriers of the FV Leiden mutation and 161 were non carriers. We further tested whether these two CpG sites were associated with biomarkers of VT risk known to be under the strong influence of FV Leiden mutation and that explores the Protein C anticoagulant pathway: the agkistrodon contortrix venom (ACV) test in the discovery population and the activated protein C resistance (APCR) in the replication sample. Strong significant associations were observed (all $p < 10\text{-}6$) which subsequently vanished ($p > 0.10$) after adjusting for FV Leiden mutation. These results may add novel insights into the epigenetic regulation of the coagulation cascade and provide evidence for the usefulness of peripheral blood methylation levels in studying biomarkers of coagulation-relevant epigenetic mechanisms.

2169W

Sex-specific effects of CAD SNPs in sudden cardiac death. *F.N. Ashar[1], C. Albert[2], S.S. Chugh[3], A. Cupples[4], M. Eigelsheim[5], P. Goyette[6], A. Huertas-Vazquez[3], H. Huikuri[7], J. Jintilla[7], X. Jouven[8], S. Kääb[9], M. Kortelainen[10], P. Kwok[11], T. Lehtimäki[12], L. Lyytikäinen[12], M. Müller-Nurasyid[13], C. Newton-Cheh[14], B. Psaty[15], S. Pulit[14], D. Siscovick[15], B. Stricker[5], N. Sotoodehnia[15], D.E. Arking[1], CHARGE-SCD.* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 2) Division of Preventive Medicine, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) The Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; 4) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 5) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 6) Université de Montréal, Montreal, Quebec, Canada; 7) Dept. of Internal Medicine, University of Oulu and University Central Hospital, Oulu, Finland; 8) Université Paris Descartes, Assistance Publique-Hopitaux de Paris, Hôpital Européen Georges Pompidou, Paris, France; 9) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany; 10) Department of Forensic Medicine, University of Oulu, Oulu, Finland; 11) Department of Dermatology, Cardiovascular Research Institute, and Institute for Human Genetics, University of California, San Francisco, California, USA; 12) Department of Clinical Chemistry, Fimlab Laboratories, Shool of Medicine, University of Tampere, Finland; 13) Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany; 14) Center for Human Genetic Research, Cardiovacular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA; 15) Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA.

Sudden cardiac death (SCD), a major cause of mortality among Western populations, is a clinically heterogeneous condition broadly defined as a sudden unexpected pulseless condition due to a ventricular arrhythmia. While the immediate underlying cause of SCD is considered to be electrical instability, the majority of SCD occurs in the setting of coronary artery disease (CAD). Given this link between CAD and SCD, we chose to investigate the effects of 45 single nucleotide polymorphisms (SNPs) previously associated at genome-wide significance with myocardial infarction/CAD for their potential association with SCD in a meta-analysis of 4496 cases and over 25,000 controls from 10 studies of European descent. We report a significant enrichment of genetic effects at these SNPs for SCD in the same direction as reported for CAD (32/45, P=0.005). Since both sex and age influence CAD burden and risk of SCD, we subsequently conducted analyses stratified by sex, as well as limited to samples less than or equal to 65 years of age. As seen in our overall analysis, we continue to observe the enrichment of effects in males (33/45, P=0.002) and younger samples (age<=65) (32/45, P=0.005). However, there is complete loss of the enrichment signal in the female-only subgroup (23/45, P=0.88). To investigate whether the enrichment was driven primarily by the presence of underlying CAD, we stratified the analysis on the basis of whether controls were population-based (7 studies) or had been selected to have CAD (2 studies). We confirmed the loss of the enrichment signal in the analysis using the CAD controls (20/45, P=0.54). These findings provide compelling genetic evidence for previous pathological studies that have reported a difference in etiology of disease between males and females, with a greater atherosclerotic burden seen in males than in females, and highlight the importance of assessing genetic associations with SCD stratified by sex.

2170T

An extreme phenotype approach to identify genes in Caribbean Hispanics for carotid intima-medial thickness, a preclinical marker of atherosclerosis. *S.H. Blanton[1,2], A.H. Beecham[1], L. Wang[1], C. Dong[2], D. Cabral[2], R.L. Sacco[1,2], T. Rundek[2].* 1) Dept of Human Genetics, Univ Miami, Miami, FL; 2) Dept of Neurology, Univ Miami, Miami, FL.

Carotid intima-medial thickness (cIMT) is a recognized risk factor for cardiovascular disease. The genetic underpinnings of cIMT are not known. The aim of this study was to identify risk/protective loci for cIMT using an extreme phenotype approach in the analysis of SNPs from our genome-wide association study. In individuals from the population-based Northern Manhattan Study, cIMT was measured by high-resolution B-mode ultrasound and expressed as total cIMT, a composite measure of IMT at different carotid sites. Genotyping was done with the Affymetrix 6.0 SNP array. Among the 908 Hispanics, we first calculated the cIMT residual score for each participant by regressing cIMT on known risk factors (age, sex, body mass index, fasting glucose, cigarette smoking pack-years, dyslipidemia, and hypertension). To enrich the genetic effects and increase our power to detect associations, we next identified the individuals in the extreme 10% and 20% of the residual distribution, for a total of 90 and 200 individuals on each side respectively, thus ensuring that the phenotypic extremes were largely unexplained by these known risk factors. For each of the extreme thresholds, we performed logistic regression analysis on 877K SNPs, while controlling for the top 3 PCAs. A gene based analysis was then performed based on the SNP results using VEGAS. The top gene from the 10% threshold (p=3.15E-04) was regenerating islet-derived 1 alpha (*REG1A*), which is over expressed in the early stages of Alzheimer disease and may play a role in both celiac disease and rheumatoid arthritis. Its relationship to cIMT is unclear. The top gene from the 20% threshold (p=1.5E-05) was family with sequence similarity in 120B (*FAM120B*). While this gene is not well characterized, it is a co-activator of the proliferator-activated receptor gamma family, which plays a crucial role in cholesterol and lipid metabolism. In addition to the gene based analysis, we also conducted a pathway based analysis using genes with a p<0.05 as input into WebGestalt. After adjusting for multiple testing, several pathways of interest were identified. In particular for the 20% threshold, the pathway of fat digestion and absorption was the most significant (p=1.0E-04). Novel SNPs reported here support the critical link between inflammation and cholesterol metabolism with arterial wall biology leading to subclinical atherosclerosis or to protection against atherosclerosis.

2171F

The link between hepcidin, iron and atherosclerosis: a Mendelian randomization approach. *T.E. Galesloot[1], L.L. Janss[2], D.W. Swinkels[3], S.H. Vermeulen[1].* 1) Department for Health Evidence, Radboud University Medical Centre, Nijmegen, The Netherlands; 2) Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark; 3) Laboratory of Genetic, Endocrine and Metabolic Diseases, Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands.

Background The 'iron hypothesis' states that people with elevated serum iron levels face a greater risk of cardiovascular disease, but epidemiologic studies on associations between iron depletion and cardiovascular risk remain inconclusive. Hepcidin, central regulatory molecule of systemic iron homeostasis, might play a role in the progression of atherosclerosis. In this study, we will investigate the link between hepcidin, iron and atherosclerosis using a Mendelian randomization approach. Methods We will include participants of the Nijmegen Biomedical Study (NBS) aged 46-67 years for whom measurements of hepcidin, iron parameters and non-invasive measurements of atherosclerosis as well as GWAS data are available (N±800). We will derive heritability estimates for hepcidin, ferritin and iron and estimate the genetic correlation between these traits using GWAS data by application of Bayesian modeling techniques. Subsequently, we will apply a multivariate GWAS analysis for hepcidin, ferritin and iron using a multivariate test of association (MQFAM) implemented in PLINK. We will identify both shared (pleiotropic) and non-shared (independent) genetic regulators of hepcidin, ferritin and iron. The independent genetic variants and a combination thereof will be used as instrumental variables in a Mendelian randomization approach to study their association with atherosclerosis. Expected results Our study will apply up-to-date techniques to identify independent genetic determinants of correlated phenotypes. The identified differences in genetic etiology will be exploited to create insight in the causal roles of hepcidin, ferritin and iron in the development of atherosclerosis.

2172W

Eleven novel loci influencing blood pressure. *B. Keating[1], V. Tragante[2,3], M. Barnes[4], S. Ganesh[5], I.H. Gho[2], X. Zhu[6], D. Levy[7], H. Hakonarson[1], P.I.W. de Bakker[3,8,9,10], F.W. Asselbergs[2,8,11], P.B. Munroe[12], The IBC BP consortium.* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; 4) Bioinformatics group at the William Harvey Research Institute, Queen Mary University of London, London; 5) Division of Cardiovascular Medicine, University of Michigan Health System, Ann Arbor, MI, USA; 6) Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 7) Center for Population Studies, National Heart, Lung, and Blood Institute, Framingham, MA, USA; 8) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 9) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 10) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts; 11) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands; 12) Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK;.

Blood pressure (BP) is a heritable determinant of risk for cardiovascular disease. To investigate genetic associations with systolic BP (SBP), diastolic BP (DBP), mean arterial pressure (MAP), and pulse pressure (PP), we genotyped ~50,000 single nucleotide polymorphisms (SNPs) that capture variation in ~2,100 candidate genes for cardiovascular phenotypes in up to 87,736 individuals of European ancestry from 36 studies. We identified 11 novel associations in independent loci containing 31 genes including PDE1A, HLA-DQB1, CDK6, PRKAG2, VCL, H19, NUCB2, RELA, HOXC@ complex, FBN1 and NFAT5, and 16 previously known associations with SBP, DBP, MAP, or PP, confirmed at the Bonferroni-corrected array-wide significance threshold ($P < 6 \times 10\text{-}7$) in a combined meta-analysis with an additional independent set of up to 68,368 individuals of European ancestry. An in depth bioinformatic analysis of the genes and variants in the 11 loci provided functional evidence to support several genes, including cis eQTL associations between SNPs and the expression of HLA-DQB1 and NUCB2. Analysis of ENCODE data identified histone modifications and motifs related to these and other novel BP signals. Druggability analysis for associated genes in this study using existing public resources, including databases of small molecules, shows that ten genes are predicted to be modified by small molecule therapeutics. In summary, we identified novel loci associated with BP and confirmed multiple previously reported associations. Our findings extend our understanding of genes involved in BP regulation, some of which may provide new targets for therapeutic intervention or drug response stratification.

2173T

Genetic and Phenotypic Architecture of Coronary Artery Disease in a Diverse Population. *I. Konidari[1], L. Wang[1], Z. Liu[1], N. Sikka[1], P. Goldschmidt[2], M.A. Pericak-Vance[1], D. Seo[1,2], G.W. Beecham[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) University of Miami, Department of Medicine, Division of Cardiology, Miami, FL, USA.

Coronary heart disease is the leading cause of death in the United States, with atherosclerosis of the coronary arteries (CAD) being one of the major underlying etiologies. Though the role of genetics on CAD has been studied in European ancestry populations, its role in Hispanic populations is unclear. To further determine the genetic etiology of CAD, we are performing a GWAS of CAD using a diverse set of cardiac catheterization patients. This study utilizes the Miami Cardiovascular Registry of the University of Miami, Miller School of Medicine. Patients were ascertained through cardiac catheterization labs; consisting of over 65% Hispanic and 15% African American. All 2,000 patients were phenotyped with a research grade read of coronary angiograms, describing the CAD in each of the 16 major coronary artery branches. Genotype data were generated using the Affymetrix 6.0 array and Illumina HumanExome beadchip platforms. Extensive QC tests were performed to ensure the integrity of the data, including sample filters and SNP filters; EIGENSTRAT methods were used to detect and subset based on race/ethnicity. Association was tested using regression-based methods, with age and sex included as covariates. Genotypes were initially analyzed within the three racial groups (White non-Hispanic, Black, Hispanic/Latino). Results indicate significantly associated variants in/near the GBA3, PMEPA1, and EPB41L2 genes. Additional analyses include tests for racial differences in the architecture of disease; that is: do different populations have different coronary branches affected at different levels? Additionally, are testing for association with disease in specific major branches; results will be presented. This study will yield important results, both about the phenotypic differences and similarities between populations, and about underlying genetic etiology of CAD.

2174F

Local ancestry inference in a genome-wide association study of a genetically diverse population with coronary artery disease. *Z. Liu[1], L. Wang[1,2], N. Vasudeva[2], P.J. Goldschmidt-Clermont[3], M.A. Pericak-Vance[1,2], D.M. Seo[3], G.W. Beecham[1,2].* 1) Dr. John T. Macdonald Foundation, Department of Human Genetics, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 3) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL.

Coronary artery disease (CAD) is a leading cause of mortality in the United States with an estimated 500,000 deaths each year. Recent studies have shown the prevalence of CAD is different among different racial/ethnic populations with Native Americans being the highest (11.6%), followed by African Americans (6.5%), Hispanics (6.1%), whites (5.8%) and Asian Pacific (3.9%). Genetic risks underlying CAD are not well understood, especially those that are racially/ethnically specific. To study the genetic risk of CAD, we conducted a genome-wide association study (GWAS) of 2,000 catheterization patients from the Miami Cardiovascular Registry of the University of Miami. This dataset represents the diversity of South Florida populations, including European, African, and Hispanic/Latinos ancestry. Genotyping was performed on Affymetrix 6.0 platform, and standard QC were performed to ensure data integrity. To address population substructure, we performed local ancestry inference to identify the ancestral state of each locus. Haplotypes were estimated from publicly available reference datasets using ShapeIT. Given phasing uncertainty, we generated 100 phasing iterations for each individual and use them to infer local ancestry with the LAMP-LD/LAMP-ANC software. The 100 iterations were averaged to generate the percentage of ancestry of European, African or Native American at each locus. To validate this approach, we estimated global ancestry for each individual by averaging the results across all chromosomes and comparing it with the estimates derived from a principal component analyses (PCA). Global ancestry estimated by our approach and PCA are highly correlated (correlation coefficient between eigenvector 1 and African ancestry component $r2=0.99$; for eigenvector 2 and Native American/Asian, $r2= -0.87$), suggesting a reliable ancestry estimation by our approach. By summing across individual haplotypes, our samples composed of approximately 75% European descent, 20% African descent and 5% Native American/Asian descent. We will utilize this local ancestry information to study the contribution of ancestry specific minor alleles to the patient's anatomic atherosclerotic burden and blood pressure. The advantage of this study is that it allows us to determine common variants associated with cardiovascular disease traits that are common in all populations and unique to a specific race/ethnicity, therefore improving our understanding of CAD genetic risks.

2175W

Genetic variants affecting the expression of *DRAM2* at 1p13.3 are associated with acute myocardial infarction with different effects for STEMI and NSTEMI.
P. Salo[1], J. Sinisalo[2], J. Kettunen[3,1], A. Havulinna[4], A. Sarin[3,1], T. Hiekkalinna[1,3], S. Ripatti[3,1], P.J. Karhunen[5,6], H. Huikuri[7], M. Lokki[8], V. Salomaa[4], M. Nieminen[2], M. Perola[1,3,9]. 1) Public Health Genomics Unit, Natl Inst Health & Welfare, Helsinki, Finland; 2) Division of Cardiology, Heart and Lung Center HUCH, Helsinki University Central Hospital, Helsinki, Finland; 3) Institute for Molecular Medicine Finland (FIMM,) Helsinki, Finland; 4) Chronic Disease Epidemiology and Prevention Unit, Natl Inst Health & Welfare, Helsinki, Finland; 5) School of Medicine, University of Tampere, Tampere, Finland; 6) Fimlab Laboratories Ltd, Tampere University Hospital Region, Tampere, Finland; 7) Department of Internal Medicine, Oulu University Hospital and University of Oulu; 8) Transplantation Laboratory, Haartman Institute, University of Helsinki, Helsinki, Finland; 9) Estonian Genome Center of the University of Tartu, Tartu, Estonia.

Myocardial infarction (MI) is usually caused by coronary artery disease (CAD). The narrow-sense heritability of MI has been estimated to be 0.35 and 0.38 for males and females. Known genetic risk variants explain 11% of the additive genetic variance in susceptibility to CAD.

Despite advances in understanding the genetic basis of MI, little attention has been paid to the distinction between ST-segment elevation MI (STEMI) and non-ST-segment elevation MI (NSTEMI). The division is made based on characteristic ECG changes and the two are known to present with different characteristics. In most cases STEMI results from a complete occlusion of a coronary artery whereas NSTEMI is caused by partial or transient blockage. NSTEMI tends to present with greater minimal luminal area in the culprit artery and smaller infarct size. Curiously, recurrent infarctions are often of the same type, suggesting that some individuals may be particularly susceptible to either STEMI or NSTEMI. The possible differences in their genetic risk factors, however, remain virtually unknown.

We performed a genome-wide association study of MI (1579 cases, 1576 controls) with stratification into NSTEMI and STEMI and replicated the results in two independent study samples. We identified a novel risk locus for acute MI at 1p13.11 associated with NSTEMI (OR=1.56, P=4×10^{-10}) but not with STEMI (OR=1.14, P=0.08). We further show that the SNPs conferring risk to NSTEMI are also associated with the expression of *DRAM2* contained within the NSTEMI-associated region and provide evidence for two distinct but partially overlapping association signals at the locus.

The exact function of *DRAM2* remains unknown, but it has been shown to be required for efficient autophagy induction. Silencing *DRAM2* inhibits autophagy under starvation and attenuates p53-mediated cell death. The association of variants with both NSTEMI and *DRAM2* expression thus suggests regulation of autophagy may affect risk for NSTEMI, raising the possibility that *DRAM2* has an effect on the survival of myocardial cells during the ischemic conditions of acute MI. Furthermore, the results also show that genetic factors in part determine whether coronary artery disease results in NSTEMI rather than STEMI.

2176T

A novel MMP12 locus is associated with large artery atherosclerotic stroke using a genome-wide age-at-onset informed approach.
M. Traylor[1], K-M. Mäkelä[2,3], L.L. Kilarski[1], E.G. Holliday[4,5], W.J. Devan[6,7], M.A. Nalls[8], K.L. Wiggins[9], Y-C. Cheng[10,11], S. Achterberg[12], R. Malik[13], C. Sudlow[14], E. Raitoharju[2,3], V. Thijs[15], A. Lindgren[16,17], A. Slowik[18], M. Walters[19], P. Sharma[20], J.R. Attia[4], G.B. Boncoraglio[21], P.M. Rothwell[22], P.I.W. de Bakker[7,23,24,25], J.C. Bis[9], S.J. Kittner[11], J. Rosand[6,7], J.F. Meschia[26], C. Levi[27], M. Dichgans[12,28], T. Lehtimäki[2,3], C.M. Lewis[29,30], H.S. Markus[1], METASTROKE, WTCCC2. 1) Research Centre for Stroke & Dementia, St. George's University of London, London, UK; 2) Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital, Tampere, Finland; 3) School of Medicine, University of Tampere, Tampere, Finland; 4) Center for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle, New South Wales, Australia; 5) Center for Bioinformatics, Biomarker Discovery and Information-Based Medicine, Hunter Medical Research Institute, Newcastle, New South Wales, Australia; 6) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 7) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 8) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA; 9) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 10) University of Maryland School of Medicine, Dept of Medicine, Maryland, USA; 11) Department of Neurology, Veterans Affairs Maryland Health Care System, Baltimore, Maryland, USA; 12) Department of Neurology and Neurosurgery, Utrecht Stroke Centre, Rudolf Magnus Institute of Neuroscience, University Medical Centre, Utrecht, The Netherlands; 13) Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians-Universität, Munich, Germany; 14) Division of Clinical Neurosciences and Insititute of Genetics and Molecular Medicine, University of Edinburgh, UK; 15) Department of Neurology, University Hospitals Leuven, Leuven, Belgium; 16) Department of Clinical Sciences Lund, Neurology, Lund University, Sweden; 17) Department of Neurology, Skåne University Hospital, Lund, Sweden; 18) Department of Neurology, Jagiellonian University, Krakow, Poland; 19) Institute of Cardiovascular and Medical Sciences, University of Glasgow, UK; 20) Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College London, UK; 21) Department of Cereberovascular Disease, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Istituto Neurologico Carlo Besta, Milan, Italy; 22) Stroke Prevention Research Unit, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK; 23) Julius Centre for Health Sciences and Primary Care, University Medical Centre, Utrecht, The Netherlands; 24) Department of Medical Genetics, University Medical Centre, Utrecht, The Netherlands; 25) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 26) Department of Neurology, Mayo Clinic, Jacksonville, FL, USA; 27) Hunter Medical Research Institute, New South Wales, Australia; 28) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 29) Department of Medical & Molecular Genetics, King's College London, London, UK; 30) Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK.

Genome-wide association studies (GWAS) have begun to identify the common genetic component to ischaemic stroke. However, ischaemic stroke (IS) is a syndrome with considerable heterogeneity. In analyses where clinical covariates explain a large fraction of disease risk, covariate informed designs have been shown to increase power to detect associations. As prevalence rates in IS are dramatically affected by age, and younger onset cases are likely to have higher stroke-specific genetic predisposition, we investigated whether an age-at-onset informed approach could detect novel associations. We used an age-at-onset informed design to identify novel variants associated with IS and its subtypes; cardioembolic (CE), large vessel disease (LVD) and small vessel disease (SVD) stroke in 6,784 cases of European ancestry and 12,095 ancestry-matched controls. Regression analysis to identify SNP associations was performed per centre on posterior liabilities after conditioning on age-at-onset and affection status. We sought further evidence of a novel association with LVD in an independent replication cohort of 1,526 cases, and followed up the association by examining mRNA expression levels of the nearby genes in atherosclerotic carotid artery plaques from the Tampere Vascular Study. We identified a novel association with an MMP12 locus in LVD (rs660599; p=1.6×10-7) using an age-at-onset informed approach. We found independent replication of this locus in a second population (p=0.0048), reaching genome-wide significance overall (p=1.7×10-8). The nearby gene, MMP12, was found to be significantly overexpressed in carotid plaques compared to atherosclerosis-free control arteries (p=1.2×10-15; fold change=335.6). Furthermore, all previously identified associations were found to be more significant using the age-at-onset informed approach. Our results indicate that a covariate-informed design, conditioning on age-at-onset of stroke, is likely to unearth novel variants not identified by conventional GWAS. We show this explicitly by identifying an association with a novel MMP12 locus in large artery stroke, supported by evidence of increased mRNA expression of the implicated gene in carotid plaques.

2177F

Genetic Architecture of Hypertension in a Multi-Ethnic Population.
N. Vasudeva[1], L. Wang[1], A. Beecham[1], P. Goldschmidt[2], M. Pericak-Vance[1], D. Seo[1,2], G. Beecham[1]. 1) Hussman Institute for Human Genomics, University of Miami; 2) Department of Medicine, University of Miami.

Previous genetic studies on hypertension have focused mostly on the European ancestry populations. We sought to dissect the genetic basis for hypertension in a diverse population of Hispanics, European ancestry, and African ancestry using a genome-wide association study. The 2,000 samples are ascertained through the clinic-based Miami Cardiovascular Registry, a dataset consisting of ~55% Hispanic and were genotyped using the Affymetrix SNP array 6.0. Quality control tests were performed to ensure data integrity. Linear regression was used to evaluate single variant association with Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Pulse Pressure (PP=SBP-DBP), and Mean Arterial Pressure (MAP=(SBP+2DBP)/3). To account for the medication-lowered blood pressure, we added 7% and 6%; for SBP and DBP, respectively, for each antihypertensive medication taken. Age, age², gender, dyslipidemia and diabetes are significantly associated ($P<0.05$) with blood pressure in our samples and were included as covariates. Eigenstrat analysis was used to divide samples into Hispanic, White non-Hispanic, and Black subgroups. Analysis was performed within each subgroup and meta-analyzed across subgroups. Strong association ($P < 0.00001$) was found at multiple SNPs for each trait, with the strongest associations at rs9584521 near the MBNL2 gene for PP ($P=7\times10\text{-}8$) and rs7278181 in CLDN14 for SBP ($P=9\times10\text{-}8$). Interestingly, both of these loci were monomorphic in the European ancestry subgroup, and would not have shown association in a combined analysis. Additional analyses are underway, including analysis of multiple variants in aggregate and replication in an additional dataset. Our study provides essential data in a diverse population and may lead to additional insights into the genetics of hypertension.

2178W

12 Novel genome wide associations for human cardiac repolarization.
N. Verweij[1], W.G. Wieringa[1], I. Mateo Leach[1], D.J. van Veldhuisen[1], W.H. van Gilst[1], H.L. Hillege[2], R.S.N. Fehrmann[3], P.I.W. de Bakker[4], R.A. de Boer[1], L. Franke[3], P. van der Harst[1]. 1) Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen , Netherlands; 2) Trial Coordination Center, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; 3) Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen , Netherlands; 4) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands.

Repolarization abnormalities of the heart have been linked to ventricular arrhythmias and sudden cardiac death. Although previous studies of cardiac repolarization have focused largely on interval duration on the electrocardiogram (ECG), changes in the amplitudes may be important as well. Therefore, we performed a GWA meta-analysis on amplitudes of the ST-segment and T-wave (ST-T wave) in 15,943 Dutch individuals. Amplitudes were measured trough the lateral, inferior, anterior, septal and aVR leads of the 12 lead electrocardiogram to define the spatial location. We evaluated 2 time points resulting in 10 traits: 5 for the ST segment (at 80ms after J point) and 5 for the T wave. Across 10 traits we identified 41 genotype-phenotype associations ($p < 6.25 \times 10\text{-}9$ based on 8 independent phenotypes) clustered within 17 independent loci. Twelve of the loci are novel determinants for cardiac repolarization. We observed that SNPs with P<6.25E-09 were 3.3 fold enriched for Dnase I hypersensitivity sites in human fetal heart (Z=3.2; compared to 337 other cell types and tissues) and >10 fold enriched for p300 enhancers in the human heart (P<4E-7). This suggests that functional regulatory variants may contribute to several of the associations observed in this study. Furthermore the associations provide insights on the spatiotemporal contribution of genetic variation involved in cardiac repolarization: SNPs at KCND3, TNKS and KCNB1 showed specific associations with the ST-wave. Other loci (XPO1, KCNH2 and TBX3) were specific for the septal leads, possibly related to the function and location of structures such as the AV bundle and its branches. These analyses provide novel insights into the biology of cardiac repolarization and may help to elucidate diseases of cardiac repolarization. Supported by the Netherlands Heart Foundation (grant NHS2010B280).

2179T

A Three-stage Genome-wide Association Study Combining Multilocus Test and Gene Expression Analysis for Young-Onset Hypertension on Taiwan Han Chinese.
K. Chiang[1,2,3], H. Yang[4], J. Chen[5], W. Pan[1,2,3]. 1) Institute of Biomedical sciences, Academia Sinica, Taipei, Taiwan; 2) Graduate Institute of Life Science, National Defense Medical Center, Taipei, Taiwan; 3) Division of Preventive Medicine and Health Services Research, Institute of Population Health Sciences, National Health Research Institutes, Taiwan; 4) Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan; 5) National Yang-Ming University School of Medical and Taipei Veterans General Hospital, Taipei, Taiwan.

Hypertension is a common and complex disorder. To identify genetic variants will eventually contribute to screening and management efficacy. Although many large-scale genome-wide association studies have been performed, a few studies have successfully identified replicable, large-impact hypertension loci, not to mention the scanty Chinese studies. Young-onset hypertension (YOH) is considered as a more promising target disorder to investigate than the late-onset one due to its stronger genetic component. We performed a three-stage genome-wide association study combining multilocus test and gene expression analysis to map YOH genetic variants. In the first stage, using single locus and multi-locus association tests, we analyzed Illumina HumanHap550 data from 399 YOH cases and 399 age and gender matched controls. In the second stage, differential gene expression analysis was carried out for genes flanked by 14 SNP septets identified in the first stage, using lymphoblast cell lines of the same set of case-control pairs. In the third stage, we re-examined the genetic association for four genes with both allelic and expression association signals in an independent set of 592 YOH cases and 592 controls and verified four genes (C1orf135, GSN, LARS, ACTN4) from the multi-locus test by either the independent sample or a joint analysis that combined the data from 1st and 3rd stages. All four are potential novel hypertension susceptibility genes. Among them, the same septet flanking ACTN4 was replicated by Hong Kong Hypertension Study (HKHS) and by WTCCC hypertension study (WTCCCHS), which has been associated with kidney related diseases, such as focal segmental glomerulosclerosis. LRAS was replicated in the HKHS, but not in the WTCCCHS. GSN which has been implicated in the inflammatory processes in animals may be specific to Taiwanese populations, since it was not validated by HKHS and WTCCCHS.

2180F

Genetic effects on levels of growth differentiation factor 15 - A PLATO genomics study. *N. Eriksson[1, 2], A. Johansson[1, 3], E. Hagström[1, 2], T. Axelsson[4], B.J. Barratt[5], R.C. Becker[6], M. Bertilsson[1], A. Himmelmann[7], S.K. James[1, 2], H.A. Katus[8], A. Siegbahn[1, 9], P.G. Steg[10], R.F. Storey[11], A-C. Syvänen[4], C. Varenhorst[1, 2], A. Åkerblom[1, 2], L. Wallentin[1, 2].* 1) Uppsala Clinical Research Center, Uppsala, Uppsala, Sweden; 2) Department of Medical Sciences, Cardiology, Uppsala University, Uppsala, Sweden; 3) Department of Immunoloy, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 4) Department of Medical Sciences, Molecular medicine, Uppsala University, Uppsala, Sweden; 5) AstraZeneca R&D, Alderley Park, Cheshire, UK; 6) Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA; 7) AstraZeneca Research and Development, Mölndal, Sweden; 8) Medizinishe Klinik, Universitätsklinikum Heidelberg, Heidelberg, Germany; 9) Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden; 10) INSERM-Unité 698, Paris, France; Assistance Publique-Hôpitaux de Paris, Hôpital Bichat, Paris, France; Université Paris-Diderot, Sorbonne-Paris Cité, Paris, France; 11) Department of Cardiovascular Science, University of Sheffield, Sheffield, UK.

Background: Growth differentiation factor 15 (GDF-15), a cytokine produced in cardiovascular cells under conditions of oxidative stress and inflammation, is an independent prognostic biomarker for cardiac events. Genetic markers located close or in the *GDF15* gene have been associated with circulating GDF-15 concentrations. In this genome-wide association study (GWAS) we assess the genetic effects on GDF-15 in a large clinical study on patients with acute coronary syndrome (ACS). **Methods:** We performed a GWAS of 10 013 ACS patients enrolled in the PLATelet inhibition and patient Outcomes study (PLATO). In total 9448 patients passed genotype quality control and had measurements of GDF-15 levels at baseline. Of these, 3133 patients had additional GDF-15 measurement after one month. Clinical and genetic correlates of GDF-15 were assessed in multivariable analyses. **Results:** We identified three independent SNPs with genome-wide significance associated with baseline levels of GDF-15: rs17725099 (P-value = 1.47E-107), rs74180880 (P-value = 6.93E-46) and rs1055150 (P-value = 5.89E-10). All three SNPs were in the same region of chromosome 19, close to the *GDF15* gene, which has been reported previously to affect levels of GDF-15. The effect on levels of GDF-15 at baseline and at one month was similar for the three SNPs (adjusted beta estimates on log scale for rs17725099, rs74180880 and rs1055150 at baseline 0.16, 0.09, 0.08 and at one month 0.16, 0.10 and 0.08, respectively). The identified SNPs are located in predicted regulatory regions, such as an active promoter and strong enhancer, or overlaps with regions predicted to be occupied by transcription factors. Clinical correlates (the top five being age, diabetes, smoking, chronic renal disease and ACS type) accounted for 26.3% of the variation (R2) in log(GDF-15) and the addition of the three top SNPs increased the explained variation by 5.3% to 31.6%. **Conclusion:** We identified three independent SNPs affecting GDF-15 concentrations. The effect was similar at baseline and at one month follow-up, indicating that the genetic effects on GDF-15 are not restricted to the acute phase of ACS. The large effect on GDF-15 levels by genetic variation might influence the GDF-15 based risk-prediction in patients with ACS. Combining these three SNPs in Mendelian-randomization studies could be used to assess the causal effect of GDF-15 on clinical endpoints in longitudinal cohort studies.

2181W

An eQTL mapping strategy identifies the FN3KRP gene as a risk factor for elevated systolic blood pressure the Norfolk Island Genetic Isolate. *L.R. Griffiths[1], M. Benton[1], R.A. Lea[1], D. Mccartney-Coxson[2], M. Carless[3], H. Goring[3], C. Bellis[3], M. Hanna[1], D. Eccles[1], G. Chambers[4], J. Curran[3], J. Blangero[3].* 1) Genomics Research Centre, Griffith Health Institute, Griffith University Gold Coast, Southport, Australia; 2) Kenepuru Science Centre, Institute of Environmental Science and Research, Wellington, New Zealand; 3) Texas Biomedical Research Institute, San Antonio, Texas USA; 4) School of Biological Science, Victoria University of Wellington, New Zealand.

Hypertension affects millions of people worldwide and is influenced by a number of attributing elements, including environmental, lifestyle and genetic factors. Expression quantitative trait loci (eQTL) influence gene expression and therefore make good candidates for functional involvement in risk of hypertension. Pedigrees within founder effect populations may provide some advantage for mapping CVD genes. This study aimed to identify eQTLs associated with hypertension-related traits in the Polynesian founder-effect population of Norfolk Island (NI). We measured genome-wide transcript levels of lymphocytes from 330 individuals forming a single large pedigree from NI. Heritability analysis was used to identify significantly heritable transcripts. A pedigree-based GWAS was conducted on these transcripts to identify cis-acting eQTLs. These eQTLs and their respective transcripts were then tested for association with blood pressure traits - systolic, diastolic, arterial and pulse pressure measures. Results revealed 1712 significantly heritable transcripts within the NI pedigree. The heritability (H2) values ranged from 0.18 to 0.83 with a median of 0.32 ($P<0.05$). A pedigree-based GWAS of the heritable transcripts revealed the presence of 200 cis-acting eQTLs ($P < 1.84\times10^{-7}$). An eQTL-centric association analysis against the blood pressure traits revealed hits at HMBOX1 and FN3KRP genes for SBP, and HOXB2 gene for DBP. Integrative regression modelling of these traits incorporating both transcript level and eQTL genotype, as well as age sex and structure, verified the FN3KRP gene on chr 17q25 as a functional risk factor for elevated SBP ($R2=0.25$, $P=3.3\times10^{-15}$). This study has utilised an eQTL mapping strategy in a large pedigree from a Polynesian founder effect population to functionally implicate the FN3KRP gene in hypertension risk. The FN3KRP gene encodes the fructosamine-3-kinase, an enzyme involved in glucose metabolism. Further studies are required to assess the association of FN3KRP and its biochemical pathways with hypertension comorbidities such as diabetes and obesity.

2182T

Known SNPs in ADAMTS7, the 9p21 region, ZFAND6 and UBE2E interact with type 2 diabetes status to modify the risk of coronary artery disease in large populations. *N.R. van Zuydam[1,2], B. Voight[3], C. Ladenvall[4], R. Strawbridge[5], S. Willems[6], E. van Iperen[7], J. Hartiala[8], E. Vlachopoulou[9], E. Mihailov[10], L. Kwee[11], C. Nelson[12], M. Kleber[13], L. Qu[3], A. Goel[2], J. Kumar[15], S. Kanoni[14], N.W. Rayner[2,14], SUMMIT and CARDIoGRAMplusC4D.* 1) University of Dundee, Dundee, United Kingdom; 2) University of Oxford, Oxford, United Kingdom; 3) University of Pennsylvania, Philadelphia, United States of America; 4) University of Lund, Lund, Sweden; 5) Karolinska Institut, Stockholm, Sweden; 6) Erasmus University Medical Center, Rotterdam, The Netherlands; 7) Academic Medical Center, Amsterdam, The Netherlands; 8) Cleveland Clinic, United States of America; 9) University of Helsinki, Helsinki, Finland; 10) University of Tartu, Talin, Estonia; 11) Duke University, Durham, United States of America; 12) University of Leicester, Leicester, United Kingdom; 13) LURIC Study, Freiburg im Breisgau, Germany; 14) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 15) University of Uppsala, Uppsala, Sweden.

Patients with type 2 diabetes (T2D) are more likely to suffer from coronary artery disease (CAD) than non-diabetic individuals. In this study we aimed to identify loci that modify the risk of CAD in patients with T2D. We combined summary statistics for 2,295,146 (imputed and directly typed) SNPs from 21,623 patients with T2D (8,456 CAD cases and 13,167 CAD free controls) and 36,104 non-diabetic individuals (14,774 CAD cases and 21,330 CAD free controls) in a fixed effects meta-analysis stratified by T2D; we tested known CAD and T2D loci for an interaction with T2D status and performed a pathway analysis on the meta-analysis summary results. The meta-analysis of allelic effects from CAD case-control associations in patients with T2D identified a signal in *ADAMTS7* represented by two independent SNPs previously reported for CAD: rs11072811 (OR=1.2, frequency=0.53, p=8E-12) and rs11634042 (OR=1.1, frequency=0.57, p=4E-08). Rs11072811 had a smaller effect on CAD risk in non-diabetic individuals (OR=1.1, p=1E-02) when compared to its effect in patients with T2D, and this interaction with T2D status was significant (p=9E-03). The meta-analysis of allelic effects from CAD case-control associations in non-diabetic individuals identified the previously published signal in the 9p21 region represented by rs1333042, (OR=1.2, effect allele frequency=0.49, p=5E-20). Rs1333042 had a significant interaction with T2D status (p=3E-04), with a smaller allelic effect in patients with T2D (OR=1.1, p=4E-05). Evaluation of known T2D loci revealed that the T2D risk alleles of rs7612463 (T2D: OR=0.9, P=8E-03; ND: OR= 1.1, P=1E-02) in *UBE2E2* and rs11634397 (T2D: OR=0.9, P=3E-02; ND: OR=1.1, P=3E-03) in *ZFAND6* were associated with decreased risk of CAD in patients with T2D but with increased risk of CAD in non-diabetic individuals ($P_{interaction}$ < 5E-03), when compared to the T2D reference allele. Pathway analysis of the combined T2D and non-diabetic CAD associations identified a REACTOME pathway involved in platelet plug formation associated with CAD overall (p=9E-06), while analysis of CAD associations in patients with T2D found associations (p<1E-03) with pathways involved with cell apoptosis (BIOCARTA), fibroblast growth factor signalling (REACTOME) , nitric oxide biosynthesis and phospholipase A2 activity (GOTERM). This study suggests that known CAD SNPs in *ADAMTS7* and 9p21, and known T2D SNPs in *UBE2E* and *ZFAND6* may differentially modify CAD risk based on T2D status.

2183F

Association of Metabochip variants to systolic blood pressure in African Americans from a biorepository linked to de-identified electronic medical records. *L. Wiley[1,2], R. Goodloe[1], E. Farber-Eger[1], J. Boston[1], D. Crawford[1,3], W. Bush[1,2].* 1) Center for Human Genetics Research, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN.

Blood pressure is a heritable human trait related to a variety of clinical conditions, and hypertensive individuals have increased risk for multiple cardiovascular disorders. The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) BioVU is a subset of the Vanderbilt University biorepository of ~15,000 DNA samples from non-European Americans representing African Americans (n=11,521), Hispanics (n=1,714), and Asians (n=1,122). These samples were genotyped using the Illumina Metabochip to examine influence of these variants on a variety of clinical outcomes. We examined the African American samples to identify genetic variants associated with median outpatient systolic and diastolic blood pressure measurements (excluding patients on blood pressuring lowering medications). There were 11,165 patients (63% female) with all available outcomes and covariates. The median BMI of this population was 25, average systolic and diastolic blood pressure were 74.8 (sd=65.8) and 43.9 (sd=41.6) respectively. Median blood pressure measurements were regressed on 192,270 markers from the Illumina Metabochip adjusting for gender, age, age^2, body mass index (BMI) and the three most significant principal components to adjust for population stratification. Two variants were associated with systolic blood pressure: rs4945935 and rs10857636. The intragenic variant, rs4945935 (MAF=0.02), was associated with a 4.5 point decrease in systolic blood pressure (p=3.4e-6). The other variant, rs10857636 (MAF=0.06), is intronic to *WDFY4* and is associated with a 2.4 point increase in systolic blood pressure (p=7.4e-6). Other variants in this gene have been associated with cardiovascular traits including lipid levels and quantitative subclinical atherosclerosis traits. Additionally this variant is in high linkage disequilbrium (r^2=1 in European populations) with a variant that is an eQTL for the gene *LRRC18*, which has been associated with abdominal aortic calcification. Because blood pressure changes are comorbid with multiple clinical conditions, we further adjusted these models for a global composite of all clinical diagnoses. Both associations are robust to this adjustment indicating that these SNPs directly impact systolic blood pressure and are not mediated through association with another condition.

2184W

Genome-wide association study of leukocyte telomere length identifies casein kinase 2(*CSNK2A2*) to be associated with shorter telomere length and increased cardiovascular disease risk in diabetes. *P. Natt[1], R. Saxena[2], P. Dib[1], S. Ralhan[3], G. Wander[3], D. Stowell[4], M. Payton[4], D. Sanghera[1].* 1) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Massachusetts General Hospital, Harvard Medical School, Boston, MA; 3) Hero DMC Heart Institute, Ludhiana, Punjab, India; 4) Department of Surgery, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Telomere shortening has been associated with multiple diseases including hypertension, myocardial infarction, osteoporosis, type 2 diabetes (T2D), and Alzheimer's disease. In this investigation, we have performed a genome-wide scan (GWAS) to identify the common genetic variation that may influence the relative leucocyte telomere length (LTL) and cardiometabolic risk in a diabetic cohort of Punjabi Sikhs from India and the US. Our results revealed a significant association of LTL with coronary heart disease (CHD) the mean LTL showed a gradual decline from healthy subjects to individuals with T2D and CHD showing respective mean LTL being 2.10 in healthy, 1.95 in T2D, 1.69 in CHD, and 1.59 in T2D+CHD. GWAS analysis of discovery cohort (n=1,616/842 T2D cases) identified 338 top independent signals ($p<10^{-4}$) to be significantly associated with LTL. Most promising 48 SNPs were further replicated through genotyping in an additional Punjabi Sikh sample (n=2,397/1,108 cases). On combined meta-analysis in Sikh populations (n=4,013/1,946 cases), we identified a novel locus in association with LTL at 16q21 represented by an intronic SNP rs74019837 in the *CSNK2A2* gene ($p = 4.4\times10^{-8}$). We also found significant association signals near *SPATA4*, *C5orf42*, and **FER** genes for affecting LTL with p values ranging from 2.7×10^{-5} to 3.3×10^{-6}. Our findings report an independent association of shorter telomere length with T2D and cardio metabolic risk. Interestingly, telomeric repeat binding factor 1(TRF1) serves as a substrate for *CSNK2A2*, which phosphorylates and initiates its binding to telomere. *CSNK2A2* also interacts with multiple genes and miRNAs in pathway controlling telomere length and cardiovascular disease. Future functional studies may provide clinically important insights on the interplay of genetic variation in *CSNK2A2* and environmental and lifestyle factors for affecting LTL and cardiometabolic risk in diabetes.

2185T

A Rho-GTPase pathway related gene is associated with chronic kidney disease via an interaction with coronary artery disease. *E. Hauser[1], C. Ward-Caviness[1], M. Winn[1,2], C. Blach[1], C. Haynes[1], E. Dowdy[1], S. Gregory[1], S. Shah[1,2], W. Kraus[1,2].* 1) Center for Human Genetics, Duke Univ Med Ctr, Durham, NC; 2) Department of Medicine, Duke Univ Med Ctr, Durham, NC.

Chronic kidney disease (CKD) and coronary artery disease (CAD) are intimately linked and often co-occur, implying they may share some aspects of their genetic basis. The genes related to the Rho-GTPase family been implicated in the pathogenesis of CAD in multiple studies. As regulators of the actin cytoskeleton Rho-GTPase genes have an important function in many cellular processes and potentially a function in the pathogenesis of CAD and CKD. Here we present a Rho-GTPase gene that interacts with CAD and is associated with CKD. To discover this association we conducted a genome-wide interaction study in a cardiac catheterization cohort, CATHGEN. We used an ordinal measure of atherosclerosis assessed during the catheterization, CADindex, as a measure of the extent of severity of CAD to test the hypothesis that specific genetic variants (SNPs) that interact with measures of CAD are associated with CKD prevalence CATHGEN. We used race-stratified cohorts of 2202 whites (EA) and 663 blacks (AA) to identify relevant interactions, and a subsequent meta-analysis filtered for consistent associations. We estimated 905,956 SNP-CADindex interactions via a logistic regression model, with CKD as the dependent variable (outcome). A Score test was used to calculate P-values (P), and we adjusted for age, race specific principal components, sex, BMI, hypertension, hyperlipidemia, diabetes, and smoking. We identified the 10 most significant interactions in each cohort, then filtered for those that validated, i.e. $P < 0.05$ and consistent direction of association in the other cohort. The most significant EA SNP to validate was rs11014215 located in the 5' regulatory region of the Rho-GTPase gene *ARHGAP21* (EA $P=5.6\times10^{-6}$, AA $P=0.045$, meta-analysis $P=1.3\times10^{-6}$). *ARHGAP21* is a cytoskeletal protein involved in cellular differentiation and is an activator of Cdc42, a cell cycle protein. Examining other interactions in *ARHGAP21* revealed one additional association, rs12219181 (*ARHGAP21* intron, EA $P = 0.032$). These analyses use a novel interaction analysis to identify rs11014215, located 5' of *ARGHGAP21*, as associated with CKD in a SNP-CADindex interaction model. This SNP was strongly associated with CKD in the EA cohort and validated in the AA cohort. Given the previous associations of Rho-GTPase family genes with CAD and the co-occurrence of CAD and CKD we believe this interaction brings novel information to understanding these two important chronic diseases.

2186F

Identifying interactions among social, psychosocial, and genetic factors that influence blood pressure in three multi-ethnic epidemiological cohorts: the Multi-Ethnic Study of Atherosclerosis (MESA), the Atherosclerosis Risk in Communities Study (ARIC), and the Health and Retirement Study (HRS). *J.A. Smith[1,2], W. Zhao[1,2], B. Needham[1-3], J.D. Faul[4], T.E. Seeman[5], E. Boerwinkle[6], A. Chakravarti[7], D.R. Weir[4], S.L.R. Kardia[1,2], A.V. Diez Roux[1-3].* 1) Department of Epidemiology, University of Michigan, Ann Arbor, MI; 2) Center for Integrative Approaches to Health Disparities, University of Michigan, Ann Arbor, MI; 3) Center for Social Epidemiology and Population Health, University of Michigan, Ann Arbor, MI; 4) Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI; 5) University of California Los Angeles, Division of Geriatrics, Los Angeles, CA; 6) Human Genetics and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas; 7) Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.

Substantial health disparities in the prevalence, severity, and age of onset of hypertension exist across ethnic groups in the United States. Traditional approaches for examining the causes of these disparities have been fragmented into 'biological' and 'social' approaches. However, blood pressure levels are likely influenced by the interaction between multiple biological and social/psychosocial risk factors. We used data from 20,495 non-Hispanic whites and 6,307 African Americans from three large multiethnic cohorts (the Multi-Ethnic Study of Atherosclerosis (MESA), the Atherosclerosis Risk in Communities Study (ARIC), and the Health and Retirement Study (HRS)) to examine the effects of interactions between common SNP variation at 29 known blood pressure loci and social factors (education, parental education) or psychosocial factors (anger) on blood pressure levels. In gene-based analysis using the Gene-Environment Set Association Test (GESAT), we found evidence of gene-by-anger interaction for systolic blood pressure at two loci in whites (*C10orf107*, p=0.007; *PLEKHA7*, p=0.007) and two loci in African Americans (*NT5C2*, p=0.009; *ULK4*; p=0.003). For diastolic blood pressure, gene-by-anger interaction was detected at two loci in whites (*C10orf107*, p=0.0002; *PLCE1*, p=0.005). Identifying the factors that lead to the persistent differences in the blood pressures among these populations is critical for developing effective strategies to reduce disparities.

2187W

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Heritability of Serum Lipids. *T.J. Hoffmann[1], M.N. Kvale[1], Y. Banda[1], E. Jorgenson[2], S.E. Hesselson[1], L. Walter[2], S. Sciortino[2], D.K. Ranatunga[2], D. Ludwig[2], C. Iribarren[2], P. Kwok[1], C. Schaefer[2], N. Risch[1,2].* 1) University of California San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern Division of Research.

Plasma lipid concentrations are an important heritable risk factor for cardiovascular diseases that are often targets for therapeutic interventions. The Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort contains extensive longitudinal lipid, prescription medication, and genotypes (over 675,000 markers) for nearly all of its 100,000 participants. Approximately 1 million HDL and LDL cholesterol measures from electronic health records were linked to genetic data for this analysis.

Genome-wide association studies have failed to account for a large amount of the heritability that has been estimated from family-based heritability calculations. Recently, an approach has been developed to use genome-wide association data to use the heritability of less-related individuals to estimate the heritability of a trait. This approach has not been systematically compared to heritability estimates derived from first, second, and third-degree cryptic related individuals identified in a large GWAS cohort. Preliminary results for total cholesterol show an apparent maternal effect of a higher mother-child heritability of .606 (.065 sd) than father-child (genetically determined) of .472 (.084), and a sibling heritability of .463 (.055). Overall, first degree relatives had a heritability of .486 (.038) and second degree .299 (.109), but kinship estimates from King robust were not stable enough for further degree relatives. Heritability estimates utilizing all genotyped SNPs from GCTA are .282(.018), while that of the independent significant genotyped SNPs was .12, and of independent significant imputed SNPs was .14. After regressing off the significant imputed SNPs, GCTA gave the expected reduced heritability of .141 (se .018), father-child .335 (.081), mother-child .404 (.065), full sibs .411 (.056), and first-degree .353 (.038). Using the significant SNPs, we also tested for epistasis and dominance interactions, but neither appreciably increased the variance explained. Nor did reducing the threshold for genome-wide significance to 1×10^{-6}. Similar analysis for LDL, HDL, and triglycerides will also be presented.

2188T

A family-based linkage and association studies reveal new variants near *THBS2* and *ACE* for blood pressure. *H. Kim[1], S. Im[1], S. Lee[1,2], S. Cho[3], J. Sung[3], K. Kim[1,2,4], J. Seo[1,2,5].* 1) Medical Research Center, Genomic Medicine Institute (GMI), Seoul National University; 2) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul 110-799, Korea; 3) Seoul National University School of Public Health, Seoul 151-742, Korea; 4) Psoma Therapeutics Inc., Seoul 153-781, Korea; 5) Macrogen Inc., Seoul 153-781, Korea.

High blood pressure (BP) causes global health problem as a major risk factor for cardiovascular disease and death. BP is heritable, with heritability ranging from 18% to 37%. Although recent genome-wide association studies (GWAS) have substantially contributed to understand the genetic architecture for BP, discovering new variants for explaining missing heritability still remains a major challenge. Moreover, GWAS might miss the true-positive associations within linkage regions because of too strict significance level for multiple testing. To uncover the novel candidates for BP, we performed fine mapping study under suggestive linkage regions as well as family-based GWAS using an isolated Mongolian sample of 751 individuals from 53 families. Our GWAS revealed 3 single-nucleotide polymorphisms (SNP)s for SBP and 5 SNPs for DBP reaching a significance level ($P < 1 \times10^{-5}$), which included several candidate genes such as *ACE*, *BMPER*, and *IL5*. Of these, rs4459609, an intergenic SNP near *ACE* gene (5.5 Kb away) showed a significant effect on both SBP ($P = 4.75\times10^{-6}$) and DBP ($P = 8.59\times10^{-6}$). For fine mapping association test under linkage peaks, we firstly observed two suggestive linkage regions on chromosome 6q27 (LOD for SBP = 2.4) and 5q35 (LOD for DBP = 2.0), which were consistent with previously reported linkage peaks in other populations. The subsequent association test in presence of linkage found 7 and 2 SNPs in each linkage region, respectively ($P < 5\times10^{-3}$). Top association signal under linkage peak was found for rs9393162, an intergenic SNP near THBS2 (21.3 kb away). We also confirmed that the 27 loci identified in previous GWAS were nominally replicated ($P < 0.05$), and of these loci, ADM and ZNF652 for SBP were still significant after multiple testing corrections ($P <$ (0.05/ the number of effective SNPs)). In addition, we examined associations with 10 anti-hypertensive drug target genes, and *CA1* and *SCNN1A* were statistically significant in Mongolian families. In conclusion, our strategic analysis combining linkage and association enabled the discovery of novel variants associated with BP near *THBS2* and *ACE* in Mongolian families.

2189F

Higher Degree of African Ancestry in Transfused African American Sickle Cell Patients is Associated with Red Cell Alloimmunization. *RA. Kittles, R. Machado, NA. Ellis, K. Batai, JN. Garcia, SA. Campbell-Lee.* University of Illinois Hospital and Health Sciences System, Chicago, IL.

Background: Red cell alloimmunization in transfused sickle cell disease patients occurs in approximately 30%. Because red blood cell (RBC) transfusion is an important part of therapy, the need for additional antigen matching once alloimmunization occurs is problematic, and leads to therapeutic limitations. Thus, it would be important to identify risk factors for alloimmunization in this patient population. We performed a genome wide association study (GWAS) in an attempt to identify possible genes associated with RBC alloimmunization. Methods: 157 sickle cell disease patients, including 69 who developed alloimmunization and 88 who did not were genotyped for 2,217,402 SNPs using the Affymetrix PanAFR Array. SNPs with a call rate <90%, Hardy Weinberg Equilibrium P<0.0001, or having a minor allele frequency (MAF) <1% were excluded from the analysis. In total, 2,113,177 SNPs were analyzed. Population structure was examined by principal components analysis (PCA) using the GWAS SNPs. Statistical analyses were performed using Golden Helix SVS 7. To correct for multiple testing, the genome wide significance threshold was set at $P<2.4 \times 10^{-8}$. We also selected 2,104 unlinked ancestry informative markers from the Affymetrix PanAFR array and estimated West African Ancestry (WAA) using STRUCTURE. Results: Logistic regression analyses revealed that PC 1 was significantly associated with RBC alloimmunization (P=0.04), and SCD patients in the highest ancestry quartile (>85% WAA) had significantly increased risk of RBC alloimmunization compared to patients in the lowest quartile (<75% WAA) (OR=2.7, 95% C.I.; 1.02-7.07). In the lowest WAA quartile, 18.8% of the patients were alloimmunized, while 29.5% of the patients were not. In the highest WAA quartile, 29.0% of patients were alloimmunized and 20.5% were not. No SNPs reached genome-wide significance for the association with RBC alloimmunization, although three SNPs, rs1507094 (*RALYL*), rs4345649 (*POLR1E*), and rs10941424 (*FYB*) almost reached genome-wide significance (P=1.31 × 10-6; 2.82 × 10-6; and 3.21 × 10-6, respectively). Of note, the *FYB* gene on 5p13.1 is a binding protein involved in platelet activation and expression of IL-2. Conclusions: In our study, patients with higher WAA were more likely to develop RBC alloantibodies. A larger sample size would be necessary to find SNPs which reach genome-wide significance for RBC alloimmunization.

2190W

Identifying population specific dyslipidemia variants using cross-population GWAS. *A. Ko[1], D. Weissglas-Volkov[1], R.M. Cantor[1], E. Nikkola[1], K.A. Deere[1], J.S. Sinsheimer[1], B. Pasaniuc[2], R. Brown[3], V. Salomaa[4], J. Kaprio[4,5,6], A. Loukola[5], A. Jula[4], M. Jauhiainen[4], M. Heliövaara[4], O. Raitakari[7,8], T. Lehtimäki[9], J.G. Eriksson[4,10,11], M. Perola[4,6], L. Riba[12,13], T. Tusie-Luna[12,13], C.A. Aguilar-Salinas[12], P. Pajukanta[1].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) 2Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA; 3) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Public Health, Hjelt Institute, University of Helsinki, Helsinki, Finland; 6) Institute for Molecular Medicine Finland FIMM, University of Helsinki,Helsinki, Finland; 7) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 8) Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 9) Department of Clinical chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland; 10) Folkhälsan Research Center, University of Helsinki, Helsinki, Finland; 11) Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland; 12) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 13) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

The high prevalence of dyslipidemia in Mexicans is a serious health burden due to the increased risk for cardiovascular disease and medical cost. Identifying Hispanic-specific variants can accelerate our understanding of lipid disease pathogenesis in this rapidly growing U.S. minority. However, studies in admixed populations are hindered by complex population substructure that can reduce power. To address this, we conducted a genome wide cross-population study using subsamples that were selected by triglyceride (TG) levels (total n=9,065) from Mexican and Finnish cohorts. The Finnish low TG controls were frequency matched for age, BMI, and sex to eliminate known TG confounders, because the Mexican low TG controls were significantly more obese (p=2.20×10-16) despite their younger age. To increase power by reducing the candidate SNPs to those that are likely to be specific to Mexicans, we compared allele frequencies of the Mexican and Finnish low TG controls using a p-value<0.001. About 90% of the SNPs passing this threshold were ancestry informative markers (AIMs) based on the CEU and MEX HapMap data, suggesting that ~10% additional trait-related variants can be found using the cross-population GWAS approach. Using this tailored set of variants to compare Mexican high TG cases and low TG controls on a pilot study on chromosome 8 with 18,426 SNPs, we discovered two experiment-wide significant TG-associated SNPs (rs4128744, p=2.61×10-8 and rs2688358, p=3.77×10-6). To evaluate this approach, we employed local ancestry inference using LAMP-LD as an independent method to examine ancestral origins across chromosome 8. Remarkably, 7 of the top 10 TG-associated SNPs from our cross-population pilot study resided in regions with enrichment for Amerindian ancestry in the Mexican TG cases versus controls, suggesting that they represent underlying susceptibility variants in the Hispanic-specific, trait-linked regions. Given this result, our design should significantly help replication and fine mapping efforts. We also explored the top 10 SNPs for a cis effect of nearby genes on Genevar analysis of ~850 publicly available adipose RNA microarray samples. Five SNPs showed nominal to significant evidence for a regulatory cis-eQTL effect in adipose. In summary, we demonstrate that the cross-population GWAS strategy in combination with local ancestry inference is a powerful approach to identify population-specific variants that may otherwise be missed by conventional GWAS.

2191T

Genome wide association study for Arteriosclerosis Obliterans in a Japanese population. *K. Ozaki[1], T. Morizono[2], Y. Onouchi[1], A. Takahashi[3], T. Tsunoda[2], M. Kubo[4], Y. Nakamura[5,6], T. Tanaka[1,7].* 1) Laboratory for Cardiovascular Diseases, RIKEN, Ctr Integrative Medical Science, Yokohama, Kanagawa, Japan; 2) Laboratory for Medical Science Mathematics, RIKEN, Ctr Integrative Medical Science, Yokohama, Kanagawa, Japan; 3) Laboratory for Statistical Analysis, RIKEN, Ctr Integrative Medical Science,; 4) Laboratory for Genotyping Development, RIKEN, Ctr Integrative Medical Science,; 5) Human Genome Center, Institute of Medical Science, the University of Tokyo, Tokyo, Japan; 6) Section of Hematology/Oncology, The University of Chicago; 7) Department of Human Genetics and Disease Diversity, Tokyo Medical and Dental University.

Arteriosclerosis obliterans (ASO) is the most common cause of peripheral vascular disease affecting the lower limbs. ASO mainly occurs from atherosclerosis, resulting in obstruction of the blood supply to the lower or upper extremities. ASO might result from complex interactions multiple genetic and environmental factors. To reveal genetic backgrounds in the pathogenesis for ASO, we performed a genome wide association study (GWAS) in a Japanese population. Genotyping was performed by the Illumina HumanHap610-Quand BeadChip for cases and by Illumina HumanHap550v3 for controls. We applied stringent quality-control criteria and tested 789 cases and 3,383 controls for 510,687 autosomal SNPs commonly available on both BeadChip. The inflation of test statistics, $\lambda_{genomic\ control}$ (λ_{gc}) was 1.04. Through combination of this GWAS and a following staged analysis with a total of 2,647 Japanese subjects with ASO and 20,560 control subjects, we have identified two susceptible loci for this disorder on chromosome 13 and 4 (P values < 10^{-8}). To our knowledge, these are the first genetic risk factors identified for ASO.

2192F

Common autosomal variants are associated with bicuspid aortic valve in Turner Syndrome. *S.K. Prakash[1], M. Silberbach[2], S. Hooker[3], D.C. Guo[1], C. Maslen[2], C.A. Bondy[4], D.M. Milewicz[1], GenTAC Investigators.* 1) University of Texas Health Science Center at Houston, Houston, TX; 2) Oregon Health & Science University, Portland, OR; 3) Baylor College of Medicine, Houston, TX; 4) National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD.

Background: The prevalence of bicuspid aortic valves (BAV) is enriched thirty-fold in women with Turner Syndrome (TS) in comparison with the general population. We hypothesize that common autosomal variants influence the development of BAV in TS women, who may be uniquely sensitized to these variants by the loss of one X chromosome. We sought to identify autosomal BAV susceptibility genes in a cohort of TS women (average age 33 years, 38% BAV, 15% coarctation). Methods: 73 TS women of European ancestry with BAV and 120 TS women with tricuspid aortic valves were genotyped using Illumina Omni-Express arrays (668,262 SNPs). Tests of association were performed using logistic regression without adjustment for covariates. Results: TS women with 45,X karyotypes were not significantly more likely to have a BAV than those with isochromosomes, rings or Xp deletions (*P*=0.09). After removal of 11 outlier samples for the first two multidimensional scaling components, the genomic inflation factor was 1. The strongest association signals were observed on chromosomes 2, 5 and 22 and did not overlap with previously reported loci for BAV. A total of 16 SNPs in the *ITGA1* locus on 5q11.2 were positively associated with BAV (OR=4.3) with a minimum *P* value of 1.5×10-5. *ITGA1*, which encodes the alpha-1 intergrin, is a promising BAV candidate gene because it is expressed in valve tissue and is required for vascular cell adhesion and myofibroblast differentiation. Replication of these regions in independent groups of cases is ongoing. Conclusion: This is the first study to demonstrate that autosomal variants are associated with BAV in TS women, and provides evidence for gene-gene interactions in BAV formation.

2193W

Exome-wide Coding Variation and Myocardial Infarction. *H. Zhang[1], O.L. Holmen[2,3], E. Schmidt[1,4], M. Lochen[5], C. Platou[2], E.B. Mathiesen[6,7], L. Vatten[8], T. Wilsgaard[5], F. Skorpen[9], M. Boehnke[10], G.R. Abecasis[10], I. Njolstad[5], K. Hveem[2,11], C.J. Willer[1,12].* 1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway; 3) St. Olav Hospital, Trondheim University Hospital, Trondheim, Norway; 4) Program in Bioinformatics, University of Michigan, Ann Arbor, Michigan; 5) Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway; 6) Brain and Circulation Research Group, Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway; 7) Department of Neurology and Neurophysiology, University Hospital of North Norway, Trosmø, Norway; 8) Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway; 9) Department of Laboratory Medicine, Children's and Women's Health Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway; 10) Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan; 11) Levanger Hospital, Levanger, Norway; 12) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan.

Genome-wide association studies have identified common variants that are associated with coronary artery disease or myocardial infarction (MI). While these common variants are robustly associated with CAD, they have small effect sizes and collectively explain less than 10% of the heritability in MI (currently estimated to be 40-60%). One explanation for the 'missing heritability' is the existence of low frequency (1-5%) or rare (<1%) variants with larger effect sizes. However, the majority of low frequency variants have not been included in modern GWAS arrays and cannot be well imputed from modern reference panels based on HapMap or 1000 Genome Project samples; thus their contribution to the risk of MI has not been systematically evaluated. In this project, as a first step towards addressing the potential contribution of these low frequency coding variants on the risk of MI, we addressed the question of whether some of these variants have moderate to high effect sizes as proposed. Using the Illumina HumanExome Beadchip, we genotyped 5,937 samples from the Nord-Trøndelag Health Study (HUNT): 2,969 medical-record confirmed acute myocardial infarction cases and 2,968 healthy controls. This array allowed for direct genotyping of 80,533 polymorphic coding variants as well as 44 known GWAS SNPs. We followed-up 21 coding variants in an additional 4,668 samples from the Tromsø study (Tromsø 4) including 2,350 MI cases and 2,318 controls, all from Norway. evaluated the array coverage in these Norwegian samples using whole genome sequencing combined with exome enrichment in 162 samples. The exome array provided successful genotyping for an estimated 74% of Norwegian loss-of-function or missense variants with frequency > 5%, 73% of coding variants with frequency 1-5% and 51% of rare coding variants with frequency < 1%. Despite sufficient power to detect association with low frequency variants (1-5% frequency) with moderate to high effect sizes (odds ratio 1.38-1.91) in the combined sample size of 10,333 individuals, we did not identify any novel genes or single variants that reached significance. However, we identified several strong candidate genes hovering below study-wide significance levels suggesting that low frequency and rare variants with intermediate effect sizes will likely be identified with larger samples.

2194T

Using common genetic variants to predict dyslipidemia. *C.M. van Duijn[1], S.M. Willems[1], E.J.F. Sijbrands[2], A. Dehghan[1], A. Hofman[1], B.A. Oostra[3], O.H. Franco[1], A. Isaacs[1].* 1) Dep Epidemiology, ErasmusMC, Rotterdam, Netherlands; 2) Department Internal Medicine, ErasmusMC, Rotterdam, Netherlands; 3) Department Clinical Genetics, ErasmusMC, Rotterdam, Netherlands.

Hypercholesterolemia (HC) is an important modifiable cardiovascular disease risk factor. We determined the ability of genotypic risk scores to identify individuals at increased risk of hypercholesterolemia. We calculated TC, LDL-C, HDL-C and TG risk scores in the Rotterdam Study (RS, n=10,072) and Erasmus Rucphen Family Study (ERF, n=2,715) using 157 SNPs discovered in recent meta-analyses of ≥100,000 individuals. Adding the TC score to a clinical model based on age, sex and BMI improved the prediction of HC (c-index in RS increased from 0.64 to 0.68 (P=2.1×10-31) and in ERF from 0.71 to 0.75 (P=2.9×10-9)). The LDL-C, HDL-C and TG scores combined improved prediction similarly. In RS, age specific risk showed an approximately doubled cumulative incidence of HC in the highest TC risk score quartile compared to the lowest. We compared the risk curves with carriers of Mendelian forms of dyslipidemia.In both cohorts, overall and in age strata, prevalence of HC increased with TC risk score quartile (Ptrend= 1.3×10-96 to 0.002). Odds ratios also increased with quartile, from 1.62 for the second to 3.5 for the highest quartile compared to the lowest in the total sample. Maximum OR was 3.9 for the highest quartile in the 65-74 years stratum (P=2.4×10-32). Our results show that genetic risk scores can be useful in identifying individuals at increased risk of hypercholesterolemia. Our findings also suggest that a genetic risk score can help identify individuals that would benefit from earlier and more intensive monitoring of cholesterol levels.

2195F

Genome-wide association study of laboratory markers in Kawasaki disease. *J. Kim, J. Lee, Korean Kawasaki Disease Genetics Consortium.* Asan Institute for Life Sciences, University of Ulsan College of Medicine, Seoul, South Korea.

Kawasaki disease (KD) is an acute self-limited vasculitis of infants and children, manifested by fever and signs of mucocutaneous inflammation. Children with KD show laboratory abnormalities of inflammation, such as elevated white blood cell (WBC) count, C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), and they also may have anemia, thrombocytosis, hypoalbuminemia, and elevated serum transaminases, etc. To identify genetic loci influencing important laboratory markers in KD - WBC count, neutrophil count, platelet count, CRP, ESR, hemoglobin (Hb), aspartate aminotransferase (AST), alanine aminotransferase (ALT), albumin, and total protein, we performed association studies using our previous genome-wide association study (GWAS) data of KD. Linear regression analyses after adjustment for age and sex were carried out for the 10 quantitative traits in 178 KD patients. A total of 165 loci passed our arbitrary stage 1 threshold for replication ($P < 1 * 10^{-5}$). For CRP, variants in the *CRP* gene showed the most significant associations ($P = 1.20 * 10^{-6} \sim 5.69 * 10^{-6}$) and for WBC count, a variant in the *TNFRSF13B* gene reached genome-wide significance level ($P = 2.80 * 10^{-8}$). For Hb, variants in the *SOX5* gene were the most significantly associated ($P = 9.40 * 10^{-7} \sim 4.41 * 10^{-6}$) and for albumin, the most compelling association involved a variant in the *CDKAL1* gene which was reported as being associated with albuminuria ($P < 8.73 * 10^{-6}$). Of the significant variants from the stage 1 GWA analysis, 22 were selected as candidates considering their functional importance. Further replication studies are planned to validate these results.

2196W

Novel association of endothelial function with a variant in *PEAR1*. *A.S. Fisch[1], P. Donnelly[1], M. Drolet[1], S. Newcomer[1], A. Parihar[1], K.A. Ryan[1], W. Herzog[2], A.R. Shuldiner[1], J.P. Lewis[1].* 1) Program in Personalized and Genomic Medicine, and the Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Department of Medicine, Johns Hopkins Medical Institute, Baltimore, MD.

The anti-platelet agent aspirin is one of the most commonly prescribed medications for prevention and treatment of coronary heart disease. However, there is great inter-individual variation in response to aspirin as measured by on-treatment residual platelet aggregation and cardiovascular outcomes. In the Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study, a genome-wide association study (GWAS) in 670 healthy Old Order Amish (OOA) participants, we identified a genome-wide significant association between on-aspirin collagen-stimulated platelet aggregation and *PEAR1* genotype (rs12041331). The rs12041331 variant was also associated with cardiovascular event risk in two independent populations. *PEAR1* (platelet endothelial aggregation receptor 1) is a platelet-platelet contact receptor that modulates platelet aggregation. In addition to platelets, *PEAR1* is highly expressed in endothelial cells, which led us to hypothesize that the variant's association with cardiovascular risk may be mediated through effects on endothelial function. To investigate this hypothesis, we genotyped rs12041331 in participants of the Heredity And Phenotype Intervention (HAPI) Heart Study, a clinical intervention study with 868 healthy OOA participants in whom cardiovascular traits were assessed, including platelet aggregation and endothelial function by brachial artery flow-mediated dilation (FMD). In this cohort, we confirmed significant association between on-aspirin collagen-stimulated platelet aggregation and *PEAR1* genotype (effect size = 6.56, p = 0.037), with the minor A allele associated with decreased platelet aggregation. In addition, we found significant association between *PEAR1* genotype and FMD (effect size = 1.14, p = 0.039), with the A allele associated with increased FMD, which represents greater endothelial function. These results indicate that *PEAR1* likely plays a dual functional role in platelets and endothelium, and that *PEAR1* rs12041331 genotype is associated with both. Future studies will be required to more precisely understand the mechanisms by which *PEAR1* and its genetic variants influence aspirin response, ultimately leading to the development of more individualized and effective anti-platelet therapy.

2197T

Novel Susceptibility Loci for Resting Heart Rate: Common Genetic Variant in KREMEN1 Is a Replicable and Genome-Wide Significant in Korean and Mongolian Populations. *B. Gombojav[1, 2], D.H. Lee[3], M.K. Lee[3], M. Kho[3], Y.J. Yang[3], E. Purevdorj[7], Y.M. Song[5], K. Lee[6], J.I. Kim[2,4], J. Sung[1,3], J.S. Seo[2,4].* 1) Epidemiology, Seoul National University, Seoul, Seoul, South Korea; 2) Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Korea; 3) Complex Disease and Genome Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Seoul, Korea; 4) Department of Biochemistry and Molecular Biology, College of Medicine, Seoul National University, Seoul, Korea; 5) 5 Department of Family Medicine, Samsung Medical Center, and Center for Clinical Research, Samsung Biomedical Research Institute, School of Medicine, Sung Kyun Kwan University, Seoul, Korea; 6) Department of Family Medicine, Busan Paik Hospital, College of Medicine, Inje University, Busan, Korea; 7) Department of Higher Education, Ministry of Education and Science, Ulaanbaatar, Mongolia.

Background - Previous studies have identified an increased resting heart rate as an independent risk factor of cardiovascular morbidity and mortality. Although heritable factors play a substantial role in regulating the heart rate, evidence about specific genetic determinants is limited. Objective -To identify common genetic variants associated with the resting heart rate, we did a genome-wide association study (GWAS). Methods - A total of 537,159 markers were tested for their association with the resting heart rate in a GWAS. The GWAS sample consisted of 3,029 healthy individuals from 661 families of the Healthy Twin Study, Korea (HT). The most promising markers from the GWAS were subsequently tested in a Mongolian replication samples comprised of 1,013 individuals from 73 extended families of the GENDISCAN study. Results - The SNP in the locus 22q12.1 at KREMEN1 [rs469983, minor allele frequency (MAF) = 0.130; an odds ratio = 1.74, 95% CI 1.27-2.37; combined P = 4.6×10-8] reached the genome-wide significance in combined analysis. In aggregate, this variant explains 0.6 % of resting heart rate variance. In addition, we replicated four novel susceptible loci in the replication sample: 17p12 at ADORA2B; 14q21.1 at LRFN5; 5q31.2 at SPOCK1 and 11q14.3 at NOX4. Conclusion - We have identified five novel susceptible loci for resting heart rate and kringle-containing transmembrane protein 1, KREMEN1 reached the genome-wide significance. Further studies are required to replicate these findings and to elucidate the biological mechanisms.

2198F

PMA(phorbol 12-myristate 13-acetate) regulates in vitro alternative splicing of ORL1, a gene, involved in atherogenesis and tumorigenesis.
E. Morini[1,2], J.R. Tejedor[3], B. Rizzacasa[2], M.C. Bellocchi[4], F. Ferrè[5], A. Botta[2], D. Caporossi[1], J. Valcárcel[3], G. Novelli[2,6,7], F. Amati[2]. 1) Dept. of Movement, Humanities and Health Sciences, "Foro Italico" University, Rome, RM, Italy; 2) Dept. of Biomedicine and Prevention, Tor Vergata University, Rome, RM, Italy; 3) Centre de Regulació Genòmica and Universitat Pompeu Fabra, Barcelona, Spain; 4) Dept. of Sperimental Medicine and Surgery, Tor Vergata University, Rome, RM, Italy; 5) Dept. of Biology, Tor Vergata University, Rome, RM, Italy; 6) Dept. of Internal Medicine, University of Arkansas for Medical Sciences and Central Arkansas Veterans Healthcare System, Little Rock, AR, USA; 7) St. Peter Fatebenefratelli Hospital, Rome, RM, Italy.

Lectin-like oxidized low-density lipoprotein receptor-1 (LOX-1), encoded by the OLR1 gene, is the major endothelial receptor for ox-LDL and plays a fundamental role in the pathogenesis of atherosclerosis. Furthermore recently, it has been demonstrated that an increased activity of LOX-1 is associated with cancer cell invasion. OLR1 is subjected to a physiological alternative splicing; its isoform, Loxin, lacks exon 5 and encodes for a putative truncated receptor (LOXIN) with impaired binding activity. LOXIN is considered as a natural inhibitor of LOX-1-mediated signalling and each effort to increase LOXIN expression may have a potential therapeutic effect. The expression of OLR1 and Loxin is regulated by six intronic SNPs in linkage disequilibrium and the H-risk haplotype is associated to acute myocardial infarction (AMI) and atherosclerosis. The macrophages of individuals with H-risk haplotype present a OLR1/Loxin mRNA ratio 33% higher compared to individuals with L-risk haplotype. In order to study in detail the regulation of OLR1/Loxin, we have analyzed the expression of the two isoforms in in vitro cell models (THP-1 and HeLa). We found that treatment with 100nM PMA increases OLR1 and Loxin expression levels; this increment is due to a PMA-mediated induction of OLR1 promoter. However, we observed that Loxin splicing was significantly increased after 3h ($p<0.00001$), while OLR1 splicing after 30' ($p<0.000001$). These results suggest a dual regulation of OLR1 splicing by PMA. To test this hypothesis we transfected HeLa cells with two minigenes (HIGH-risk and LOW-risk) containing the genomic sequences spanning from exon 4 to the whole 3'UTR of the OLR1 gene. These minigenes carry two different haplotypes of the six SNPs regulating OLR1 splicing. We performed a dose-curve PMA treatment at 3h and interestingly, we found that 10nM PMA up-regulates Loxin expression (FC10nM=+1.6, $p<0.05$) in HeLa transfected with H-risk (Hela-H); Loxin increase was not followed by OLR1 up-regulation and the OLR1/Loxin ratio after 10 nM of PMA was lowered to 28.1% compared to untreated HeLa-H. These results demonstrate a role of PMA in modulating OLR1 splicing in vitro, leading to an increase of the less functional Loxin isoform. At the best of our knowledge these are the first evidences of a regulation of OLR1 splicing, and provide new data addressing a future more selective and personalized therapy for diseases caused by OLR1 over-expression.

2199W

Identification of a predictive/prognostic genetic signature in Chagas Cardiomyopathy: A systems biology approach on the site of action.
C. Chevillard[1], L. Ferreira[2,3], S. Canbantous[1], A.F. Frade[2,3], P.C. Teixeira[2,3], B.M. Ianni[2], C.W. Pissetti[4], B. Saba[5], L.H.T. Wang[5], A. Kuramoto[2,3], L.G. Nogueira[2,3], P. Buck[2], F. Dias[6], M. Baron[2,3], S. Alves[2], E.C. Mairena[2,3], A. Schmidt[6], E. Donadi[6], J.A. Marin-Neto[5], M. Hirata[5], M. Sampaio[5], A. Fragata[5], E.A. Bocchi[7], A.N. Stolf[2], A.I. Fiorelli[2], R.H.B. Santos[2], V. Rodrigues Jr[4], A.C. Pereira[1,2], J. Kalil[2,3,7], E. Cunha-Neto[2,3,7]. 1) Fac Medicine, INSERM U906, Marseille cedex 5, France; 2) Heart Institute (InCor), University of São Paulo School of Medicine, Av. Dr. Enéas de Carvalho Aguiar, 44 Bloco 2 9º andar- 06504-000 - São Paulo, SP-Brazil; 3) Institute for Investigation in Immunology (iii), INCT, São Paulo SP, Brazil; 4) Laboratory of Immunology, Universidade Federal do Triângulo Mineiro, 40 Frei Paulino, 48036-180 Uberaba, MG-Brazil; 5) Instituto de Cardiologia Dante Pazzanese (São Paulo) Avenida Dr. Dante Pazzanese, 600 Vila Mariana - 05013-909 - São Paulo, SP-Brazil; 6) School of Medicine of Ribeirão Preto (FMRP), University of São Paulo, Av. Bandeirantes, 4900 - Monte Alegre 15059-900 - Ribeirão Preto, SP -Brazil; 7) Division of Clinical Immunology and Allergy, University of São Paulo School of Medicine, 06504-000 São Paulo, SP -Brazil.

Chagas disease, due to the protozoan Trypanosoma cruzi, is endemic in poor, rural areas of Latin America, but hundreds of thousands of patients now live in the USA. An estimated 300,000 new cases and 50,000 fatilities occur per year. Chronic Chagas disease cardiomyopathy (CCC) is an inflammatory cardiomyopathy that affects approximately 30% of infected individuals decades after infection. Familial aggregation of CCC suggests that there might be a genetic component to disease susceptibility. The outcome of Chagas disease is ultimately defined in the patients' hearts, a consequence of local inflammation and myocardial tissue response. We hypothesize that genetic polymorphisms control expression of several key genes in pathogenetically relevant pathways and networks in the myocardial tissue of CCC patients. The corollary is that it may be possible to establish a host genetic signature with prognostic value based on such polymorphic genes. Access to the largest Chagas disease study population assembled to date, as well as to a cryopreserved heart tissue collection allows us to set up a systems biology approach on Chagas disease, going from exhaustive analysis of molecular changes in hearts to study of genetic variants in disease groups, that couldn't be performed anywhere else. We used a multi-disciplinary approach to screen for genes whose differential expression in CCC hearts could be controlled by genetic polymorphisms. On heart tissue samples we performed a whole transcriptome analysis, microRNA profiling, and whole genome methylation analysis on the same samples. For each differentially expressed gene, we assessed whether it was associated to differential methylation or its expression pattern was under the control of concordantly altered microRNAs. Statistical analyses were conducted to merge proteomic, gene expression and methylation data in order to identify candidate genes for genetic studies. Differentially expressed genes that show no evidence of epigenetic control will have genetic variants in or around it studied using two approaches i) a case-control study based on common SNPs ii) characterization of nuclear families with multicases by exome sequencing to identify rare functional variants. We believe that this approach may allow establishing a host genetic signature with prognostic value, while revealing novel insights into pathogenesis and identifying novel therapeutic targets. This combined approach may be of use in other complex diseases.

2200T

Identification of susceptibility modules for coronary artery disease using a genome wide integrated network analysis.
C. Dong, X. Luo, S. Duan. Zhejiang Provincial Key Laboratory of Pathophysiology, School of Medicine, Ningbo University, Ningbo, Zhejiang 315211, China.

Although recent genome-wide association studies (GWAS) have identified a handful of variants with best significance for coronary artery disease (CAD), it remains a challenge to summarize the underlying biological information from the abundant genotyping data. Here, we propose an integrated network analysis that effectively combines GWAS genotyping dataset, protein-protein interaction (PPI) database, literature and pathway annotation information. This three-step approach was illustrated for a comprehensive network analysis of CAD as the followings. First, a network was constructed from PPI database and CAD seed genes mined from the available literatures. Then, susceptibility network modules were captured from the results of gene-based association tests. Finally, susceptibility modules were annotated with potential mechanisms for CAD via the KEGG pathway database. Our network analysis identified four susceptibility modules for CAD including a complex module consisted of 15 functional interconnected sub-modules, AGPAT3-AGPAT4-PPAP2B module, ITGA11-ITGB1 module and EMCN-SELL module. MAPK10 and COL4A2 among the top-scored focal adhesion pathway related module were the most significant genes (MAPK10: OR = 32.5, $P = 3.5\times10\text{-}11$; COL4A2: OR = 2.7, $P = 2.8\times10\text{-}10$). The significance of the two genes were further validated by other two gene-based association tests (MAPK10: P = 0.009 and 0.007; COL4A2: P =0.001 and 0.023). The susceptibility modules identified in our study might provide novel clues for the clarification of CAD pathogenesis in the future.

2201F

Systematic phenotype prediction in zebrafish identifies novel, disease-relevant cardiovascular gene functions.
G. Musso[1,2], M. Tasan[3,4], C. Mosimann[5,6,7], J.E. Beaver[3], E. Plovie[2], L.A. Carr[5,6,7], H.N. Chua[4], J. Dunham[4], K. Zuberi[4], Q. Morris[4], L. Zon[5,6], F.P. Roth[3,4,8], C.A. MacRae[1,2]. 1) Department of Medicine, Harvard Medical School, Boston, MA; 2) Cardiovascular Division, Brigham and Women's Hospital, Boston, MA; 3) Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA; 4) Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON; 5) Howard Hughes Medical Institute, Boston, MA; 6) Stem Cell Program, Children's Hospital Boston, Boston, MA; 7) Division of Hematology/Oncology, Children's Hospital Boston, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA; 8) Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto, ON.

Comprehensive functional annotation of vertebrate genomes is a central step in realizing the full potential of genome-scale technologies for fundamental and translational discovery. However, the identification of novel gene functions through forward genetic screening quickly becomes overwhelming not only due to the number of surveyable genes, but also due to the increasing number of observable phenotypes. Unbiased prediction of gene-phenotype relationships offers a strategy to direct finite experimental resources towards likely phenotypes, thus maximizing de novo discovery of gene functions. Here we prioritized genes for phenotypic assay in embryonic zebrafish through genomic data integration and machine learning, predicting the effect of loss of function of each of 15,106 zebrafish genes on 338 distinct anatomical processes. Cross-validation suggested the resulting predictions to be particularly precise for certain phenotypic categories, including cardiac phenotypes. In proof-of-concept studies we validated 16 high-confidence cardiac predictions using targeted morpholino knockdown and manual blinded phenotyping in embryonic zebrafish, confirming a significant enrichment for cardiac phenotypes as compared to morpholino controls. Subsequent detailed analyses of cardiac function confirmed these results, identifying novel physiological defects for 11 tested genes, the majority of which are conserved in humans. Among these we identified tmem88a, a recently described attenuator of Wnt signaling, as a discrete regulator of the patterning of intercellular coupling in the zebrafish cardiac epithelium. These results suggest that large-scale gene prioritization in zebrafish enriches for phenotype association, thus substantially reducing survey effort in identifying novel, disease-relevant vertebrate gene functions.

2202W

Large-scale transcriptome profiling in peripheral blood mononuclear cells (PBMCs) of early-onset myocardial infarction individuals and matched controls.
C. Müller[1], A. Schillert[2], M.O. Scheinhardt[2], S. Szymczak[2], F. Ojeda[1], C.R. Sinning[1], R. B. Schnabel[1], S. Wilde[1], P.S. Wild[3,4], K.J. Lackner[5], T. Munzel[4], A. Ziegler[2], S. Blankenberg[1], T. Zeller[1]. 1) Clinic for general and interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany; 2) Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein, Campus Lübeck, Lübeck, Germany; 3) Clinical Epidemiology, Center for Thrombosis and Haemostasis, University Medical Center Mainz, Mainz, Germany; 4) Department of Medicine II, University Medical Center Mainz, Mainz, Germany; 5) Institute for Clinical Chemistry and Laboratory Medicine, University Medical Center Mainz, Mainz, Germany.

Despite the substantial progress in diagnosis and therapy, myocardial infarction (MI) remains one of the major causes of mortality worldwide. MI in young adults is a rare phenomenon and the molecular mechanisms of the pathogenesis leading to MI at young age are still unclear. The aims of this study were to identify differentially expressed genes between young, non-acute MI individuals and healthy controls and to highlight underlying biological mechanisms by functional over-representation analyses. In 112 non-acute MI individuals (age at MI <= 50) and 112 age and sex matched healthy controls from the Gutenberg Health Study (GHS) gene expression levels in PBMCs were determined using Affymetrix GeneChip Exon ST1 Arrays. Differential gene expression was computed by applying linear mixed models adjusted for body mass index, hypertension, smoking status, diabetes and LDL/HDL ratio and the first four principal components. Functional analyses of significantly differentially expressed genes (FDR<=0.1) were conducted using Ingenuity Pathway Analysis. Associations between medical therapy and gene expression were identified in PBMCs of GHS controls (n=307). 183 genes were significantly associated with MI. The top four differentially expressed genes were G-protein-coupled receptor 15 (GPR15, up-regulated), ATP-binding cassette transporter A1 (ABCA1, down-regulated), selenoprotein O (SELO, down-regulated) and cardiotrophin 1 (CTF1, up-regulated). We observed a strong up-regulation of GPR15 with smoking behaviour, but the association with MI remained significant after adjustment. It had been shown that the GPR15 locus in leukocytes of smokers is hypomethylated, indicating a higher accessibility of the DNA and subsequent increased expression of GPR15, which is reflected in our data. Moreover, we observed a positive correlation between GPR15 expression and plaque formation. Pathway analysis revealed over-representations of differentially expressed genes within pathways of the renin-angiotensin system and pathways related to the pathogenesis of cardiac hypertrophy, all sharing a common mechanism, suggesting a de-regulation of intracellular Ca2+ levels. In summary, we observed over-representations of differentially expressed genes in key pathways of the cardiovascular system. The top gene GPR15 shows strong associations to smoking and plaques. Ongoing functional analysis will help to understand its role in cardiovascular disease, which could be promoted by smoking.

2203T

Investigating the role of HoxA3 during cardiac development.
V. Sanghez[1], D. Rux[2], S. Chan[2], L. Borges[3], N. Koyano[3], R. Perlingeiro[3], D. Garry[3], M. Kyba[2], M. Iacovino[1]. 1) Pediatric, Medical Genetics, LABiomed Research Institute at Harbor UCLA, Torrance, CA; 2) Department of Pediatrics, Lillehei Heart Institute, University of Minnesota, MN; 3) Department of Medicine, Lillehei Heart Institute, University of Minnesota, MN.

During embryoid body (EB) differentiation mesoderm patterns toward its major derivatives: lateral plate mesoderm to generate blood and vasculature and paraxial mesoderm to generate muscle and cardiac derivatives. Day 4 EB differentiation thus resembles the gastrulating embryos at DPC 7 (Days Post Conception). Previous studies performed by our group showed a role of HoxA3 on the generation of hematopoietic progenitor cells from haemogenic endothelium, the progenitor cell of Hematopoietic Stem Cell (HSC) in vivo. As a result of HoxA3 upregulation the genome transcriptional profile showed a repression of hematopoietic commitment and promotion of endothelial fate. Unexpectedly we found a strong repression of cardiac specific regulators both in endothelial-committed progenitors and total EBs. HoxA3 gene results to be crucial during developmental stage, indeed genetic deletion in mice results in perinatal lethality due to cardiovascular outflow tract defects as well as pulmonary defects and, as an overall phenotype, it recapitulate DiGeorge syndrome. Fate mapping has shown contribution of HoxA3 expressing cells to the outflow tract as well as right atria and ventricle. We demonstrate with functional assays and at molecular level, that upregulation of HoxA3 during mesoderm patterning strongly represses both cardiac commitment and differentiation. We are dissecting the molecular mechanism underlying this repression and we show preliminary evidence that HoxA3-dependent cardiac repression is mediated through regulation of the Notch ligand, Jagged1.

2204F

Joint Association of 31 Mitochondrial Variants with Type 2 Diabetes: The Strong Heart Family Study. *Y. Zhu[1], ET. Lee[2], SA. Cole[3], K. Haack[3], LG. Best[4], BV. Howard[4], J. Zhao[1].* 1) Dept Epidemiology, Tulane Univ Pub Health & Tropical Med, New Orleans, LA; 2) Center for American Indian Health Research, University of Oklahoma Health Sciences Center; 3) Texas Biomedical Research Institute, San Antonia, TX; 4) Missouri Breaks Industries Research Inc, Timber Lake, SD.

Background: Multiple mitochondrial variants have been associated with T2D, but a single SNP usually confers a small risk to disease. A gene-family approach taking into account the joint contribution of multiple variants may have greater power in detecting genetic associations. Objective: To evaluate the cumulative impact of 31 tagging SNPs in 13 mitochondria-related genes in American Indians participating in the Strong Heart Study (SHS). Methods: Thirty-one tagging SNPs in thirteen mitochondrial-related genes were genotyped in 1,221 American Indians residing in South/North Dakota. All subjects were recruited and examined in 1989-1991 by the SHS. We first conducted single SNP analysis by logistic regression, adjusting for age, sex, BMI, socioeconomic status, lifestyle factors, (physical activity, alcohol intake and smoking status), estimated glomerular filtration rate (eGFR) and history of cardiovascular disease or hypertension. Gene-based association was assessed by combining p-values of all SNPs within a gene based on single SNP analysis using a weighted truncated product method. The gene-family association was performed by combining p-values of all 13 genes using the same method. Sensitivity analysis was conducted to evaluate whether the observed associations were primarily driven by the most significant SNPs. Multiple testing was corrected using false discover rate (FDR). Results: Multiple SNPs showed nominal or marginal association with T2D or insulin resistance, but the association diminished after correction for multiple testing. Four genes, including CYTB (p=0.04), COX3 (p=0.04), ND1 (p=0.037) and TRNC (p=0.018), were weakly associated with T2D, but the associations disappeared after correction for multiple testing. However, a gene-family analysis comprising all 31 SNPs from the 13 genes showed significant association with T2D, suggesting that mitochondrial alleles defined by haplotypes of these SNPs contribute to T2D susceptibility. This association was not driven by the most significant SNPs. Conclusion: Although a single SNP confers only a small risk to T2D, multiple variants in the mitochondrial pathway may jointly contribute to T2D, suggesting the importance of modeling the joint impact of multiple variants on complex disease susceptibility. Our findings may provide useful information for risk classification or personalized approach to diabetes management in American Indians.

2205W

RNA sequencing of human adipose tissue identifies allele-specific expression associated with dyslipidemia. *M. Alvarez[1], R.M. Cantor[1], P.M. Reddy[1], Y. Hasin[2,3], D. Weissglas-Volkov[1], E. Nikkola[1], J. Kuusisto[4], A.J. Lusis[1,2,3], M. Laakso[4], P. Pajukanta[1].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA; 3) Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, California; 4) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Cardiovascular disease (CVD) is a common cause of death world-wide. To accelerate early intervention and treatment of CVD, it is important to functionally identify mechanisms that contribute to CVD risk. A large amount of phenotype differences among individuals can be attributed to sequence variants affecting gene expression. We used adipose RNA sequencing (RNAseq) to screen for likely regulatory variants associated with high serum triglyceride (TG) levels, a CVD risk factor. Adipose tissue is highly relevant for TG metabolism, as TGs are stored there, and hydrolysis of TGs occurs predominantly in adipose tissue. We sequenced adipose RNAs of individuals with low and high serum TG levels (n=104) from the Finnish population cohort METabolic Syndrome In Men (METSIM) using 50 bp, paired-end reads and the Illumina HiSeq2000 platform at the depth of 3-4 samples/lane. We aligned and mapped the reads uniquely using STAR, and calculated read counts and called the variants using SAMtools. To investigate whether allele-specific expression (ASE) is a common functional mechanism regulating adipose gene expression in dyslipidemia, we screened variants for ASE in this preliminary sample of 104 by filtering for heterozygous variant sites with >10 mapped reads and >2 reads with the variant allele using a minor allele frequency (MAF) of >10%. We observed 17,884 potential SNPs with ASE. Among the 17,884 SNPs, 8,479 variants with significant ASE were seen in the total sample, while 4,299 variants were seen in subjects with high TG and 4,350 in subjects with low TG, respectively, using a paired t-test with a Bonferroni correction. In the high TG group, of the 4,299 variants, 1,324 were unique and in the low TG group, of the 4,350 variants, 1,273 were unique. None of the variants differed in the percent of ASE between individuals with high and low TG using the Wilcoxon rank sum test, a Bonferroni correction, and MAF>20%, suggesting that for common variants, dyslipidemia-specific differences in ASE percents are rare. The David pathway tool identified 11 significant functional categories enriched in the high TG group and 12 functional categories enriched in the low TG group that passed a Bonferroni correction, indicating that ASE variants may accumulate in molecular pathways based on the dyslipidemia status. To summarize, these novel data suggest that ASE is a common mechanism underlying differential gene expression between dyslipidemic cases and controls in adipose tissue.

2206T

Genes related to *CRKL* and mouse heart development may act as genetic modifiers to congenital heart disease in human 22q11.2 deletion syndrome. *J. Chung[1], S.E. Racedo[1], T. Guo[1], D.M. McDonald-McGinn[2,3], E. Zackai[2,3], R.J. Shprintzen[4], B.S. Emanuel[2,3], B. Funke[5,6], B.E. Morrow[1].* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) The Virtual Center for Velo-Cardio-Facial Syndrome, Inc., Manlius, NY; 5) Laboratory for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, MA; 6) Department of Pathology, Massachusetts General Hospital, Boston, MA.

Congenital heart disease (CHD) occurs at a rate of 7-9:1000 live births and 80% of CHD cases have no clear etiology. Approximately 70% of patients with 22q11.2 deletion syndrome (22q11DS, also known as DiGeorge/VCFS, OMIM: 192430; 188400) have a CHD, mostly of the conotruncal type, with highly variable expressivity suggesting the presence of genetic modifiers. The majority (~85%) of patients with 22q11DS have a 3 Mb deletion encompassing ~60 genes including the T-box transcription factor, *TBX1* (OMIM: 602054) and the adaptor protein, *CRKL* (OMIM: 602007), both implicated in the etiology of CHD in the disorder. There have been extensive studies regarding CHD on *TBX1*, but few have been performed on *CRKL*, which links various signaling pathways to downstream effectors. To elucidate the independent role of *CRKL* in human CHD, we analyzed echocardiograms from 15 subjects with 22q11DS encompassing *CRKL* but not *TBX1*. We found 7/15 have cardiac outflow tract alignment or septal defects, which are strikingly similar to those seen in mutant mice harboring an allelic series of *Crkl* dosage, suggesting that *CRKL* is important in its own right, and that it may serve as a genetic modifier of CHD in subjects with the typical 3 Mb deletion.

To follow up on *CRKL* as a candidate modifier gene of 22q11DS and identify additional modifiers related to *CRKL*, we performed whole exome sequencing (WES) on 176 subjects with the typical 3 Mb deletion; 74 subjects have tetralogy of Fallot and 102 have a normal heart/aortic arch. We performed a candidate gene association test, using SKAT (PMID: 23684009), on 141 genes either in *CRKL* pathways or with a known role in the development of structures related to phenotypes in *Crkl* mutant mice. Of the 141 genes, 103 have nonsynonymous single nucleotide variants (SNVs). Six genes (*NFATC4*, *HBEGF*, *CBL*, *HSPG2*, *PXN*, and *ELMOD2*) have a nominal p-value < 0.05 and empirical p-value < 0.1 after 10,000 permutations. Our top gene is *NFATC4* (p-value = 0.0073, FDR q-value = 0.2509, OMIM: 602699). Importantly, mice with mutations in *Nfatc4* have similar heart malformations as our *Crkl* mutant mice indicating that nonsynonymous variants in *NFATC4* may genetically interact with *CRKL* to modify the CHD phenotype seen in 22q11DS. We will test for genetic interactions between *Crkl* and *Nfatc4* using mouse models and perform replication studies of our findings in an independent cohort of 22q11DS subjects.

2207F

Genetic Influence of Scavenger Receptor Class B Type 1 (*SCARB1*) on Plasma Lipid Traits in non-Hispanic White Americans. *V. Niemsiri[1], X. Wang[1], M.M. Barmada[1], J.E. Hokanson[2], R.F. Hamman[2], F.Y. Demirci[1], M.I. Kamboh[1].* 1) Human Genetics, Graduate School of Public Health , University of Pittsburgh, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO.

Coronary heart disease (CHD) is a major public health burden worldwide. A main risk factor for CHD includes low levels of high-density lipoprotein cholesterol (HDL-C). *SCARB1* is one of the candidate genes involved in HDL-C metabolism. To identify genetic influences of both common and rare variants of *SCARB1* on plasma lipid traits, we sequenced all 13 exons, exon-intron boundaries plus 1 kb of 5' and 3' flanking regions of *SCARB1* in 95 non-Hispanic White (NHW) US individuals with the upper (n = 47) and lower (n = 48) 10th percentile of HDL-C/triglyceride (TG) distribution. Sequencing analysis revealed 44 variants (MAF ≥5%, n = 12; MAF <5%, n = 32). No statistically significant difference was observed in the distribution of rare variants between the two extreme HDL-C/TG groups (P = 0.615). We genotyped 40 relevant uncommon variants and common tag SNPs identified by our sequencing (MAF ≥5%, n = 8; MAF <5%, n = 32), plus 32 additional tag SNPs from HapMap covering the entire gene and 4 reported variants in our total NHW sample (n = 623). Of the 76 genotyped variants, 69 (MAF ≥5%, n = 39; MAF <5%, n = 30) passed QC and were further evaluated for their association with lipid traits. Based on the gene-based test, suggestive association was observed with apolipoprotein B (apoB) (P = 0.042). In single-site analysis of 39 variants with MAF ≥5%, three (rs11057844, rs701106, rs838880) and four (rs2343394, rs4765615, rs2278986, rs11057820) variants were nominally associated with HDL-C and apoB, respectively (P <0.05). Rare variant and haplotype analyses are currently underway. Our findings indicate a relatively small influence of SCARB1 genetic variation on plasma lipid and apolipoprotein levels.

2208W

Study on the association between *APOH* variants and thrombophilia. *L. Tang, X. Lu, Q. Wang, J. Yu, X. Jian, T. Guo, H. Wang, H. Liu, Y. Hu.* Institute of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.

Plasma beta 2-glycoprotein I (beta 2-GPI) is the most common autoantigen for the production of antiphospholipid antibodies. It have a variety of anticoagulant properties. Beta 2-GPI inhibits the factor XI activation in the intrinsic pathway of blood coagulation, von Willebrand factor-dependent platelet aggregation, platelet-dependent prothrombinase activity, and the procoagulant activity of thrombin. Recent studies suggest that the formation of anti-beta 2-GPI antibodies is associated with the protein structure abnormality of beta 2-GPI and that a conformational change in beta 2-GPI is sufficient to induce autoantibodies, which can confer high risk for both arterial and venous thrombosis. Therefore, we investigated *APOH* encoding beta 2-GPI as a candidate gene to identify possible variants that could contribute to thrombosis susceptibility. The putative promoter, 8 exons, and 3'-untranslated region of *APOH* were screened using a PCR-sequencing method in 60 individuals who were diagnosed with unprovoked venous thrombosis. The haploid genotypes were subsequently determined in the case-control populations (1304 versus 1334) from the MAGIC study, and the odds ratios for venous thrombosis risk were calculated by logistic regression analysis. Four common polymorphisms g.5028C>A, c.422T>C (p.Ile141Thr), c.461G>A (p.Arg154His), and c.1004G>C (p.Trp335Ser) were identified by resequencing. These variants were in complete linkage disequilibrium with one another and produced 3 haploid genotypes: wild type H1, CTGG/CTGG; heterozygous type H2, CTGG/ACAC; homozygous type H3, ACAC/ACAC. The case-control study showed that the H2 type was present in 157 venous thrombosis individuals and 135 controls; while the H3 type was identified in 12 of the thrombosis group and 2 of the control group. Thus, carriers of the H2 type and the H3 type had a 1.29-fold and 6.04-fold increased risk of venous thrombosis, respectively. The common *APOH* haploid genotypes H2 and H3 were associated with an increased risk of venous thrombosis and were common genetic risk factors for thrombophilia in the Chinese population.

2209T

APOE p.Leu167del mutation in Familial Hypercholesterolemia. *Z. Awan[1], H. Choi[1], N. Stitziel[3], I. Ruel[1], R. Husa[1], MH. Gagnon[1], RH. Wang[1], N. Seidah[2], S. Kathiresan[4], J. Genest[1].* 1) The Research Institute of the McGill University Health Centre, RI-MUHC, Montréal, QC; 2) Institut de Recherches Cliniques affiliated to Université de Montréal, IRCM, Montréal, QC; 3) Department of Medicine, Division of Cardiovascular and Statistical Genomics, Washington University School of Medicine, St. Louis, MO; 4) Broad Institute, Harvard Medical School, Cambridge, MA.

PURPOSE: Autosomal dominant hypercholesterolemia (ADH) is caused by mutations in the low density lipoprotein receptor (LDLR), its ligand apoB (APOB) or proprotein convertase subtilisin/kexin type 9 (PCSK9) genes. Yet DNA sequencing does not identify mutations in these genes in a significant number of cases with a phenotype of ADH, suggesting that ADH has multiple genetic etiologies. METHODS: Through a combination of clinical examination, biochemical analysis, candidate gene approach and next-generation exome sequencing we investigated the genetic basis of an ADH phenotype in a proband of an Italian origin. RESULTS: we identified an in-frame three base-pair deletion in apolipoprotein E (APOE, Chromosome 19:45412053-55) resulting in a Leu167del mutation. The proband presented with an acute myocardial infarction at age 43, requiring urgent coronary revascularization. He had extensive tendinous xanthomas, xanthelasmas and elevated levels of total cholesterol 11.2 mmol/L, LDL-C 9.69 mmol/L, normal HDL-C 1.62 mmol/L and triglycerides levels 1.13 mmol/L. HPLC lipoprotein profile showed selective increase in LDL-C. DNA sequencing did not identify any mutation in the LDLR, PCSK9, LDLR adapter protein-1 (LDLRAP1) and exon 26 of the APOB gene. We then performed exome sequencing on three individuals from the family. Using data derived from exome chip genotypes for association with LDL-C, the strongest evidence of association was found for the identified APOE Leu167del mutation. CONCLUSIONS: The Leu167del mutation is predicted to alter the protein structure of ApoE near the α-helix within the receptor binding domain. This report confirms that ADH can be caused by mutations within the APOE gene and represents the 4th loci causing ADH.

2210F

Role of Titin gene variants in human dilated cardiomyopathy.
R.L. Begay[1], S.L. Graw[1], G. Sinagra[2], M. Merlo[2], D. Slavov[1], G. Barbati[2], A. Di Lenarda[2], X. Zhu[1], L. Mestroni[1], M. Taylor[1]. 1) Department of Medicine, University of Colorado - CU Cardiovascular Institute, Aurora, CO; 2) Cardiovascular Department, University of Trieste, Italy.

Background: The Titin gene (TTN) is the second largest human gene and encodes the largest known human protein, titin, that plays a central role in sarcomere organization. Mutations in TTN have been difficult to study due to the large size of the gene, however recent data have shown that TTN truncation mutations lead to a dilated cardiomyopathy phenotype. The large size of TTN means that almost all individuals carry missense TTN variants, the majority of which are presumably benign yet raise challenges in the evaluation of TTN in patients with overt dilated cardiomyopathy. Methods: Our group sequenced 313 exons covering the N2B and N2BA cardiac isoforms of TTN in 135 subjects with dilated cardiomyopathy. Bioinformatic filtering for 'severe' missense variants was based on SIFT score, GERP, Polyphen2 HDVAR, low frequencies in the 1000 genome project, and Exome Sequencing Project. Results and Conclusions: Ultimately, 45 TTN 'severe' missense variants were noted in 38 probands. Segregation of 'severe' TTN missense variants was demonstrated in larger families. Seven subjects were compound TTN heterozygotes, and six subjects were double heterozygotes with a 'severe' TTN variant and a pathogenic variant in another dilated cardiomyopathy gene. Genotype-Phenotype analyses between TTN 'severe' variant carriers and non-carriers was also performed which showed a trend towards better survival in TTN 'severe' missense carriers (p=0.071). These data provide important insight for researchers and laboratory into the prevalence and phenotypic consequences of 'severe' TTN variants in dilated cardiomyopathy patients.

2211W

Desmosomal and titin gene variants in arrhythmogenic right ventricular cardiomyopathy: genotype-phenotype correlations.
F. Brun[1,2], C. Barnes[1], G. Sinagra[2], D. Slavov[1], G. Barbati[2], X. Zhu[1], R. Begay[1], S. Graw[1], B. Pinamonti[2], E. Salcedo[1], M. Taylor[1], L. Mestroni[1], Familial Cardiomyopathy Registry. 1) Cardiovascular Institute, University of Colorado, Aurora, CO; 2) Department of Cardiology, Hospital and University of Trieste, Trieste, Italy.

Background Arrhythmogenic right ventricular cardiomyopathy (ARVC) is caused by abnormalities in desmosomal proteins of the intercalated disc and desmosomal gene mutations are the principal cause of ARVC. Recently novel variants were discovered in the sarcomeric gene titin (*TTN*) that are associated with ARVC. Whether known mutation carriers differ clinically from non-carriers is not well known. To address this question, we analyzed clinical outcomes in our ARVC population based on mutation status.

Methods Thirty-eight ARVC families (66 patients) were analyzed, with a median follow-up of 77 months. Genotype-phenotype association analysis was performed, and multiple variables including symptoms, electrocardiogram/echocardiogram abnormalities, arrhythmias, pacemaker and/or ICD implantation and survival time free from death or heart transplant were compared between desmosomal mutation carriers, *TTN* carriers and non-carriers.

Results Seven patients (11%) harbored rare genetic variants in desmosomal genes (*DSP, PKP2, DSG2*, and *DSC2*), 14 (21%) carried *TTN* variants and 45 (68%) were non-carriers. Desmosomal carriers (DC) had a higher prevalence of inverted T waves in V2-3 in the absence of RBBB (100% vs. 32%, p=0.001) and epsilon waves (57% vs. 14%, p=0.007) compared to non-carriers. The *TTN* group had significantly more supraventricular arrhythmias (atrial fibrillation, atrial tachycardia) (43% vs. 0%, p=0.04) and required more pacemakers (57% vs. 0%, p=0.018). Conversely, DC required more heart transplants relative to non-carriers (57% vs. 11%, p=0.03) and exhibited a worse survival free from death or heart transplant (63% vs. 88% at 30 years and 42% vs. 88% at 50 years, p<0.001).

Conclusions This study provides valuable insights into the clinical consequences of gene mutations in individuals with ARVC. *TTN* rare variants confer greater risk for supraventricular arrhythmias and the need for pacemaker implantation relative to DC, while DC portends a greater risk for electrocardiogram abnormalities and the combined end-point of heart transplant or death compared to non-carriers.

2212T

Burden of rare variants in *PON1* is associated with ischemic stroke.
D.S. Kim[1,2], D.R. Crosslin[1,2], P.L. Auer[3,4], A.A. Burt[1], A.S. Gordon[2], C.E. Furlong[1,2], J.F. Meschia[5], M. Nalls[6], U. Peters[3], S.S. Rich[7], D.A. Nickerson[2], G.P. Jarvik[1,2], NHLBI Exome Sequencing Project (ESP). 1) Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, WA; 2) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 4) Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA; 5) Department of Neurology, Mayo Clinic, Jacksonville, FL; 6) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD; 7) Center for Public Health Genomics, University of Virginia, Charlottesville, VA.

Background: HDL-associated paraoxonase 1 (encoded by *PON1*) is an enzyme with broad substrate specificity and whose activity has been associated with numerous human diseases, including ischemic stroke. Common variants in *PON1* have not consistently been associated with stroke. Rare coding variation that is predicted to alter *PON1* enzyme function may be more strongly associated with ischemic stroke. **Methods and Results:** The NHLBI Exome Sequencing Project (ESP) sequenced the coding regions of the genome (exome) in 6,503 participants ascertained for heart, lung and blood phenotypes. In this sample of 4,204 participants, 496 had verified ischemic stroke. A total of 29 nonsynonymous variants were identified in *PON1*. Sequence kernel association testing (SKAT) analysis adjusting for covariates identified evidence for association ($p=1.29\times10^{-3}$; permutation $p=3.01\times10^{-3}$) with stroke. The association of *PON1* with ischemic stroke was stronger in participants with African ancestry (AA $p=5.73\times10^{-4}$; permutation $p=5.03\times10^{-3}$). Ethnic differences in the association between *PON1* with ischemic stroke could be due, in part, to the effects of $PON1_{Val109Ile}$ missense variant (overall $p=7.88\times10^{-3}$; AA $p=6.52\times10^{-4}$), found at higher frequency in AA participants (1.16%) than to European ancestry (0.04%) participants. **Conclusions:** Rare nonsynonymous exomic variation in *PON1* was associated with ischemic stroke, with stronger associations identified in those of African ancestry. Given the high morbidity and mortality of stroke and the higher prevalence of stroke in individuals with African Ancestry, further studies should replicate these findings and also functionally validate the effects of the rare nonsynonymous variants on PON1 enzyme function.

2213F

The spectrum and prevalence of genetic background noise in patients with arrhythmogenic (right ventricular) cardiomyopathy.
T.T. Koopmann[1], S. Walker[2], G. Kaur[2], S.W. Scherer[2,3], C.R. Marshall[2,3], R.M. Hamilton[4]. 1) Physiology and Experimental Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 2) The Centre for Applied Genomics and Genetics and Genome Biology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 3) McLaughlin Centre and Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 4) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.

Introduction: Arrhythmogenic (right ventricular) cardiomyopathy (ACM, previously known as ARVC) is histopathologically characterized by progressive fibrofatty replacement of myocardium, primarily of the right ventricle. Mutations associated with the disorder have been identified in genes mostly encoding proteins that form desmosomes; however, in about 50% of patients no mutation is found in the known ACM associated genes and the pathogenicity of many identified variants is unclear. To determine the spectrum and prevalence of 'genetic background noise', we sequenced the exomes of 19 ACM cases and 35 controls and performed mutation burden analysis. **Methods:** In this study, we performed whole exome sequencing of 19 unrelated ACM patients from European descent and 35 unrelated ethnicity matched control individuals and focused on the contribution of genetic variation in 52 genes previously associated with ACM, or interaction partners of these genes, to ACM susceptibility. **Results:** The mean number of variants identified in the 52 candidate genes in ACM patients was 53.8 per individual; controls had an average of 41.7 variants (*P* = NS). Mutation burden (defined as the average number of mutations per person) analysis of the identified variants revealed that non-synonymous rare variants (MAF<1%) in the 52 genes were not significantly more present in patients compared to controls (mean = 2.47 in cases versus 2.34 in controls). Non-synonymous rare variants in the previously ACM associated genes (PKP2, DSG2, DSP, DSC2, JUP, RYR2, PLN, TMEM43, TTN) also showed no significant difference in mutation burden (mean = 1.84 in cases versus 1.54 in controls). Interestingly, the overall yield of variants in the ACM genes was 73.7% in cases versus 80% in controls. Of these, 71.4% of patients and 57.1% of controls had >1 mutation. **Discussion:** This study is an evaluation of genetic variation in ACM patients versus healthy controls for the ACM susceptibility and candidate genes. Further research will be required to elucidate the role of the identified variants in ACM.

2214W

Copy number variants in patients with abdominal aortic aneurysms.
D. Majoor-Krakauer[1], K.M van de Luijtgaarden[2], H.J.M Verhagen[2], G.C.M Huijbregts[1], L. Koster[1], A.M Bertolli Avella[1], H.B Beverloo[1], L.J.C.M van Zutven[1]. 1) Dept. of Clinical Genetics, Erasmus University MC, Rotterdam, Netherlands; 2) Dept. of Vascular Surgery, Erasmus University MC, Rotterdam, The Netherlands.

Abdominal aortic aneurysms (AAA) are a major cause of morbidity and mortality in the elderly and are characterized by extracellular matrix (ECM) degeneration, dysregulation of vascular smooth muscle (SMC) cells, inflammation and atherosclerosis. In approximately 20% of cases abdominal aneurysm are familial (fAAA). No genes involved in fAAA have been identified so far. The aim of the current study was to investigate if copy number copy number variants (CNV) may add in discovering pathofysiologic pathways and genes involved in AAA formation. Methods: AAA patients referred for genetic counseling in 2011 and 2012 were included. Patients were classified as familial when at least one first-degree was reported to have an aortic aneurysm. Patients without a positive family history were classified as sporadic (spAAA). Microarray was performed in all index patients (i.e. first relative diagnosed with AAA). Microarray was performed by using the Illumina HumanCytoSNP-12v2.1 and analyzed using Nexus CopyNumber Discovery v6.1 software. Results were evaluated using the UCSC Genome Browser March 2006, the Database of Genomic Variants and an in-house database of controls. Rare CNV regions larger than 5Mb were classified as variants and CNV regions larger than 3Mb that have been observed previously in control populations at least three times were classified as polymorphisms. GeneNetworks and Ingenuity pathway (IPA) analysis was performed of the genes identified in CNV variants. Results: The study population consisted of 67 index AAA patients; 54 were classified as fAAA and 13 as spAAA. We found 12 rare CNV variants in 11 (20%) of the fAAA patients. In patients with spAAA no rare CNV were found. Tissue expression, biological process and KEGG of the genes identified in rare CNV in fAAA was performed showing frequent involvement in inflammation, glucose metabolism, smooth muscle cell or cartilage functioning. Expression of a number of genes in cicatrisation may indicate links with collagen pathways. Conclusion: These findings together with the IPA showing that a number of CNV genes are involved in the TGF-β pathway, may indicate that in familial AAA similar mechanisms may be activated as in the known genetic TGF-β, collagen, or smooth muscle cell related genetic aneurysm syndromes.

2215T

Whole-Genome Sequencing of Families with Early-Onset Myocardial Infarction.
A. Mehta[1], N. Schnetz-Boutaud2[2], D.J. Van Booven[1], L. Wang[1], Z. Liu[1], N. Vasudeva[1], M.A. Pericak-Vance1[1], J.L. Haines[2], G.W. Beecham[1]. 1) University of Miami, Institute for Human Genomics; 2) Vanderbilt University, Center for Human Genetics Research.

Coronary heart disease is the leading cause of death in the United States, with myocardial infarction (MI) being a primary clinical presentation. While genome-wide association studies of common variation have revealed a number of risk loci for MI, most of the loci have very small risk effects (odds ratios between 1.06-1.20) and they collectively explain less than 10% of the heritability of MI. To investigate the role of rarer variation in MI, we are performing whole-genome sequencing on 28 affected individuals with MI across 6 different families. Each family has at four or more individuals affected by MI, with at least one affected individual with onset under age 55. Whole-genome sequencing of these individuals will be performed using the Illumina Hiseq 2000 sequencer. Each individual will be sequenced using four lanes of the flowcell to achieve ~40x depth of coverage across the genome. The alignment will be conducted using the GATK pipeline, and variant and indel calling will be done using the Unified Genotype caller from the GATK pipeline. The variants and indels will be annotated using SeattleSeq for common annotations and Annovar for custom annotations (e.g., cardiovascular related loci, ENCODE elements, and enhancers, etc). Together, these tools provide comprehensive annotations, like functional group, conservation scores and functional prediction scores (Sift and polyphen2), frequency data from the 1,000 Genomes Project and the NHLBI Exome Sequencing Project. Additional family members that do not have WGS data will be analyzed using a genome-wide genotyping platform (i.e., exome chip) to establish phase and haplotype sharing within families, using the MERLIN software. Shared variants will be filtered based on annotation to identify those most likely to contribute to CHD. The most high-priority variants will be assessed in an independent dataset.

2216F

Exome sequencing identifies a novel candidate gene, NRG1, for serum cholesterol levels in Mexicans.
E. Nikkola[1], M. Alvarez[1], M.V.P. Linga Reddy[1], A. Ko[1], D. Weissglas-Volkov[1], C. Gutierrez-Cirlos[2], L. Riba[2,3], M.L. Ordoñez Sánchez[2], Y. Segura Kato[2], T. Tusie-Luna[2,3], C. Aguilar-Salinas[2], P. Pajukanta[1]. 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles; 2) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 3) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

Background: It is important to identify and functionally characterize genes decreasing serum low-density lipoprotein cholesterol (LDL-C) levels to better prevent and treat hypercholesterolemia, one of the most important risk factors of coronary heart disease. The Mexican population is especially prone to hypercholesterolemia with 43% of Mexicans exhibiting total cholesterol levels >200mg/dl in a recent national survey. Familial hypobetalipoproteinemia (FHBL) is an autosomal dominant disorder characterized by low levels of plasma LDL-C, total cholesterol (TC), and apolipoprotein B (apoB). FHBL has previously been shown to be caused by mutations in the APOB, PCSK9, and ANGPTL3 genes, and loci on chromosomes 10 and 3p21 have also been linked to FHBL. However, in most FHBL cases the underlying cause is unknown. To identify genes involved in FHBL and cholesterol metabolism in Mexicans, we explored two Mexican FHBL families without known mutations in the previously identified genes. The probands of the families had a similar kinetic profile with an increased apoB-LDL and apoB-VLDL catabolic rate. **Methods and Results:** The affected and unaffected individuals from the Mexican FHBL families were exome sequenced by capturing 62Mb coding and regulatory regions. We focused on the variants shared by all affected family members and not present in the family members with normal apoB levels. We also filtered the variants based on their type, frequency, gene expression, and functional predictions using SIFT and PolyPhen. In family 1, filtering for novel missense variants and variants with minor allele frequency (MAF)<5% as well as applying the filters for gene expression and functional predictions, reduced the variant pool to 10 nonsynonymous variants. In family 2, the same filtering resulted in 2 nonsynonymous variants. To further investigate whether any of these 12 genes regulates cholesterol levels generally in Mexicans, we tested their common variants for association with TC in 3,700 Mexican individuals. Variants in one gene, neuregulin 1 (NRG1), passed the Bonferroni correction for association with TC, with rs1383966 providing the strongest signal. **Conclusions:** Exome sequencing of FHBL families and a follow up in 3,700 individuals identified the NRG1 gene for serum cholesterol levels in Mexicans. Replication in additional Mexican cohorts and functional studies are warranted to further investigate the role of NRG1 in cholesterol metabolism.

2217W

Clinical presentation and functional characterization of a c.532+1G>A splice site mutation in *CASQ2* in a patient with catecholaminergic polymorphic ventricular tachycardia.
K. Patel[1], M. Delio[1], T.V. McDonald[2], C. Montagna[1], B. Morrow[1]. 1) Department of Genetics, Albert Einstein College of Medicine, New York, NY; 2) Department of Medicine (Cardiology), Albert Einstein College of Medicine; Montefiore Medical Center, Bronx, NY.

Catecholaminergic polymorphic ventricular tachycardia (CPVT) (Ventricular tachycardia, stress-induced polymorphic; VTSIP; OMIM#604772) is an inherited disorder characterized by episodic syncope triggered through physical or emotional stress. Morphologically, the heart is normal and patients usually present with mild symptoms such as dizziness. However, if left untreated, CPVT can be highly lethal causing sudden death from ventricular tachycardia that may degenerate into ventricular fibrillation. Approximately 70% of CPVT cases are caused by dominant mutations in the Ryanodine receptor 2 (*RYR2*) gene and a further 7% are accounted for by recessive mutations in Calsequestrin-2 (*CASQ2*). Here, we report a patient who first presented with episodic palpitations and syncope related to exercise at 11 years of age. Due to intermittent QT interval prolongation, she was given the provisional diagnosis of Long QT syndrome or CPVT. None of her family members had displayed cardiovascular symptoms. We have recently identified this patient to be compound heterozygous for mutations in the *CASQ2* gene through a research based next generation sequencing initiative at our institute. We designed a custom, multi-disease target capture panel, covering 650 disease causing genes and revealed that this patient has a novel c.199C>T (p.Gln67*) mutation and a previously reported splice site mutation c.532+1G>A. No other mutations were detected in any genes linked to Long QT syndrome. Although the c.532+1G>A mutation has not yet been demonstrated to alter mRNA splicing, the mutation has been observed in a previously reported consanguineous CPVT family (PMID: 12386154) and is therefore considered most likely to be pathogenic. We are currently exploring a protein-based immuno-assay to detect aberrant *CASQ2* products, which may be formed as a result of this splice site mutation. Therefore by using our research based target panel we have been able to confirm the clinical diagnosis of CPVT in this patient. *In vitro* functional investigations will help us understand the pathogenic mechanism, which leads to the clinical presentation seen in our patient. As a result of this work, additional at-risk family members can now be offered cascade testing.

2218T

Survival is determined by mutation type and molecular mechanism in Ehlers Danlos Syndrome (EDS) type IV. *M.G. Pepin[1], U. Schwarze[1], D. Leistritz[1], M. Liu[2], P.H. Byers[1].* 1) Dept Pathology, Univ Washington, Seattle, WA; 2) Dept Biostatistics, Univ Washington, Seattle, WA.

EDS type IV, the vascular type, is caused by dominant mutations in COL3A1. Affected individuals experience life-threatening arterial dissection and rupture, bowel perforation, and uterine rupture. We present a summary of natural history and genotype-phenotype relationships for 1232 individuals (632 index cases and 600 relatives) with EDS type IV. Median survival was 51 years. Arterial rupture and dissection were the leading cause of first complications (70%) and the leading cause of death (90%). Survival differed significantly by type of mutation (haploinsufficiency, substitutions for glycine residues in the triple helix, and splice site mutations) and even the mildest genotype (null mutations) had lower survival than the US normal. Hazard ratios for donor site mutations (8.8) and any substitution for a triple helical glycine (4.6) compared to null mutations were significant. Among substitutions for glycine in the triple helical domain, substitutions that introduced serine had a 2.6 fold higher hazard than null mutations, while for substitutions by valine the hazard was 7.2 fold higher. The location of substitutions for glycine residues in the triple helical domain was not a significant factor in survival. The median survival for females was significantly greater (54 yrs) than for males (46 yrs). Pregnancy did not alter overall survival risk (Murray et al in press). Aortic rupture in young males was one factor that contributed to the poorer survival in males. Both intracranial aneurysm and coronary artery dissection presenting as "heart attack", particularly in women in the fourth decade, were more prevalent among those affected than previously appreciated. Emergency surgical repairs of arterial rupture or dissection were often unsuccessful, while planned surgery of an identified aneurysm had significantly better outcome. In instances when a death occurred during surgical repair for colon rupture (15% of recorded cases), unexpected vascular rupture was often noted. All data suggests that surveillance of the arterial tree by MRI from the time of diagnosis may be beneficial. The findings provide some guidance for surveillance and indicate that in the creation of clinical trials attention has to be paid to mutation type to balance treatment and control groups. (Supported in part by funds from the Freudmann Fund at the University of Washington).

2219F

GENOMIC ANALYSIS OF LONG QT SYNDROME IN INDIAN COHORT. *S.F. Qureshi[1], C. Narasimhan[2], A. Ali[1], A. Venkateshwari[3], K. Thangaraj[4], P. Nallari[1].* 1) Department of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India; 2) Care Hospital, Hyderabad; 3) Institute of Genetics and Hospital for Genetic Diseases, Osmania University, Hyderabad; 4) Centre for Cellular and Molecular Biology, Hyderabad.

Long QT Syndome, characterized by prolonged QTc along with congenital bilateral deafness is a result of homozygous or genetic compound variations in either of the K+ channel genes, KCNQ1 and KCNE1. Tumor Necrosis factor-alpha, a known inflammatory marker is one of the modifier genes in CVDs. Hence, genomic analysis of KCNQ1, KCNE1 and G-308A SNP of TNF-alpha gene/s was carried out in 32 LQT blood samples and 100 control samples. Molecular analyses were carried out on genomic DNA of 100 controls (50M; 50F), 32 LQTS patients and the available family members followed by In-silico analyses to establish the genotype-phenotype correlations. Genetic compounds of an intronic SNP (A2737320G) and a synonymous (S546S) SNP in KCNQ1; and a neutral substitution (G38S) SNP in KCNE1 were observed apart from genotyping of G-308A SNP of TNF-alpha gene. In KCNQ1 intronic and exonic SNPs (OR-0.36, p=0.014); and TNF-alpha G-308A SNP (OR-0.27, p=0.001), the AG genotype was found to confer protection implying the heterozygote advantage in the genetic compounds. High degree of Linkage disequilibrium was observed (D'=0.9998, 0.9278, 0.7143) indicating a strong synergistic association between the four genotypes with LQTS. In-silico analysis further confirmed mRNA secondary structure variations and altered protein binding during spliceosome processing. Codon bias and usage for KCNQ1 synonymous SNP (S546S) also revealed the higher usage of the rare variant (UCA=0.15) compared to common UCG codon (0.05) confirming the altered regulation of the protein. G38S missense substitution of KCNE1 leads to changes in stem 4 and 5 of mRNA secondary structure wherein finally a polar hydrophilic serine is being replaced by a non-polar, hydrophobic glycine disrupting the structure and function of the transmembrane protein. It is thus predicted that the variations may alter the KCNQ1 and KCNE1 transmembrane protein causing prolonged QTc. Hence, along with the potassium channel genes KCNQ1and KCNE1, TNF-alpha can be considered as a biomarker in understanding the pathophysiological features of poorly understood Long QT Syndrome. The study highlights the role of genetic compounds and their additive gene action.

2220W

Novel sarcomeric gene variations in Hypertrophic cardiomyopathy patients of Indian Cohort. *A. Rangaraju[1], M.L Satyanarayana[1], C. Narasimhan[2], P. Nallari[1].* 1) Department of Genetics, Osmania University, Hyderabad, Andhra Pradesh, India; 2) Cardiologist, CARE Hospitals, Hyderbad, Andhra Pradesh, India.

Hypertrophic Cardiomyopathy (HCM) is a major cause of morbidity and mortality, characterized by hypertrophy of the left/right ventricle and interventricular septum with an autosomal dominant mode of inheritance. It has a prevalence of 1 in 500 and exhibits a variable clinical course ranging from being asymptomatic to severe heart failure and sudden cardiac death. Mutations in the sarcomeric genes that code for myofilaments and their supporting proteins have been identified as a cause of HCM. Despite the identification of 900 different causative mutations, the frequency of mutation rate is very low, alternatively, the effect of modifier genes could influence the degree of hypertrophy developed in response to the causal mutation. MYBPC3 and MYH7 are the two genes which have been largely implicated accounting to 50% of HCM cases. Objective: Considering the phenotypic and genotypic heterogeneity of HCM and the ethnic/geographic variation existing in the Indian population, screening for the possible genetic variations in MYBPC3 and MYH7 in HCM patients of South Indian cohort was carried out. Methodology: Blood samples of 100 controls and 100 patient samples were collected after obtaining the informed written consent along with the clinical data of the patients for genomic DNA isolation. For screening of MYBPC3 and MYH7genes, PCR based SSCP analysis was carried out. The samples with variation were further confirmed by automated DNA sequencing. Results: The present study revealed 3 novel intronic and 2 novel synonymous variations and 2 cases of compound heterozygosity, apart from 1 reported missense mutation and 1 reported synonymous variation in MYBPC3 gene. Screening of MYH7 revealed 3 novel missense variations and 1 novel synonymous variation which will be discussed. Although these variations are not highly pathogenic but might have a regulatory role in the splicing mechanism, and their effect might be influenced in the presence of modifier genes and environmental factors.

2221T

A family-based whole genome sequencing study to identify genetic determinants of platelet hyper-aggregation following aspirin in African Americans at high risk for coronary artery disease. *M.A. Taub[1], L.R. Yanek[2,3], I. Ruczinski[1], L. Huang[2], D.M. Becker[2,3], L.C. Becker[2,3], R.A. Mathias[2,3], The GeneSTAR Research Program.* 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Johns Hopkins School of Medicine, Baltimore, MD; 3) The GeneSTAR Program, Johns Hopkins School of Medicine, Baltimore, MD.

Background: Low dose aspirin (ASA) is given to inhibit platelet aggregation in individuals at risk for arterial thromboses and is considered standard of care for secondary prevention of coronary artery disease (CAD). We have previously identified common genetic determinants of platelet aggregation following 14-days of ASA, 81 mg/day, using a GWAS approach, and here build on a long-standing study in families with increased risk for CAD (GeneSTAR) applying a whole genome sequencing (WGS) approach. Methods: 5 African American families segregating the phenotypes of hyper-aggregators (>90th percentile of trait distribution in the full sample, N=869) and normal-aggregators (25th-75th percentile) to collagen (5 ug/ml) post ASA. 30 subjects underwent WGS by Illumina, Inc. (depth >30X). Variants were filtered (GQ>20 and segmental duplication regions masked). We determined the set of variants showing perfect co-segregation with phenotype in each family. We tabulated the number of families with perfect co-segregation at (i) each variant site; and (ii) variants collapsed over the units of CCDS transcripts focusing on variants novel to GeneSTAR, i.e. not observed in The Thousand Genomes Project. Results: We observed ~4 million (M) variants per genome, of which ~113K were private variants not shared with another family member. Of the 12.7M sites present across all the samples, 1.6M were novel to GeneSTAR. While we found no sites that showed perfect co-segregation with phenotype in all 5 families, variants in the promoters of HK1 and SLC16A11 co-segregated with phenotype in 3 families. No transcript was commonly implicated across all 5 families. However, gene ontology (GO) analysis of the 218 transcripts that co-segregated with phenotype across 4/5 families includes cell adhesion and wound healing as two of the most significant hits (hypergeometric p-value< 0.01). Restriction to only non-intronic variants within a transcript led to 141 genes in 2 or more families with a GO enrichment in epithelial cell-cell adhesion and cell adhesion genes (p-value < 0.01). Conclusions: This analysis suggests a potential role for rare sequence-identified variants in pathways of genes as determinants of platelet hyper-aggregation post ASA. Analysis is underway to asses statistical significance for these variants, extend annotation beyond coding sequence by leveraging ENCODE data, and test for replication in a sample of 100 independent WGS African American samples from GeneSTAR.

2222F

The molecular dissection of familial dilated cardiomyopathy. *J.M. Taylor[1], S.L. Reid[1], J. Hayesmoore[1], J. Woodley[1], K. McGuire[1], K. Thomson[1], E. Blair[2], H. Watkins[3], A. Seller[1].* 1) Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Trust , Oxford., Oxfordshire, United Kingdom; 2) Department of Clinical Genetics, Oxford University Hospitals NHS Trust , Oxford., Oxfordshire, United Kingdom; 3) Department of Cardiovascular Medicine, West Wing, John Radcliffe Hospital, Oxford OX3 9DU,.

Introduction: Familial dilated cardiomyopathy (DCM) exhibits considerable clinical heterogeneity both within and between families: penetrance is incomplete and age dependant. Combined clinical screening and molecular analysis is therefore essential to identify affected individuals. Currently, there are in excess of 30 genes reported to cause nonsyndromic familial DCM. This study focuses on the genetic heterogeneity of 50 familial DCM patients from a single referral centre and examines the frequency and distribution of clinically relevant variants within this discrete patient cohort.

Method: Molecular analysis of 28 DCM-related genes was undertaken within the Oxford Genetics Laboratory. Enrichment of targeted genes was undertaken using a custom built Haloplex™ kit (Agilent) and sequenced using an Illumina MiSeq. A commercial software package (NextGENe™, by Softgenetics) was used for data alignment, which was then filtered using an in-house pipeline.

Results: On average, 97.5% of target regions were consistently covered to x30 read depth and clinically relevant variants were detected in ~40% of patients. The test sensitivity is higher than expected (previously reported as 17-35% [1,2]). Pathogenic variants were found within genes responsible for the sarcomere and desmosome structure, and the sodium ion channel.

Conclusion: This study demonstrates the utility of an extensive DCM gene panel and further highlights the genetic heterogeneity of familial DCM. Clinically relevant novel genotype-phenotype associations will be discussed.

[1] Hershberger & Siegfried, Am Coll Cardiol 2011; 57:1641-9. [2] Lakdawala et al., J Card Fail. 2012; 18(4):296-303.

2223W

Novel homozygous missense mutations in the SLC2A10 gene in a Turkish pediatric patient with arterial tortuosity syndrome. *S.G. TEMEL[1], O. BOSTAN[2], D. PROOST[3], L. VAN LAER[3], E. CL[2], B. LOEYS[3].* 1) Histology & Embryology, University of Uludag, Faculty of Medicine, BURSA, Turkey; 2) Department of Paediatric Cardiology, University of Uludag, Faculty of Medicine, Bursa,Turkey; 3) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium.

Arterial tortuosity syndrome (ATS) is a rare autosomal recessive connective tissue disorder, mainly characterized by tortuosity and elongation of the large- and medium-sized arteries with predisposition to stenoses and aneurysms. ATS is caused by mutations in the SLC2A10 gene, encoding for the facilitative glucose transporter 10 (GLUT10) and is described typically in pediatric patients. We report a 2 years old boy with ATS who initially presented with a cardiac murmur. Echocardiography at the age of one month demonstrated dilatation of the aortic root (15mm) and tortousity in the descending aorta. The main clinical findings included elongated face, saggy cheeks, micrognathia, malar hypoplasia, joint hypermobility and hyperextensible skin. Follow-up echocardiography showed an increase of the diameter of the aortic root to 32mm and 35 mm at the age of 8 months and 2 years old, respectively. Angiography and computerized tomography angiography showed fusiform aneurysmatic dilatation of ascending aorta, marked tortuosity of both pulmonary arteries and thoracic aorta, stenosis in vena cava inferior and collateral structure associated with right hepatic vein. Sequencing of the SLC2A10 gene in the proband revealed the presence of a novel pathogenic homozygous missense variant (c.727C>A). Heterozygous SLC2A10 mutations were shown in her mother and father, demonstrating true homozygosity. The pathogenic variant leads to a p.Gln243Lys change in the seventh transmembrane domain of GLUT10. The proband underwent aortic root replacement surgery. ATS resembles Loeys-Dietz and Marfan related disorders, so timely differential diagnosis is extremely important for early diagnosis and intervention of aneurysms to prevent serious vascular complications.

2224T

Mutation detection in aortopathy and other vasculopathies complicating hereditary disorders of connective tissue by next generation sequencing. *J. Vandrovcova[1], R.A. Weerakkody[1], J. Biggs[1], P.J. Norsworthy[1], C. Neuwirth[1], L. Game[1], A. Vandersteen[2], F.M. Pope[2], N.J. Cheshire[3], T.J. Aitman[1].* 1) Physiological Genomics and Medicine Group and Genomics Laboratory, MRC Clinical Sciences Centre, Imperial College London, UK; 2) National Ehlers-Danlos Syndrome Diagnostic Service & North West Thames Regional Genetics Service, North West London Hospital NHS Trust, UK; 3) Imperial Vascular Unit, St Mary's Hospital, Imperial College, London, UK.

Hereditary disorders of connective tissue are a heterogeneous group of disorders with overlapping clinical features. A proportion of these patients are at risk of early arterial rupture. Early diagnosis, screening of family members and subsequent antihypertensive therapy and vascular surveillance can therefore be life-saving. Some patients with connective tissue disorders have clinical features that suggest a specific diagnosis such as Marfan syndrome, or vascular Ehlers-Danlos syndromes (EDS), but most patients do not clearly fit such a diagnosis. Genetic testing is available but due to the cost and time associated with Sanger sequencing these tests are usually limited to familial and syndromic forms of the disease. DNA samples were obtained from 300 patients with thoracic aortic aneurysms and dissections (TAAD) and/or with a vascular EDS phenotype. To test the utility of next generation sequencing for detecting mutations in these disorders, amplicons were designed to cover exons and exon-intron boundaries of genes known to be associated with TAAD (*FBN1*, *TGFBR1*, *TGFBR2*, *MYH11*, *ACTA2*, *MYLK*) and EDS (*COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *TNXB*). Samples were amplified using the Access Array System (Fluidigm) and sequenced by MiSeq (Illumina). In the pilot study of 126 samples 98.1% of bases were covered at 25x and 89.1% were covered at 500x. Rare likely pathogenic variants were identified in 15% of aortopathy patients (n=70) and 10% of vascular EDS patients (n=56). Analyses on a further 174 patients are ongoing. Our data indicate that pathogenic mutations in genes previously thought to be associated with only rare familial or syndromic forms of aortopathy may be implicated in a significant proportion of sporadic cases and demonstrate the potential utility of next generation sequencing for routine diagnostics of hereditary disorders of connective tissue.

2225F

New frame-shift deletion in EMD gene causes Emery-Dreifuss muscular dystrophy with severe cardiomyopathy required heart transplantation. *E. Zaklyazminskaya[1], O. Blagova[2], M. Yakovleva[1], V. Rumyantseva[1], A. Nedostup[2], D. Shumakov[3], S. Dzemeshkevich[1].* 1) Medical Genetics Laboratory, Petrovsky Russian Research Centre of Surgery RAMS, Moscow, Russian Federation; 2) Sechenov Moscow State Medical University, Moscow, Russian Federation; 3) Federal Research Centre of Transplantation and Artificial Organs, Moscow, Russian Federation.

Background. Emery-Dreifuss muscular dystrophy (EDMD) is a clinically heterogeneous disease characterized by progressive muscular atrophy, joint contractures, normal intelligence, and variable cardiac involvement. At least 6 genes are known to cause EDMD when mutated (EMD, LMNA, SYNE1, SYNE2, FHL1, and TMEM43). **Clinical case**. Male proband, 38 y.o. , has addressed to genetic counseling because of low progressive skeletal myopathy from the childhood, having walking difficulties but ambulant, moderate knees and elbows contractures, high CK level, normal intelligence, arrhythmias (HR 83-96 bpm, paroxismal atrial fibrillation, AVB I-II, RBBB, premature ventricular beats), and dilated cardiomyopathy (EF LV 33%). Proband's mother has no clinical signs of myopathy but got a pacemaker at the age of 53 y.o. Two sons, 4 y.o. and 11 y.o. are healthy. Informed consent for genetic testing was taken in accordance with Helsinki declaration. Mutations analysis was performed by PCR-based direct Sanger sequencing of coding area and adjacent intronic areas of genes of interest. Two genetic variants were detected: frame-shift deletion c.del619C in EMD gene and c.IVS4-13T>A in LMNA gene. Both variants were not found in control group of 100 healthy volunteers. Taking into account progressing of cardiac arrhythmias and high risk of sudden cardiac death in patients with EDMD, dual-chamber ICD was implanted. Patient had repeated appropriate shocks during 6 month after ICD implantation despite anti-arrhythmic therapy, and progressive impairment of cardiac pump function. Heart transplantation was proposed and successfully performed. There was no anesthetic complication during the procedure. For now, the follow-up period is 3 month with significant health improvement. **Conclusion**. We detected a novel frame-shift mutation c.del619C in EMD gene causing premature stop-codon appearance and protein shortening (p.236X). We suggest pathogenic role of this mutation in familial case of EDMD with predominance of cardiac symptoms. The role and possible modulating effect of rare novel VUCS c.IVS4-13T>A in LMNA gene is to be elucidated. Study of possible effect on LMNA mRNA processing is in progress now.

2226W

Assessment of the enrichment for rare coding variants in 16 related cases of fibromuscular dysplasia. *N. Bouatia-Naji[1,2], R. Kiando[1,2], P-F. Plouin[2,3], X. Jeunemaitre[1,2,3].* 1) Paris Cardiovascular Research Center U970 HEGP research Center, INSERM, Paris, France; 2) Paris Descartes University, Paris Cité Sorbonne, Paris France; 3) AP-HP, Centre for Rare Vascular Diseases, Hopital Européen Georges Pompidou, Paris, France.

Fibromuscular dysplasia (FMD) is an arterial disease characterized by non-atherosclerotic stenoses reported mostly in renal (70%) and extra-cranial carotid (30%) arteries. FMD predisposes to hypertension, renal ischemia and stroke, the first cause of disability worldwide. The causes of FMD are unknown and it occurs predominantly (80% of cases) in females under 50 yrs with a prevalence of ˜4/1000. The genetics of FMD is under-investigated because of the lack of large families and cohorts due to the rarity of the disease and the complexity of the diagnosis that requires artery biopsy and/or complex imaging. The aim of this study was to assess the role of rare coding and putatively functional variants in the onset of FMD in related cases. Patients are five sib-pairs and two sib-trios, all affected (N=16) with confirmed multifocal 'string-of-beads' FMD diagnosed by angiography/scanner. Mean age diagnosis is 44yrs (min=23; max=57). Exome sequencing and variant calling was performed by IntegraGen® using Agilent's capture Kit (V4) and Hiseq2000 (Illumina®) with 54X median coverage. We identified 4,311 confident (high calling-score and depth >8X) missense rare (MAF<0.02 in EVS and/or 1000G) variants predicted to be damaging (PolyPhen2) in 3,414 genes. No gene carried mutations that were specific to all patients or at least 3 out of 7 sibpairs/trios suggesting strong genetic heterogeneity in familial FMD. In order to identify genes to prioritize that might be enriched in mutations in FMD cases, we performed a genomic burden test for highly mutated genes (>= 4 variants, N=44) and used data from ˜4,300 unselected Europeans as controls (Exome Variant Server). Fischer exact test that compared variants vs non-variant alleles counts in cases and controls revealed several genes significantly (P<10e-5) enriched for putatively functional variants at the genomic level. Important fibrosis and loss of extracellular matrix (ECM) at the media of FMD arteries has been demonstrated. Using less significant threshold (P<0.01), we identified two genes involved in ECM structure, cell shape and junction, which are interesting candidates for FMD. Validation of the enrichment of statistical and candidate genes requires the extension of the current sample to larger cohorts of patients either by exome sequencing or genotyping.

2227T

Resequencing of Renin-Angiotensin-Aldosterone-System Genes Identifies Rare Variants Associated with Blood Pressure Salt-Sensitivity: The GenSalt Study. *T.N. Kelly[1], J.E. Hixson[2], L.C. Shimmin[2], Q. Zhao[1], D. Gu[3], J. He[1].* 1) Dept Epidemiology, Tulane Univ, New Orleans, LA; 2) Human Genetics Center, University of Texas, Houston, TX; 3) Cardiovascular Institute and Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, and Chinese National Center for Cardiovascular Disease Control and Research, Beijing, China.

Genetic association studies have revealed significant associations between common variants from renin-angiotensin-aldosterone system (RAAS) genes and blood pressure (BP) responses to dietary sodium intervention. However, the role of rare RAAS variants in salt-sensitivity remains unexplored. We conducted a resequencing study to identify rare functional variants associated with BP salt-sensitivity among participants of the Genetic Epidemiology Network of Salt-Sensitivity (GenSalt) study. The GenSalt study was conducted among 1,906 participants who underwent a 7-day low-sodium (51.3 mmol sodium/day) followed by a 7-day high-sodium feeding-study (307.8 mmol sodium/day). We selected 300 GenSalt subjects with the highest and 300 GenSalt subjects with the lowest mean arterial pressure responses to the high sodium intervention to participate in the current resequencing study. Functional regions of seven RAAS genes, including ACE2, APLN, AGTR1, HSD11B1, HSD11B2, NR3C2, and RENBP were resequenced using the VariantSEQrTM system (Applied Biosystems; Foster City, CA). RAAS variants with minor allele frequencies less than 5% were collapsed according to gene and analyzed using the cohort allelic sums test (CAST). We identified significant associations between rare variants in the APLN, AGTR1, and HSD11B2 genes and BP salt-sensitivity, with p-values of 0.05, 0.03, and 0.03, respectively. Within the promoters, splice sites, exons, and 3' untranslated regions of these 3 genes, we identified 17 rare APLN variants, 20 rare AGTR1 variants, and 19 rare HSD11B2 variants. Nine percent of GenSalt participants with high salt-sensitivity were carriers of at least one of the rare APLN variants, while only 4% of salt-resistant participants were carriers. In addition, approximately 17% of participants with high salt-sensitivity were carriers of rare AGTR1 variants compared to only 11% of salt-resistant subjects. Further, 8% of those who were highly salt-sensitive compared to only 4% of those who were salt-resistant were carriers of HSD11B2 variants. In summary, we provide the first evidence for a role of rare and potentially functional RAAS variants in BP salt-sensitivity. Validation study will be needed to confirm these findings.

2228F

Whole-exome sequencing in familial calcific aortic valve stenosis. *S. LE SCOUARNEC[1,2], C. DINA[2], F. SIMONET[2], C. SCOTT[1], H. LE MAREC[2], V. PROBST[2], M. HURLES[1], N. CARTER[1], T. LE TOURNEAU[2], JJ. SCHOTT[2].* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) L'institut du thorax, Inserm UMR 1087, Nantes, France.

Calcific aortic valve stenosis (CAVS) is a complex disease under the influence of risk factors (e.g. dyslipidemia) as well as genetic polymorphisms (e.g. in APOA). The identification of rare genetic variants with a strong effect would give major insights into the mechanisms leading the bone-like phenotype. Mutations in the NOTCH1 gene have been linked to aortic valve calcification but they only explain a handful of cases and also lead to congenital malformations. Our aim is to uncover the first genes involved in 'degenerative' CAVS, using families that appear to be monogenic autosomal dominant forms of CAVS, including a very large pedigree of >50 cases. First, we sequenced the exome of 20 patients affected by severe CAVS, from 9 families with ≥ 3 cases (mean depth: ~87X per patient). The hundreds of rare variants were prioritised according to their functional consequence and their segregation with the CAVS phenotype in the families. As a validation study, we sequenced a target of 499 kb (Agilent SureSelect capture, Illumina HiSeq sequencing) in 472 CAVS cases. The 499-kb target contained exonic regions of candidate genes from our exome sequencing project and linkage analysis (69 genes) or the literature (31 genes), and regulatory regions. After applying standard frequency and quality filters, we performed burden tests to identify CAVS disease genes. The identification of rare genetic variants within this project could lead to the possibility of alternative patient management for a disease that represents a major public health problem in the ageing population.

2229W

Identification and Characterisation of a Novel Gene for Cardiomyopathy. *D. Phelan[1], G. Wilson[1], K. Pope[1], G. Gillies[1], J. Sim[1], M. Bahlo[2], P. Hickey[2], C. Bromhead[2], P. James[3], D. du Sart[3], M. Delatyki[1], R. Leventer[1], D. Amor[3], P. Lockhart[1].* 1) Murdoch Childrens Research Institute, Melbourne, Australia; 2) Walter and Eliza Hall Institute, Melbourne, Australia; 3) Victorian Clinical Genetics Services, Melbourne, Australia.

Hypertrophic cardiomyopathy (HCM) is a disease characterised by hypertrophy of the left ventricle of the heart. HCM has a prevalence of ~1 in 500 individuals and is the leading cause of sudden cardiac death among young individuals. Over 20 HCM genes have been identified but at least 20% of cases remain of unknown aetiology. An extended family presented to the Victorian Clinical Genetics Service with a syndromic form of HCM in which the affected first cousins also suffered from multiple pterygia (webbing of joints). Both affected individuals were the offspring of a consanguineous relationship therefore an autosomal recessive pattern of inheritance was expected. Copy number analysis and linkage studies were performed using data from high density SNP-array analysis. A single linkage region was identified with a LOD score >1.2. This region of 15q has not previously been associated with cardiomyopathy and achieved a LOD score of 2.9. The linkage encompassed 11.4Mb and contained 169 known genes. No CNVs of interest were identified. Whole Exome Sequencing identified 100 homozygous variants within the linkage interval of which 99 were excluded by bioinformatic filtering based on minor allele frequency and predicted functional consequence. The causative mutation was a homozygous truncating variant in a poorly characterised protein that may function in cardiomyocyte differentiation. To determine the potential contribution of the gene to HCM, we sequenced the gene in a cohort of HCM patients (n=55) previously excluded for mutations in common HCM causing genes. We did not identify any predicted pathogenic variants suggesting mutations in the gene are not a common cause of HCM. To investigate the function of the gene we have characterised the expression in human and mouse tissues. The tissues showing the highest expression levels were the heart and skeletal muscle, consistent with patient phenotype. In addition, we have characterised complex patterns of tissue specific alternative splicing. Immunocytochemical studies of human and mouse heart and skeletal muscle tissue demonstrated predominant localisation to the extracellular matrix. In conclusion we have identified a novel gene for HCM. Further characterisation of this protein and it's cellular function will lead to increased understanding of the molecular basis of cardiomyopathy and result in improved clinical outcomes for affected individuals and families.

2230T

Identification of a single nucleotide polymorphism variant in TYRO3 associated with coronary artery disease risk in the ClinSeq® Study.
H. Sung[1], B. Suktitipat[1], K. Lewis[2], D. Ng[2], SG. Gonsalves[2], JC. Mullikin[3,4], LG. Biesecker[2], AF. Wilson[1], NISC Comparative Sequencing Program. 1) Genometrics Section, Inherited Disease Research Branch, National Human Genome Research Institute, NIH, Baltimore, MD; 2) Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Comparative Genomics Unit, Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) National Institutes of Health Intramural Sequencing Center (NISC), National Human Genome Research Branch, NIH, Bethesda, MD.

ClinSeq® is a large-scale medical sequencing study designed to investigate associations of rare sequence variants with traits related to coronary artery disease (CAD). The study currently includes 1092 non-smoking participants between the ages of 45 to 65 with normal to severe coronary artery calcification scores. About 200 CAD-related traits were measured at the NIH Clinical Research Center in Bethesda, MD. Whole-exome sequencing was performed with the Agilent SureSelect 38Mb and 50Mb capture kits for 387 and 325 individuals, respectively, at NIH Intramural Sequencing Center. Single nucleotide polymorphism variants (SNVs) common to both capture regions with a call rate > 98% and minor allele frequency (MAF) > 1% were used to check for cryptic relatedness and for misspecified population stratification by multidimensional scaling analysis - 635 unrelated European Americans (EAs) were included. For each capture kit with EAs only, SNVs with at least one homozygote and a call rate ≥ 50% were included. The two capture regions with only SNVs in common were merged, yielding 439,807 SNVs. Of these SNVs, 74% and 46% had MAFs < 0.01 and < 0.001, respectively. Individuals were divided into four bins based on their 10-year Framingham risk scores for developing CAD: Bin 1 (< 5%), Bin 2 (5-10%), Bin 3 (> 10%) and Bin 4 (known CAD). The union of all SNVs in 232 individuals in Bin 1 were compared with the union of all SNVs in 89 individuals in Bin 4 using the Variant Annotation, Analysis, and Search Tool (VAAST), a probabilistic search tool in likelihood-framework based on amino acid substitution frequencies. Under the assumption of a recessive inheritance model in terms of best-scoring homozygous or heterozygous variants in the VAAST analysis, a nonsynonymous variant in TYRO3 was identified (rs62001448, MAF=0.095) with a p-value=1.01e-81. The variant allele (T) at rs62001448 was found in a heterozygous state in 16 (7%) times in 232 individuals in Bin 1 and 26 (29%) in 89 individuals in Bin 4. No variant allele (T) was found in a homozygous state. Mouse studies done by Angelillo-Scherrer at al.(2005) and Cosemans et al.(2010) indicated that a loss of function in TYRO3 would decrease thrombus formation and a gain of function would increase thrombus formation. This finding suggests the importance of TYRO3 gene in CAD and provides insight into some of the underlying genetic mechanisms of CAD in the ClinSeq® study.

2231F

Rare variants underlying a common, complex disease: exome sequencing in early-onset myocardial infarction.
H.-H. Won[1,2,3,4], N. Stitziel[5,6], R. Do[1,2,3,4] for NHLBI's Exome Sequencing Project and Myocardial Infarction Genetics Exome Sequencing Consortium. 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 5) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO; 6) Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO.

Myocardial infarction (MI) is the leading cause of death and disability worldwide and is heritable. Studies of common genetic variation have discovered more than 45 genes and loci associated with risk for MI. Whereas rare genetic variation has long been recognized to underlie the inheritance of Mendelian diseases, the extent to which rare variation contributes to complex genetic disorders such as MI has been widely debated. The purpose of this study was to test the hypothesis that rare (frequency < 1%) DNA sequence variation contributes to risk for MI. The protein-coding regions of ~18,500 genes (the 'exome') were deeply sequenced in each of 4,000 individuals with early-onset MI (men with MI ≤ age 50; women with MI ≤ age 60) and in 3,800 individuals without MI. From this initial discovery study, we sought replication using targeted re-sequencing in larger populations (n=13,432). Rare non-synonymous mutations in two genes were more frequent in early-onset MI patients versus MI-free controls at a level of statistical significance exceeding a stringent correction for the number of genes in the genome (threshold = 2×10^{-6}). Deleterious mutations in the low-density lipoprotein receptor (*LDLR*) gene were present in 4.0% of cases compared with 1.3% of controls, representing an odds ratio for disease of 3.1 for mutation carriers ($p = 4\times10^{-10}$). In the gene apolipoprotein A-V (*APOA5*), 1.3% of cases carried a non-synonymous mutation compared with 0.6% of controls, representing an odds ratio for disease of 2.2 for mutation carriers ($p = 5\times10^{-7}$). The mutations associated with MI risk in *LDLR* and *APOA5* prominently relate to elevations in low-density lipoprotein cholesterol and triglyceride levels, respectively. We demonstrate, for the first time, that a burden of rare variants in individual genes contributes to risk for a common, complex disease and definitively establish *APOA5* as a gene responsible for MI in the population.

2232W

Rare and Common Exome Chip Variants are Associated with Fasting Glucose and Insulin Levels - The CHARGE-S Exome Chip and Sequencing Study. *J. Wessel[1], B. Cornes[2], S. Wang[3], R. Jensen[4], P. An[5], L. Lange[6], E. Lange[6], M. Nalls[7], J. Wilson[8], B. Cade[2], Y. Lu[9], R. Loos[9], V.J. Gudnason[10], J. Pankow[11], S. Willems[12], C. van Duijn[12], L. Bielak[13], X. Guo[14], L. Rasmussen-Torvik[15], M. Province[5], I. Borecki[5], J. Dupuis[3], J. Rotter[14], D. Siscovick[4], M. Goodarzi[14], J. Meigs[2], CHARGE Consortium.* 1) Dept. of Epidemiology, Indiana Univ Sch Public Health, Indianapolis, IN; 2) Department of Medicine, Harvard School of Medicine, Boston, MA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) Department of Medicine, University of Washington, Seattle, WA; 5) Department of Genetics, Washington University School of Medicine, St Louis, MO; 6) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC; 7) Laboratory of Neurogenetics, National Institute of Aging, Bethesda, MD; 8) Department of Medicine, University of Mississippi Medical Center; 9) The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai; 10) Icelandic Heart Association; 11) Department of Epidemiology, University of Minnesota School of Public Health; 12) Department of Epidemiology, Erasmus University Medical Center; 13) Department of Epidemiology, University of Michigan School of Public Health; 14) Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA; 15) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine.

Genome-wide association studies (GWAS) have identified 56 loci in which common single nucleotide variants (SNVs) are associated with fasting glucose (FG) or fasting insulin (FI) levels. These SNVs explain <5% of FG or FI variation and are largely non-exonic. To identify new and rare (MAF<1%) SNVs in protein-coding regions, we tested association of FG and FI with SNVs represented on the Illumina Exome Chip. We hypothesized that some known FG- and FI-associated loci identified by GWAS harbor rare protein-coding SNVs, and that large-scale exome-wide analysis will reveal novel loci associated with FG and FI. We tested associations of 161,720 exome chip variants with FG and FI (adjusted for BMI) in up to 13,535 non-diabetic individuals of European (EA) ancestry from 3 studies in the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) consortium. For common SNVs (MAF ≥1%), we used single variant additive association tests. For rare SNVs we used the Sequence Kernel Association Test (SKAT) and the Weighted-Sum (W-S) tests, which aggregate variants across each gene, and combined results across cohorts using the R package, SKATMeta. Models were adjusted for age, sex, site, cohort and principal components covariates. We used SNV annotation from SNPInfo and RegulomeDB to interpret predicted functional significance of SNVs. We considered single variants with $P \leq 2\times10^{-7}$ and SKAT or W-S test results with $P \leq 2.0\times10^{-6}$ to be statistically significant. Common SNVs in known GWAS loci were associated with FG in or near MTNR1B, G6PC2 and GCK had P values$<2\times10^{-7}$; for FI, SNVs in GCKR ($P=3\times10^{-5}$) and PPARG ($P=9\times10^{-7}$) approached significance. None of the FG SNVs were in exons, but the FI SNVs were coding nonsynonymous. 76% of exome chip SNVs had MAF<1%. W-S tests identified 10 novel SNVs at the known FG locus G6PC2 that were suggestively associated with FG ($P=3\times10^{-5}$), including several predicted to be damaging and one (rs146779637), observed in all 3 studies (MAF=.003), that results in a premature stop codon (R283*). W-S testing identified 5 SNVs at a novel locus, PRKG1, that together were suggestively associated with decreased FG ($P=2\times10^{-5}$). We conclude that Exome Chip analyses confirm multiple common SNVs and identify potentially novel rare coding SNVs associated with FG and FI. These findings extend the allelic spectrum of genetic variation underlying diabetes-related quantitative traits and suggest that coding variation may influence glycemic regulation.

2233T

Novel Dilated Cardiomyopathy Mutations Identified by Whole-Exome Sequencing. *N. Chami[1,2], R. Tadros[1,2], M. Beaudoin[1], K. Sin Lo[1], L. Robb[1], F. Lemarbre[1], M. Talajic[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montreal, Québec, Canada; 2) Faculté de Médecine, Université de Montréal, Montréal, Québec, Canada.

Introduction: Dilated cardiomyopathy (DCM) is a disorder of the heart muscle that is characterized by a dilated and impaired contraction of the left ventricle and may ultimately lead to sudden death, heart failure or transplantation. Genetic factors play a major role in DCM; however, known mutations explain less than 20% of familial cases. We hypothesized that a whole-exome approach to identify novel mutations would improve understanding of the genetic basis of DCM. Methods: We performed whole exome DNA sequencing in 26 clinically well-characterized French-Canadian families that are followed at the Montreal Heart Institute's genetic clinic. The inclusion criteria comprised four main conditions: patients had left ventricular ejection fraction, LVEF ≤ 45%, left ventricular dilatation, absence of other known causes of cardiomyopathy, as well as a family history of DCM. Next-generation DNA sequencing was done using the Illumina HiSeq2000 instrument and a paired-ends 2×101 bp protocol. Results: We identified 24,714 novel variants out of which 4,863 are coding. We show that 19% of the families in our study carry unique and novel nonsense mutations in the TTN gene, consistent with previous reports. Further, we identified a nonsense mutation in BAG3, Arg309X, which co-segregated with DCM in three families. The mutation was shared by all 14 affected members and absent in all healthy individuals. Importantly, we show that BAG3 truncating mutations are significantly correlated with early onset of DCM. Conclusion: Mutations in TTN and BAG3 account for a significant proportion of familial dilated cardiomyopathy at the Montreal Heart Institute. Our results highlight the importance of whole exome sequencing in identifying novel causal mutations in DCM which could lead to a better understanding of the disease and improve clinical testing permitting earlier diagnosis in affected families.

2234F

Whole-exome sequencing of a large cohort of patients with congenital heart disease in the Quebec founder population. *C. Preuss[1], M. Capredon[1], C. Prive[1], JC. Grenier[1], M. Samuels[2], P. Chetaille[3], P. Awadalla[1], G. Andelfinger[1].* 1) Department of Pediatrics, Centre Hospitalier Universitaire Sainte Justine, Montreal, QC, Canada; 2) Centre de Recherche de l'Hôpital Ste-Justine, Université de Montréal, Montréal, Quebec, Canada; 3) Cardiology Service, Centre Mère-Enfants, Centre Hospitalier Universitaire de Québec, Université de Laval, Québec City, Québec, Canada.

Congenital heart disease (CHD) is the most common birth defect, the most common cause of infant morbidity and the second most common cause of infant mortality in Canada. It is defined as a structural malformation of the heart and/or great vessels that is present at birth and is of functional significance. Mendelian transmission of congenital heart disease in large families is rare, even in those for which a variable expression of penetrance is observed. We have whole-exome sequenced 96 individuals with left ventricular outflow tract obstructions such as septal defects, aortic valve lesions, coarctation, and supraventricular arrhythmias which have been recruited in Quebec. Library preparation was performed using the TargetSeq enrichment kit and sequences were obtained using the SOLiD5 platform in our in house genome core facility. For several members of multiple affected families genetic linkage mapping has been performed. However, since genetic and phenotypic heterogenity is the norm even for rare Mendelian disease such as CHD, it remains a challenging tasks to make accurate and clinically useful predictions regarding the disease phenotype using classical genetic approaches. Here, we present our recent results from whole-exome sequencing highlighting that genomic, phenotypic and population context matters. The filtering for rare, potential damaging disease variants revealed a number of loci enriched for disease related pathways including histone modification, NOTCH signaling and metalloproteinases involved in heart development. Many of these rare variants identified in newly implicated candidate genes are the subject of ongoing validation and functional studies and might jointly determine the phenotype. In order to test our hypothesis whether the excess of rare, population specific variants contributed to disease risk in individuals, we will validate the identified variants and perform a population wide screening in an independent control cohort from Quebec (CARTaGENE). The heterogenous nature and often not complete penetrance of CHD suggests that composite effects and mutation clustering within families may be rather the norm than the exception for traits that have been primarily studied under the Mendelian mode of disease inheritance.

2235W

A comprehensive association study of *APOE-C1-C4-C2* gene cluster variation with plasma lipoprotein traits in U.S. Whites. *Z.H. Radwan[1], X. Wang[1], F. Waqar[1], J.E. Hokanson[2], R.F. Hamman[2], M.M. Barmada[1], F.Y. Demirci[1], M.I. Kamboh[1].* 1) Human Genetics, GSPH, Univ Pittsburgh, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, Univ Colorado Denver, Aurora, CO.

Apolipoproteins (APO) are major determinants of plasma lipoprotein-lipid distribution and their genetic variation have significant impact on cardiovascular disease (CVD) risk, which is the leading cause of morbidity and mortality worldwide. Therefore, understanding the genetic of plasma lipid traits and determining subjects with high risk of developing CVD is of great public health importance. Although the influence of *APOE* common polymorphisms on plasma lipid traits has been extensively evaluated in diverse ethnic groups, little attention has been paid on other genes located in the same genes cluster on chromosome 19. The objective of this study was to characterize common and rare genetic variation through comprehensive sequencing of the *APOE*, *APOC1*, *APOC4*, *APOC2* genes along with their hepatic regulatory regions *HCR-1*, and *HCR-2* in 95 subjects from US non-Hispanic Whites (NHWs) with extreme HDL-C/TG levels (47 in the upper 10th percentile, and 48 in the lower 10th percentile), and then to genotype the identified tagSNPs and rare variants along with HapMap tagSNPs to cover the intergenic regions in 623 NHWs to evaluate their associations with lipid traits. A total of 115 variants (105 substitution, 10 indels) were identified in the sequencing subset of which 74% were reported, 26% were novel, 54% were common (MAF≥5%), and 46% were rare or less common (MAF<5%). The number of observed unique rare variants was higher in the high HDL-C/low TG group than in the low HDL-C/high TG group (23 vs. 15). Seventy variants (65 sequencing, and 5 HapMap SNPs) were successfully genotyped in 623 subjects followed by association analyses. Gene-based association results confirm the major contribution of this gene cluster on variation in plasma LDL-C, TC, and apoB levels. Single-locus association analysis revealed 20 significant associations with at least one lipid trait (9.65E-13≤p≤0.046). Rare variants analyses showed more significant associations with TC than with other lipid traits. Our findings confirm the significant contribution of this gene cluster in affecting plasma lipid profile and consequently the CVD risk.

2236T

Whole genome sequencing in African American families to identify genetic determinants of extreme obesity. *L. Yanek[1], M. Taub[2], I. Ruczinski[2], L. Huang[3], D. Vaidya[1], L. Becker[1], D. Becker[1], R. Mathias[1], The GeneSTAR Research Program.* 1) The GeneSTAR Research Program, Johns Hopkins School of Medicine, Baltimore, MD; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Johns Hopkins School of Medicine, Baltimore, MD.

Background: African Americans (AA) have the highest rate of obesity in the US. We have noted a dramatically high prevalence of extreme obesity (BMI≥35 kg/m^2) in AAs in the GeneSTAR study of families with a history of early-onset coronary artery disease: 64 (19%) of the 339 AA GeneSTAR families had >1 relative pair with extreme obesity. We leverage whole genome sequence (WGS) data available on 3 AA families containing one or more members with extreme obesity and one or more with normal BMI (18.5-25 kg/m^2) to search for rare genetic determinants of extreme obesity. Methods: Deep (>30X) WGS was performed by Illumina, Inc, in 130 GeneSTAR subjects; analysis was restricted to 17 individuals from 3 families segregating extreme obesity. Variants were filtered (GQ>20 and segmental duplication regions were masked), and we determined the set of variants with 'perfect' co-segregation with phenotype in each family. We tabulated the number of families with 'perfect' co-segregation at (i) each variant site; and (ii) variants collapsed over the unit of each CCDS transcript. Current analysis is restricted to 'novel' sites, i.e., sites not previously noted in the Thousand Genomes Project. Results: There were ~4 million (M) sites per sample, each had ~ 2.7M heterozygous sites and as many as ~217K private variants. Cumulatively, 10.5M variant sites were noted in the 17 individuals, 1.1M of which were novel. We found 18 sites to perfectly co-segregate with the extreme obesity phenotype across all 3 families. Of these, one may affect a binding site of SPI1 to chr12:113740876..113741076. When collapsed over all non-intronic variants within transcripts, 5 transcripts mapping to three distinct genes were found to contain genetic variants perfectly co-segregating with extreme obesity in all 3 families: TJAP1, MUC6 and COL2A1. MUC6 regulates gastric mucus secretion while TJAP1 is a tight junction associated protein and COL2A1 is the alpha 1 chain of collagen Type II. Conclusions: This analysis from the WGS of 3 families suggests there are rare sequence-identified variants that may influence extreme obesity in AA families. Analysis is underway to assess (i) statistical significance of these variants; (ii) consideration of genes with known sequencing issues such as MUC6; (iii) extended analysis leveraging ENCODE data to fully assess the non-coding regions of the genome; and (iv) replication of these signals in a sample of 100 independent AA with WGS data from GeneSTAR.

2237F

***TGFBR1* mutations in a British cohort of Thoracic Aortic Aneurysm and Dissection (TAAD) patients.** *J.A. Aragon-Martin[1], A.M. Mohamed[1], A. Wan[1], A. Saggar[2], M. Jahangiri[3], A.H. Child[1].* 1) Cardiovascular Sciences Research Centre, St George's University of London, Cranmer Terrace, London SW17 0RE, UK; 2) Clinical Genetics Unit, St George's University of London, Cranmer Terrace, London SW17 0RE, UK; 3) Department of Cardiothoracic Surgery, St. George's Healthcare NHS Trust, Cranmer Terrace, London SW17 0RE, UK.

Background: TAAD is inherited as autosomal dominant with variable expression (interfamilial and intrafamilial) and reduced penetrance. Of all non-syndromic TAADs 20% have a family history for aneurysm. TAAD occurs as syndromic (MFS - FBN1, LDS - TGFBR1/2, AOS - SMAD3, EDS - COL3A1) and non-syndromic (SMAD3, TGFBR1/2, ACTA2, MYH11, MYLK) forms; with unidentified genes in 4 known loci (AAT1, AAT2, TAAD3, TAAD5). Recent studies have started to elucidate the genetic basis of non-syndromic TAAD and its genetic heterogeneity. Mutations have been found in genes encoding proteins involved in the extracellular matrix, the TGFb pathway and the contractile apparatus of aortic smooth muscle cells. Candidate genes could be those that interact with the known genes or are part of the same pathway. The aim of this study was to determine the frequency of TGFBR1 mutation in a consecutive series of TAAD patients attending a British TAAD cardiac genetics clinic. Diagnosis was based on pedigree, physical examination, echocardiogram, CT and MRI studies and operative reports.

Method: A total of 101 UK patients (80M:21F, mean age 53±13) with known TAAD who did not fulfill the revised Ghent criteria for MFS, and with no demonstrable mutations in FBN1, TGFBR2, ACTA2 or SMAD3 genes were recruited to this study. These patients were screened for mutations in all exons of TGFBR1 including intron/exon boundaries by bi-directional sequencing. Due to the unavailability of family members segregation studies were not performed. In-silico tools (GVGD, Mutation T@ster, Polyphen-2, SIFT, ESE_finder, BGDP) and human genetic variation databases (1000Genome, EVS) helped to analyse the variants.

Results: Four novel variants (p.H358Y/c.1072C>T, p.A368A/c.1104T>G, p.T375T/c.1125A>C, p.L486W/c.1457T>G) and one recurrent mutation in LDS2A (p.R487W/c.1459C>T) were identified in the serine-threonine kinase domain of the gene. In-silico analysis agrees that these 5 variants are disease causative.

Conclusion: Position 487 is a hot spot. This study suggests that mutations in TGFBR1 (4 novel mutations and 1 recurrent mutation in 5 unrelated probands) are responsible for 3-5% of TAADs and supports data from previous studies. Next-generation sequencing will help with the screening of the remaining samples to elucidate their genetic basis. Contact=jaragon@sgul.ac.uk.

2238W

RNAseq Analysis of Congenital Heart Defects. *D.M. McKean[1], D.S. Herman[1], A.G. Bick[1], D.C. Christodoulou[1], S.R. DePalma[1], J.M. Gorham[1], J. Homsy[1], J.D. Jiang[1], J.D. Overton[6], H. Wakimoto[1], S. Zaidi[6], R.E. Breitbart[2], W.K. Chung[3], R.P. Kim[4], O. Toka[5], R.P. Lifton[6], C.E. Seidman[1], J.G. Seidman[1], Pediatric Cardiac Genomic Consortium.* 1) Genetics, Harvard Medical School, Boston, MA; 2) Boston Children's Hospital, Boston, MA; 3) Pediatrics, Columbia University Medical Center, New York, NY; 4) Children's Hospital Los Angeles, Los Angeles, CA; 5) Institute of Human Genetics, Erlangen University, Erlangen, Germany; 6) Genetics, Yale University, New Haven, CT.

Congenital heart disease (CHD) is the most common type of birth defect and the leading cause of death within the first year of life. To elucidate genetic causes and transcriptional responses to CHD mutations, we performed next generation sequencing and RNAseq on cardiovascular tissues. From 119 CHD subjects, 160 cardiac tissues were obtained from 11 unique sites. For 23 subjects (46 cardiac samples), genomic sequence (whole exome or whole genome) and RNAseq data was available. High quality RNAseq reads were aligned to hg19 using Tophat1.4. Duplicate reads were removed, yielding an average of 20 million reads per subject. Gene expression levels were compared with like tissues from other subjects, from comparable neonatal mouse tissues, and from different tissue from same individuals. 76 genes expressed in neonatal mouse were not expressed in human tissues and 285 genes expressed in human tissues were not expressed in neonatal mouse. 55 genes displayed allele-specific expression (ASE) in multiple human samples (perhaps reflecting cardiac imprinting). 146 genes displayed ASE in ≥1 tissue in a single subject. Interrogation of genomic DNA is underway to consider if ASE in these subjects reflects genomic mutation. Genomic DNA was analyzed to predict RNA splice variants (Maximum Entropy, Neural Network, Splice Site Finder, Human Splice Finder, and Gene Splicer algorithms). Over 1000 unique sites were predicted to alter RNA splicing from one or more algorithms. Among >300 genes expressed in the heart, aberrant splicing was seen with 14 variants, and not seen with 9 variants. Active studies integrating allele loss with predicted splice variants are underway. Ongoing work will define the sensitivity and specificity of splice variant prediction tools.

2239T

Novel rare coding variants underlie blood lipid levels in the population: an exome array association study in 55,000 whites and blacks. *G.M. Peloso[1, 2] on behalf of the CHARGE Lipids Working Group.* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.

Circulating low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) are quantitative, heritable risk factors for coronary heart disease, and genome-wide association screens of common DNA sequence variants have identified many loci associated with these traits. Here, we test the hypothesis that rare coding and splice-site mutations contribute to inter-individual variation in blood lipid concentrations in the population. We characterized ~41,000 European Ancestry (EA) and ~14,000 African Ancestry (AA) samples from 13 studies using a new, rare-variant genotyping array ("the Exome Chip") that was designed based on the protein-coding sequences for ~18,500 genes ("the exome") in >12,000 individuals. Within each of the 13 contributing studies, we tested the association of plasma lipids with individual rare variants as well as sets of variants within each gene. To combine statistical evidence across studies, we performed fixed effects meta-analysis. For single variant association, we set the significance threshold to 7×10^{-8}, corresponding to a Bonferroni correction for 742,932 tests (3 phenotypes x 247,644 variants on the array). For the gene-based association, we set the significance threshold to 5×10^{-7}, corresponding to a Bonferroni correction for 105,444 tests (3 phenotypes x 17,574 genes on the array x 2 tests). The most significantly associated results for each trait replicated known associations in APOE for LDL-C, CETP for HDL-C, and APOA5 for TG. We identified two new genes with rare coding variants associated with lipid levels in EA samples: ANGPTL8 for HDL-C (Q46X, 0.1% frequency, beta=10 mg/dl, P= 9×10^{-11}); and PAFAH1B2 for triglycerides (S161L, 1.7% frequency, beta=-10%, P=4×10^{-09}). Additionally, we found two new gene associations in the AA samples: PCSK7 for HDL-C (R504H, 0.2% frequency, beta=18 mg/dl, P=3×10^{-20}); and COL18A1 for triglycerides (G111R, 1.9% frequency, beta=-16%, P=4×10^{-17}). With the Exome Chip rare variant genotyping array, we have discovered novel associations between rare variants and lipid levels.

2240F

Serum 25 (OH) D deficiency, cardiovascular disease and APOE genotypes: Increased 25 (OH) D status in APOE ε4 carriers does not protect from cardio-metabolic risk. *D. Sanghera[1], H. Finley[1], P. Natt[1], A. Subramanian[1], P. Blackett[2].* 1) Dept Pediatrics, Sect Gen, Univ Oklahoma HSC, Oklahoma City, OK; 2) Dept Pediatrics, Sect Endo, Univ Oklahoma HSC, Oklahoma City, OK.

Our earlier findings reported high prevalence of serum 25(OH) vitamin D deficiency (50<nm/L)to be associated with cardiovascular risk factors including obesity, hypertension, and type 2 diabetes (T2D) in South Asians Sikhs from India. Epidemiological and experimental evidence suggests that APOE ε4 allele protects against vitamin D deficiency, however, it appears contradictory with regard to association APOE ε4 with chronic inflammation and cardiovascular disease. The purpose of this investigation was to examine the distribution of serum 25 (OH) D in an extended sample of individuals from Punjabi ancestry residing in India and US, and to examine role of genetic variation in APOE with vitamin D levels and cardio-metabolic risk. We measured serum 25 (OH) D levels in serum sample of a total of 3,300 Punjabi participants (1,793 T2D cases and 1,560 controls and also genotyped these individuals for APOE polymorphism. A total of 78% individuals were deficient (<50nmol/L) and 56% were seriously deficient (<30nmol/L) in vitamin D levels. A highly significant, inverse relationship between vitamin D levels and T2D was observed (p<0.0001). Furthermore, (82%) diabetics were classified as vitamin D deficient compared to (74%) non-diabetic controls. Furthermore, low vitamin D levels were associated with significantly increased fasting glucose (p<0.007), systolic blood pressure (p<0.003), and body mass index (BMI) (p<0.0001), independent of T2D status. Distribution of APOE genotypes did not differ among T2D patients compared to controls. However, prevalence of APOE ε4 carriers was significantly higher in patients with coronary artery disease (CAD) (14.2%) vs. non-CAD controls (11.9%) (p=0.022). The APOE ε4 carriers had significantly higher serum levels of 25(OH) D compared to non APOE ε4 carriers and non-carriers (30.2±2.5 vs 25.8 ±2.4nm/L, p=0.021), respectively. Despite this, APOE ε4 carriers had higher fasting glucose (p=0.021), higher mean arterial pressure (p= 0.04) and lower HDL cholesterol (p=0.031). Our findings suggest that although APOE ε4 is a modulator of vitamin D status, the adverse metabolic impact associated with APOE ε4 genotype cannot be explained by better serum 25 (OH) D status.

2241W

African ancestry is associated with hypertensive cardiomyopathy in Brazilian patients with heart failure. *C. Fridman[1], M. Cardena[1], A. Ribeiro-dos-Santos[2], S. Santos[2], A. Mansur[3], A. Pereira[3].* 1) Dept of Legal Medicine, Ethics and Occupational Health, Medical School, University of São Paulo, São Paulo, Brazil; 2) Laboratory of Human Genetics and Medicine, Federal University of Pará, Belém, Pará, Brazil; 3) Department of Cardiology, Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School, University of São Paulo, São Paulo, Brazil.

Heart failure (HF) is the final pathway for many diseases affecting the heart; it is recognized as the main risk factor for early morbidity and mortality caused by cardiovascular disease. HF is a complex trait resulting from combination of genetic, environmental, lifestyle and ethnic background, being more prevalent in black individuals. It is known that self-declared ethnicity is not a good method to ethnic classification, especially in admixed population where there is little correlation between self-declared skin color and genetic background. Herein, we investigate the association between genomic ancestry and mitochondrial haplogroups (mtDNA-hg) with different heart failure etiologies in 503 Brazilian patients. Mitochondrial haplogroups were obtained by sequencing the control region of mtDNA. Genomic ancestry was obtained using a set of 48 Ancestry Informative Markers (AIMs) INDELs type. Estimation of global ancestry proportions analyzes was done using Structure v.2.2; statistical analysis was done with SPSS 14.0. Different etiologies were observed within our sample: hypertensive(28.6%), ischemic(28.4%), valve(15.1%), chagasic(11.9%), idiopathic(9.9%) and others(6.1%). In the total sample, African mtDNA-hg was the more prevalent (46.4%), followed by Amerindian (28.2%) and European (25.4%) mtDNA-hg. Genomic ancestry showed the major contribution being European, with 57.5% (±22.1%), followed by African genomic ancestry with 28.3% (±21.7%). The analyzes of associations with different etiologies showed that individuals belonging to African mtDNA-hg and those with greater contribution of African genomic ancestry present increased risk for hypertensive cardiomyopathy development (p=0.003 OR2.05[95%I.C(1.29-3.25)]) and (p=0.002 OR6.07[95%I.C(1.96-18.9)]), respectively. Different physiologic, genetic and environmental factors are known to be related with increased propensity for sodium retention and lower bioavailability of nitric oxide in black individual, as well as increased of oxidative stress; all of them playing an important pathophysiological role in the development of hypertension. This study provided evidence of association between African ancestry (mtDNA and AIMs) and hypertensive cardiomyopathy in HF patients. Besides, this study support the importance of using AIMs and mtDNA to study the genetic of complex diseases in admixed populations, where skin color is not a good indicator of an individual's genetic ancestry. Supported: FAPESP, LIM40-HC.

2242T

Genetic Evaluation in Cardiac Neurodevelopmental Clinic: Diagnosis and Health Care Management. *P.C. Goldenberg[1], A. Parrott[1], B.J. Adler[1], J. Anixt[2], K. Mason[2], B.S. Marino[1].* 1) Cardiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Department of Development and Behavioral Pediatrics, Cincinnati Children's Hospital Medical center.

Objective: To assess burden of clinical and laboratory-confirmed genetic disorders and utility of genetics evaluation in a cardiac neurodevelopmental clinic. **Background:** There is a known high prevalence of genetic and syndromic diagnoses in the pediatric cardiac population. These disorders may have multi-system effects, which may have an important impact on neurodevelopmental outcome. Patients and families may benefit from consultation by genetic specialists in a cardiac neurodevelopmental clinic.**Methods:** A retrospective chart review was conducted of patients evaluated in a cardiac neurodevelopmental clinic from 12/6/2011-4/16/2013. All patients were seen by a cardiovascular geneticist with genetic counseling support. **Results:** A total of 216 patients were included. 30% of patients had a previously determined clinical or laboratory confirmed syndromic diagnosis (65). Following genetics evaluation an additional 18 patients (8%) were given a new genetic or syndromic diagnosis, including rare diagnoses (Ohdo, Kabuki, and Axenfeld-Rieger syndromes), prenatal teratogenic exposures (Depakote), copy number variants, malformation associations (MURCS, VACTERL), and typical genetic diagnoses in cardiology populations (Williams, Noonan, 22q11.2 deletion, CHARGE, and Rubinstein-Taybi syndromes). There was significant increase in post-visit clinical and laboratory confirmed syndromic diagnosis in comparison to pre-visit diagnosis (p<0.001). Genetic testing was recommended for 112 (52%)patients, imaging was obtained on 17 patients (8%), 73 patients had additional recommended referrals (34%), and 14 patients had clinical lab studies (6%). Syndrome-specific guidelines were available and followed for 26 patients with known diagnosis (12%), including some that had been lost to genetics follow-up. Red Book guideline asplenia recommendations were made for 6 heterotaxy patients, and 23 families affected by LV outflow tract obstruction were advised to have echocardiograms (11%).**Conclusion:** Genetics involvement in cardiac neurodevelopmental clinic can be helpful in identifying new unifying clinical or laboratory confirmed genetic diagnoses and providing syndrome-specific care, which may impact the patient's overall health status and neurodevelopmental outcome..

2243F

Incidence of aortic root dilatation in patients with 22q11 deletion: The 3-year experience of a VCF specialty clinic. *C.B. Hills[1], S.L. Dugan[2], M.E. Pierpont[3], J.D. Sidman[4].* 1) Children's Heart Clinic, Minneapolis, MN; 2) Genetics Division, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 3) Division of Pediatric Genetics and Metabolism, University of Minnesota, Minneapolis, MN; 4) Children's ENT and Facial Plastic Surgery, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN.

Background: Congenital cardiac malformations are commonly present in patients with 22q11 deletions. Frequent malformations include cono-truncal defects such as tetralogy of Fallot, interrupted aortic arch, vascular rings/ right aortic arch, and septal defects. Aortic root dilatation has been reported in this population, but the incidence and association with specific cardiac diagnoses has not been well-defined. Method and Results: A retrospective chart review was performed for all patients seen at the Velocardiofacial Syndrome Specialty Clinic at the Children's Hospitals and Clinics of Minnesota from 2009-2012. A total of 78 individual patients were identified (age range newborn to age 29 years). Echocardiogram reports and images were reviewed for each patient. Fourteen patients were referred from outside institutions and aortic root dimensions were not available. All available echocardiograms were reviewed for each patient by a pediatric cardiologist. A total of 179 echocardiograms performed on the remaining 64 patients (range 1-20 studies per patient) were diagnostic for determining the aortic root dimension at the level of the sinuses of Valsalva. Aortic root z-scores were calculated for each diagnostic study. Thirty-two (50%) of the patients were found to have at least one study in which their aortic root dimensions z-score exceeded a valve of +2. Incidence by cardiac diagnosis varied, with the highest incidence found in patients with tetralogy of Fallot (77%); vascular ring (70%); and interrupted aortic arch (67%). Six patients out of 20 (30%) with structurally normal hearts were found to have aortic root enlargement on at least 1 echocardiogram. Several patients were found to have aortic root dimensions above normal limits with subsequent normalization of the root size for their body surface area; in others, the aortic root enlargement occurred over time. None of the patients in this cohort have required aortic root replacement or have had an aortic root dissection to date. Conclusions: Aortic root enlargement is a common finding in patients with 22q11 deletions, even in the presence of a structurally normal heart. Acute aortic root dissection has been reported only rarely in patients with 22q11 deletions, but additional research is needed to determine the incidence of progressive root enlargement with time and possible need for ongoing screening studies in this population.

2244W

Identical Adolescent Twins with Progressive Thoracic Aortic Aneurysm, Recurrent Dissection and ACTA2 Mutations. *R.B. Hinton, A. Shikany, S.M. Ware.* Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Thoracic aortic aneurysm (TAA) is a subclinical disease and some cases are associated with aortic dissection and sudden death. TAA can occur as an isolated nonsyndromic condition or in conjunction with inherited connective tissue disorders such as Marfan syndrome. The genetic basis of isolated TAA is identifiable in approximately 20% of cases with up to 14% attributable to mutations in smooth muscle actin (ACTA2). We identified 17 year-old identical twin brothers with severe fusiform TAA (aortic root Z-scores +8.0 and +7.7, respectively, and ascending aorta Z-scores +7.6 and +12.7) that progressed to multiple dissections. Twin 1 presented with dissection of an abdominal aortic aneurysm (AAA). Both brothers had marked aneurysmal dilation of the internal carotid arteries and pulmonary emboli. Eye findings were notable for mydriasis. Molecular testing identified a de novo ACTA2 p.Lys328Asn heterozygous mutation. Testing of FBN1, TGFBR1, TGFBR2, and MYH11 was normal. Histologic analysis of thoracic and abdominal aorta samples showed architectural abnormalities including subintimal hyperplasia, advential fibrosis and elastic fiber fragmentation. Smooth muscle actin, the protein encoded by ACTA2, is misexpressed in both thoracic and abdominal aorta tissue types, but the abnormalities appear more severe in TAA. Given the known inflammatory component present in AAA, we examined inflammatory markers in both TAA and AAA. Interestingly, CD-45, which is diffusely present in control tissue, is selectively increased in the adventitia layer only of TAA but increased in clusters of cells in both the media and adventitia of AAA. CD-68, a macrophage marker that is normally absent from aorta, is weakly present in TAA adventitia and strongly present in AAA media and adventitia. TGFβ signaling was also examined. Canonical (p-Smad2/3) and non-canonical (p-Erk1/2) signaling was present in controls in all three layers of the aorta. In contrast, both canonical and non-canonical signaling were strongly increased in TAA and moderately increased in AAA with disproportionate increases in the adventitia. These individuals with ACTA2 mutation demonstrate that severely progressive aneurysmal disease is characterized by smooth muscle and TGFβ signaling abnormalities in conjunction with a potential role for inflammation in the thoracic aorta.

2245T

Erythrocyte sedimentation rate (ESR) in ischemic stroke and its relation with stroke risk factors. *A. Kaur[1], A. Uppal[2], K. Kaur[1].* 1) Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Uppal Neuro Hospital, Amritsar, Punjab, India.

Elevated ESR level is considered to be reliable marker of chronic and acute inflammation. Inflammation is hallmark of ischemic stroke. Cardiovascular risk factors, chronic and acute inflammation appears to be linked closely to each other and to aggravate the occurrence of various thrombotic events involved in pathophysiology of ischemic stroke. Objective: To investigate the relationship between ESR levels and various traditional risk factors in ischemic stroke. Methodology:The study included total of 107 ischemic stroke patients hospitalized in Uppal Neuro hospital, Amritsar (Punjab), diagnosis being confirmed with MRI. Any history of hypertension, diabetes, CAD and atrial fibrillation (AF) was assessed. The patients were classified into three groups according to ESR values measured within 24 hours of admission. Results and Conclusion: A total of 107 ischemic stroke patients including 61 males and 46 females were evaluated for ESR levels, hemoglobin (Hb) content, TLC, neutrophils and lymphocyte count. Patients with ESR≤10mm/hr were included in group I (n=49) with mean age 60.1±12.2. Group II (ESR 11-25) included 72 patients with mean age 65.7±12.2 and patients (n=11) in group III with ESR levels ≥ 26mm/hr had mean age 67.6±12.4. Group I showed higher number of males (79.2%) but Group III with abnormally high values of ESR included most of females (90.9%) than group II (43.1%) and group I (20.8%). Comparison among three groups with respect to age, Hb content, TLC, neutrophils and lymphocytes was carried out with ANOVA (univariate). The difference of mean Hb levels among three groups have been found to be statistically significant (F=20.02, p<0.001). No statistical difference was observed among three groups with respect to age (F=2.29, p=0.106), TLC (F=0.78, p=0.46), neutrophils (F=2.36, p=0.09) and lymphocytes (F=2.36, p=0.09). In terms of stroke risk factors including hypertension, diabetes, CAD, no significant difference was observed. It has been concluded that higher ESR values on admission are associated with low Hb content while on other hand they are not significantly related with various traditional stroke risk factors, may be due to small sample size. Large sample size (being studied) could provide some useful information regarding the relationship between ESR and stroke risk factors.

2246F

Alpha cardiac actin mutation co-segregates with hypertrophic and dilated cardiomyopathy as well as being associated with conduction disease. *J. McGaughran[1,2], L. Hunt[1], M. Christiansen[3], J. Atherton[2,4].* 1) Gen Hlth Queensland, Royal Brisbane & Women's Hosp, Brisbane, Queensland, Australia; 2) University of Queensland, School of Medicine, Brisbane, Queensland, Australia; 3) Statens Serum Institute, Copenhagen, Denmark; 4) Department of Cardiology, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia.

Mutations in alpha cardiac actin (ACTC1) have been rarely reported in nonsyndromic familial dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM). We report a family with both HCM and DCM associated with conduction disease requiring pacing that co-segregates with an ACTC1 missense mutation. A 40 year old woman with HCM presented for genetic testing. Her father was also said to have HCM and died whilst awaiting cardiac transplantation. She was aware that 2 of her paternal uncles were also said to have cardiomyopathy. An ACTC1 missense mutation (c.76 G>A) was identified on genetic screening. Further family studies were suggested to clarify the pathogenicity of the change. One uncle initially presented at the age of 51 years with atrial fibrillation associated with normal LV size, wall thickness and systolic function. He subsequently developed symptomatic atrioventricular (AV) conduction disease, dilated cardiomyopathy and clinical heart failure at 64 years of age and pacing was performed in view of 3rd-degree AV block (despite no negative chronotropic agents). LMNA mutation screening was negative. He also had the ACTC1 missense mutation. Another paternal uncle diagnosed with HCM also had the ACTC1 missense mutation. Conclusion: This case illustrates the phenotypic variation associated with sarcomere mutations. ACTC1 mutations have been previously reported in a small number of HCM and DCM families, however has not been previously reported in associated with conduction disease.

2247W

Cumulative Risks for Thoracic Aortic and Other Vascular Diseases Associated with ACTA2 Mutations. *E.S. Regalado[1], R.E. Lasky[2], D. Guo[1], T.A. Bensend[1], G. Arno[3], A. Child[3], T. Morisaki[4], D. Liang[5], D.M. Milewicz[1].* 1) Internal Medicine, University of Texas Health Science Center at Houston, Houston, TX; 2) Pediatrics, University of Texas Health Science Center at Houston, Houston, TX; 3) St. George's University of London, London, UK; 4) National Cerebral and Cardiovascular Center Research Institute, Suita, Japan; 5) Stanford University, Stanford, CA.

ACTA2 mutations are the most common cause of familial thoracic aortic aneurysms and dissections (TAAD) responsible for 10-14% of the disease and predispose to occlusive vascular diseases leading to stroke and coronary artery disease (CAD) with strong association of specific mutations with vascular diseases. Disease risk estimates are important for genetic counseling and surveillance of at-risk patients, but are not available for ACTA2 mutations. Using data collected from the largest cohort of individuals with a spectrum of ACTA2 mutations (n = 217), we determined Kaplan-Meier estimates of failure function for TAAD (aortic aneurysms requiring surgical repair or aortic dissections), CAD or myocardial infarct (MI), and stroke for all mutations and specific mutations occurring in at least 2 unrelated families. By age 70 years, 85% (95% CI= 74-93%) of individuals had TAAD. Percentage of individuals who had TAAD from birth to oldest age of follow-up was 100% by age 25 years for Arg179, 100% by age 45 years for Arg198, 93% (95% CI= 74-100%) by age 63 years for Arg258, 85% (95% CI= 53-99%) by age 62 years for Arg118, 75% (95% CI= 59-88%) by age 70 years for Arg149, 70% (95% CI= 37-96%) by age 76 years for Arg212, 68% (95% CI= 44-90%) by age 65 years for Arg39, 67% (95% CI= 32-95%) by age 56 years for Gly160, and 60% (95% CI= 17-99%) by age 77 years for Arg185. Using the logrank test, risk for TAAD was significantly higher for mutations altering Arg179 (p<0.001) and Arg258 (p=0.017) compared to other ACTA2 mutations. Eighteen percent (95% CI= 12-28%) of individuals had early onset CAD or MI and 13% (95% CI= 9-20%) had early onset stroke by age of 55 years. Mutation-specific risk estimates were also calculated for early onset stroke and CAD or MI. Risk for early onset stroke was significantly higher for mutations altering Arg179 (42%, 95% CI= 17-80%, p<0.001) and Arg258 (41%, 95% CI= 20-71%, p=0.001) but lower for Arg149 (4%, 95% CI= 1-26%, p=0.031). None of the specific mutations were significantly associated with early onset CAD or MI, although Arg149 showed borderline significant association (28%, 95% CI= 14-51%, p= 0.058). These findings demonstrate variable disease risk specifically for ACTA2 mutations altering Arg179 and Arg258 and provide detailed risk estimates that will aid in genetic counseling, vascular imaging and management of patients, and provide a guide for assessing aortic disease risk for other genes that predispose to this disease.

2248T

Immunohistochemical study of lysosomal Cathepsin A deficient mouse model shows accumulation of vasoactive peptides in brain. *V. Seyrantepe, S.C. Ozturk, Z. Timur, S. Akyildiz, O. Ozdemir.* Dept Molecular Biology and Genetics, Izmir Institute of Technology, Urla Izmir, Turkey.

Cathepsin A is lysosomal enzyme that makes complex with α-neuraminidase and β-galactosidase. In the complex, it activates α-neuraminidase and stabilizes β-galactosidase. Cathepsin A is an acidic serine carboxypeptidase with deamidase and esterase activities at neutral pH. It catalyzes the hydrolysis of C-terminal peptide bond which potentially affects vasoactive peptide actions in normal and pathological conditions. In order to elucidate the biological function of Cathepsin A in vivo, we previously generated knock-in mouse model with catalytically inactive Cathepsin A protein. We showed that Cathepsin A-/- mice have significantly increased arterial blood pressure and higher level of endothelin-I secreted from cultured brain cells. In this work, we present our immunohistochemical study of vasoactive peptides in brain and visceral organs from Cathepsin A -/- mice and their wild type-siblings. We showed higher level of endothelin-I and substance P accumulation in brain from 3 and 6 months old Cathepsin A-/- mice compared to wild-type mice. However, we found that Cathepsin A deficency did not result in significant accumulation in kidney, liver and lung. The detailed analysis of other vasoactive peptides including angiotension I, bradykinin and oxytocin in brain and visceral organs of Cathepsin A-/- mice will be valuable to understand better the physilogical role of this lysosomal enzyme in the modulation of vasoactive peptides and complex processes such as regulation of blood pressure.

2249F

The Yield of Clinical Genetic Testing for Isolated Left Ventricular Non-compaction (LVNC). *A.R. Shikany, A.S. Parrott, S. Ware, E.M. Miller.* The Heart Institute, Cincinnati Children's Hospital, Cincinnati, CO.

Left ventricular noncompaction (LVNC) is a distinct form of cardiomyopathy characterized by two layers of compacted and noncompacted myocardium with prominent trabeculations in the noncompacted layer. LVNC is clinically heterogeneous with subtypes including isolated LVNC (LVNC with otherwise normal left ventricular size, thickness, and function) and non-isolated LVNC (LVNC co-occurring with another form of cardiomyopathy). Mutations in 11 genes known to cause cardiomyopathy have been identified in 40% of cases of LVNC, not defined by subtype. This study sought to investigate the yield of genetic testing among patients with isolated LVNC. This retrospective chart review included patients evaluated in the multidisciplinary Cardiomyopathy Clinic at Cincinnati Children's Hospital between July 2009 and December 2012. Clinical data was collected on 169 patients from 144 families for whom LVNC was confirmed by a cardiologist. Seventy-one probands had a diagnosis of isolated LVNC and 73 probands had non-isolated LVNC. Of individuals with isolated LVNC, 46% (33/71) had genetic testing, of which 12% (4/33) had a disease-causing mutation. The yield of genetic testing in the non-isolated LVNC group was 33% (16/48). A variant of unknown significance (VUS) was identified in 21% (7/33) and 17% (8/48) of the isolated LVNC and non-isolated LVNC subtypes respectively. Of the individuals with isolated LVNC who had a disease-causing mutation, all (4/4) had a reported family history of LVNC, cardiomyopathy, or sudden cardiac death. Given the low yield of testing and risk for a VUS, results from this study suggest that testing for isolated LVNC may be of limited clinical utility in the context of a negative family history. Screening of first-degree relatives should be considered prior to genetic testing to confirm family history status. Healthcare providers should consider LVNC subtype in their provision of counseling regarding the benefits and limitations of genetic testing for LVNC.

2250W

Causative analysis of markers related to PR interval properties and atrial fibrillation after cardiac surgery. *M.I. Sigurdsson[1], J. Muehlschlegel[1], A. Fox[1], M. Heydarpour[1], P. Lichtner[2], T. Meitinger[2], C. Collard[3], S. Shernan[1], S.C. Body[1].* 1) Anesthesia, Perioperative and Pain Medicine, Brigham and Women's Hospital , Boston, MA; 2) Institut für Humangenetik, Neuherberg, Germany; 3) Division of Cardiovascular Anesthesia, Texas Heart Institute, Houston, TX, USA.

Several genetic loci linked to atrial fibrillation (AF) have also been associated with PR interval implying a mechanistic role. We determined the association between genetic variants and AF following coronary artery bypass surgery by pathways including, and independent, of the PR interval utilizing methodology that adjusts the association between a genetic marker and AF for properties of PR interval. Genetic variants associated with AF or PR interval from literature were tested for association with AF and PR interval in a cohort of 1314 patients undergoing coronary artery bypass grafting. Those associated with both were analyzed with association with poAF independent of the PR interval. Patients with AF had longer pre-operative PR interval, maximum PR interval and maximum change in PR interval and change in PR interval from pre-operative value. Two regions on 1q21 and 4q25 had several markers associated with both AF and the PR. All of those had association with AF independent of the PR interval. In contrast rs740178 on chromosome 16 (was associated with AF only via the PR interval. Our results highlight the utility of studying the molecular background of cardiac pathophysiology in a thoroughly phenotyped cohort of patient undergoing cardiac surgery.

2251T

Clinical and mutational spectrum in patients with cardiac glycogenosis related to mutations in the prkag2 gene. *J. Thevenon[1], G. Laurent[2], P. Charon[3], P. Laforêt[4], A. Millaire[5], D. Klug[5], L. Gouya[6], C.A. Maurage[7], S. Kacet[5], J.C. Eicher[2], X. Jeunemaitre[8], M. Desnos[8], E. Bieth[9], P. Bouvagnet[10], D. Duboc[11], L. Martin[12], P. Réant[13], F. Picard[13], C. Bonithon-Kopp[14], E. Gautier[14], C. Binquet[14], C. Thauvin-Robinet[1], J.E. Wolf[2], L. Faivre[1], P. Richard[15].* 1) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, Hôpital d'Enfants, CHU Dijon et Université de Bourgogne, France; 2) Centre de Cardiologie Clinique et Interventionnelle, Hôpital du Bocage, Centre Hospitalo-Universitaire de Dijon, Dijon, France; 3) AP-HP, Département de Génétique, Centre de référence des maladies cardiaques héréditaires, INSERM UMRS956, UPMC Université Paris 6 ; Hôpital de la Pitié-Salpêtrière, Paris, France; 4) Centre de Référence de pathologie neuromusculaire Paris-Est, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Paris, France; 5) Hôpital Cardiologique, CHRU de Lille, France; 6) INSERM U773, Centre de Recherche Biomédicale Bichat Beaujon CRB3, Université Paris 7 Denis Diderot, site Bichat, Paris, France; 7) Université Lille Nord de France, USDL, EA 1056, F-59000 Lille, France ; Department of Pathology, Lille University Hospital, F-59000 Lille, France ; INSERM U837, F-59000 Lille, France; 8) INSERM, UMRS_970, Paris Cardiovascular Research Center, France; Département de cardiologie, hôpital européen Georges-Pompidou, Assistance publique-Hôpitaux de Paris, 20, rue Leblanc, 75015 Paris, France; Inserm U 633, faculté de medicine, université Paris; 9) Hôpital Purpan, Department of Medical Genetics, Toulouse, France; 10) Laboratoire Cardiogénétique, Groupe Hospitalier Est, Hospices Civils de Lyon, Lyon, France; 11) Service de Cardiologie, Hôpital Cochin, Paris, France; 12) Laboratoire d'anatomopathologie, Plateau technique de Biologie, CHU de Dijon, France; 13) Department of Surgical Cardiology, Haut Leveque Hospital, Pessac, France; 14) Centre d'investigation clinique -épidémiologie clinique/ essais cliniques, CHU, Dijon, France; 15) APHP, UF Cardiogénétique et Myogénétique, Service de Biochimie Métabolique, Groupe Hospitalier Pitié-Salpêtrière 47-83 boulevard de l'Hôpital, 75651 Paris cedex 13, France.

Background: PRKAG2 mutations are responsible for an autosomal dominant muscular glycogenosis associating hypertrophic cardiomyopathies (HCM), ventricular pre-excitations (VPE), progressive atrio-ventricular blocks (AVB) and sometimes sudden cardiac deaths (SCD). **Objectives:** To describe the time dependent clinical occurrence of clinical manifestations of the disease. To define the indications for PRKAG2 mutation screening. **Methods and Results:** Clinical and molecular data of 34 patients from 9 families were gathered from a national French cohort of 140 patients tested for PRKAG2 gene since 2001. The pick-up rate varied from 0% when the indication was an isolated HCM to 20% when associated with a VPE (isolated short PR interval or WPW syndromes). Aside from the recurrent p.Arg302Gln (48%), we identified four new mutations. Time dependant risk of complication occurrence was estimated using a Kaplan-Meier method. At 40 years of age the chances to present with a VPE was 70% (99%-CI: 50%-87%), with a HCM was 61% (99%-CI: 41%-81%), and to die from a SCD was 20% (99%-CI: 8%-42%). At the 66 years, 32% of the patients required an implantable cardiac defibrillator and 2 required heart transplant at 55 and 60 years. No significant genotype-phenotype correlation has been found. However p.Arg302Gln mutation was less often associated with a HCM and no WPW syndrome was diagnosed in the p.Arg302Gln group. WPW radiofrequency ablation was associated with a complete AVB in 4/6 patients. **Conclusion:** Because of the risk of SCD, PRKAG2 screening may be useful in case of VPE with or without HCM.

2252F

Genetic profiling supports the causal role of type 2 diabetes, and fasting insulin and glucose in cardiovascular diseases. *E. Tikkanen[1,2], M. Pirinen[1], A.S. Havulinna[2], V. Salomaa[2], S. Ripatti[1,2,3,4].* 1) Institute for Molecular Medicine Finland FIMM, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Hjelt Institute, University of Helsinki, Helsinki, Finland; 4) Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Epidemiological studies have identified several traits associated with cardiovascular disease (CVD), but few of these have been shown to be causal risk factors. Genetic information obtained from genome-wide association studies (GWAS) could be used to assess the causality of risk factors on disease. The aim of this study was to examine the causal biological pathways for CVD by genetic profiling of several risk factor traits. By utilizing the information from published GWAS, we generated genetic risk scores (GRSs) for blood lipids, obesity, blood pressure, inflammation and type 2 diabetes -related traits in Finnish case-cohort dataset (N=1974) with extensive clinical and follow-up data. The GRSs were calculated by summing the number of risk alleles and dividing the sum with the number of SNPs, after which each GRS were standardized (mean=0, SD=1). Associations between the GRSs and cardiovascular events were tested with Cox proportional hazards models and two-sided $P<0.05$ were considered to be statistically significant. To evaluate whether the GRS effects are governed by the proportion of heritability (h2) explained by the genetic variants, we estimated h2 from genome-wide data of Finnish cohorts (N(max)=14,073) and calculated the proportion of h2 explained by the GRS for each trait. We found that the GRSs for insulin, type 2 diabetes, triglycerides and total cholesterol increased the risk for incident coronary heart disease in an analysis adjusted for age and gender. The GRSs for type 2 diabetes and glucose were associated with incident stroke. In a multivariable analysis accounting for all risk factor traits and genetic risk scores, the GRS for insulin remained associated with coronary heart disease (HR=1.44, 95% CI 1.15-1.79). The GRSs for insulin and glucose account for 19% and 45% of the trait heritability, respectively. The proportion the GRS explains of the h2 was not correlated with the association of trait-based GRS with CVD risk (r=-0.005). We conclude that these results support the causal role of type 2 diabetes -related traits for cardiovascular disease.

2253W

Left ventricular non-compaction, Ebstein's anomaly and autosomal dominant polycystic kidney disease: A novel association and review of the vascular features of ADPKD. *L. Zahavich[1], A. Dipchand[1, 2], S. Bowdin[1, 2].* 1) The Hospital for Sick Children, Toronto, Canada; 2) University of Toronto, Toronto, Canada.

Autosomal dominant polycystic kidney disease (ADPKD) is the most common inherited renal disease and one of the most common genetic disorders with a prevalence of 1/400-1/1000. There are two known genes associated with ADKD, PKD1 and PKD2 accounting for about 90% of cases. Left ventricular non-compaction (LVNC) is characterized by persistent fetal myocardium leading to trabeculations and a thickened 2-layer myocardium. In some patients, this can result in heart failure. Familial recurrence of LVNC has been reported in about 20% of cases. Genetic causes include sarcomere gene mutations, Barth syndrome and chromosomal variants. We present a case of a 2 year old boy who was diagnosed with Ebstein's anomaly and echogenic kidneys prenatally. He presented with congestive heart failure at 3 weeks of age and was then diagnosed with LVNC. At 9 months he was found to have bilateral renal cystic disease. Family history is significant for PKD, bicuspid aortic valve and aortic aneurysm in his mother as well as PKD and valvular disease in extended maternal relatives. Testing of the MYH7 gene was organized given the association between LVNC, Ebstein's anomaly and MYH7 mutations and was normal. The patient also had a chromosome microarray which was normal. He was subsequently found to have a well described mutation in PKD1 (p.Q4042X). Vascular abnormalities are frequent in patients with ADPKD, particularly hypertension and valvular disease. Common valvular anomalies include mitral valve prolapse, mitral or tricuspid valve regurgitation and aortic root dilatation. There have been reports of left ventricular hypertrophy associated wth ADKPD, including in those who are normotensive. Previous studies have also noted an increased prevalence of dilated cardiomyopathy in patients with ADPKD. To date, there have been three reported cases of LVNC in patients with ADPKD, two adults and one pediatric case. The pathogenesis of ADPKD is currently unknown, however a 'two-hit' model of cyst formation has been hypothesized. This model proposes that PKD null cells induce apoptosis of surrounding normal cells leading to cyst formation. Cardiac defects and abnormal myocyte trabeculations have been described in PKD1 mutant mice. A potential mechanism for LVNC in patients with ADPKD could involve induced apoptosis of the myocardium leading to persistent trabeculations. Further studies of the cardiac phenotype in ADKPD are required to assess the relationship between ADPKD and LVNC.

2254T

Identification of Copy Number Variation (CNVs) in patients with Isolated Conotruncal Heart Defects: a family trio study. *J. Arteaga-Vazquez[1], A. Aguayo[1], Y. Svyryd[1], G. Vargas[2], JE. Calderon[2], C. Zamora[2], O. Mutchinick[1].* 1) Genetics, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, México, D.F., México; 2) Instituto Nacional de Cardiología Ignacio Chávez. México, D.F.

Background: Copy number variations (CNVs) were initialy recognized as genomic non-functional rare rearrangements but currently around 180,000 CNVs are identified along the human genome and contribute approximately to 5% of genome variation among individuals. CNVs can impact phenotypes by gene dosis modifications. Besides, Tetralogy of Fallot (ToF) is present in 1/3000 live births and accounts for 60% of conotruncal malformations. Associated mutations have been described in early developmental genes including *NKX2-5*, *NOTCH1*, *TBX1*, *JAG1*, *NOTCH2*. Studying nuclear families, Greenway et al. identified seven CNVs associated with an increased risk for ToF, three of which involved genes previously associated with heart disease: *TBX1*, *NOTCH1* and *JAG1*. On the other hand, MLPA is a high performance and low cost technique that allows the analysis of over 40 target genome sequences with a single pair of primers in a single reaction. Aim: To identify the frequency and type of certain CNVs related to congenital heart disease, including those present in specific genes (*GATA4*, *NKX2-5*, *TBX5*, *BMP4*, *CRELD1*) and other 48 genomic sequences distributed throughout the chromosomal regions: 10p14, 4q35, 17p13, 22q11 and 22q13 by means of MLPA technique in a sample of Mexican patients with ToF, their parents and a control group. Material and Methods: 28 family trios in which the propositus had the diagnosis of ToF and 28 healthy controls were analyzed with the MLPA technique using the P311 and P250 kits from MRC-Holland®. Real Time-PCR and FISH with specific probes were used to validate the results. Results: One propositus showed a de novo CNV in the *TBX1* gene, involving exons 2 and 7, both parents were normal. We did not observe any CNV in the control group. Conclusions: We identified a *TBX1* gene deletion in 1/28 cases with ToF, which was not detected by conventional karyotype or FISH analysis for the 22q11 region. The proportion of cases with de novo CNVs in our sample is according to the observed by other research group. The *TBX1* variant in our patient could explain the presence of ToF due to the participation of this gene in early cardiac development. The MLPA technique allows the detection of chromosomal alterations such as microdeletion and microduplication not identified by conventional karyotype or FISH analysis, at a low cost and with a fast performance.

2255F

Identification of Global Methylation Markers for Myocardial Infarction in Males. *J.M. Devaney[1, 2], B.T. Harmon[2], E.P. Hoffman[2], S.E. Epstein[1].* 1) Head, Translational and Vascular Biology Research, MedStar Health Research Institute, Washington, DC; 2) Dept. Integrative Systems Biology Research Center for Genetic Medicine Children's National Medical Center 111 Michigan Ave, NW Washington DC 20010.

The leading cause of death in the world is coronary artery disease (CAD). There have been numerous risk factors identified for the development of CAD and it's main complication myocardial infarction (MI), including high plasma lipid concentrations, high blood pressure, smoking, diabetes, and markers of inflammation. We hypothesized that epigenetic markers may add to the growing etiology of myocardial infarction and provide additional biomarkers that can be used to predict MI in individuals with CAD. METHODS: The DNA was extracted from whole blood and analyzed for methylation changes using the Illumina HumanMethylation450 BeadChip (485,577 CpG sites). Statistical analysis was completed on 395,899 CpG sites (Partek Genomics Suite) comparing males with stable CAD (n = 7) and CAD and MI (n = 7). Any CpG sites with an average Beta value less than 0.1 was removed due to the sensitivity of the BeadChip assay. RESULTS: After the scrubbing of data, 7,131 CpG sites were significantly different between males with stable CAD and males with CAD with MI ($p < 0.05$). There were 1,710 CpG sites that showed a higher beta value in CAD group compared to CAD with MI. The top ranked CpG site by p-value ($p = 2.38 \times 10^{-06}$; cg08053846) was located in the promoter of the gene SERPINE1 and showed 81 percent higher beta value in individuals with stable CAD. CONCLUSIONS: This work provides new insight into the biological differences between individuals with stable CAD and CAD with a myocardial infarction. Additionally, we can utilize epigenetic markers to discern novel pathways involved in MI disease etiology.

2256W

Evaluation of the Familiality of Cardiovascular Diseases among Patients in a Large Healthcare System. *S. Knight[1,2], B. Horne[1,2].* 1) Intermountain Heart Institute, Intermountain Medical Center, Murray, UT; 2) Division of Genetic Epidemiology, University of Utah,Salt Lake City, UT.

Objective: To determine the familiality of cardiovascular diseases(CVD) Methods: The Intermountain Genealogy Registry, which contains genealogies for 700,000 patients(pts) in the Intermountain Healthcare System, was used. One-way and two-way genealogy index of familiality (GIF) were generated. Results: The overall pts GIF (pGIF) was 0.443. The GIF for patent foramen ovale (PFO) was >40% and myocardial infarction (MI) was >10% larger than the pGIF. No two-way GIF were >10% larger than the pGIF. Conclusions: PFO and MI had a stronger familial component than other CVD. Narrowing phenotypes may increase the GIF.

GIF	MI	HF	CAD	HTN	PFO	Valve
MI	0.488	0.462	0.470	0.453	0.474	0.466
HF		0.471	0.457	0.446	0.453	0.458
CAD			0.476	0.450	0.465	0.460
HTN				0.461	0.451	0.447
PFO					0.628	0.459
Valve						0.476

.

2257T

Genome-wide linkage and positional association study of blood lipid phenotypes: The GenSalt study. *C. Li[1], J.E. Hixson[2], L.C. Shimmin[2], D.C. Rao[3], D. Gu[4], J. He[1], T.N. Kelly[1].* 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; 2) Department of Epidemiology, University of Texas School of Public Health, Houston, TX, USA; 3) Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA; 4) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

The genetic mechanisms underlying lipid levels remain largely unknown. We conducted a genome-wide linkage scan and positional association analysis of blood lipid phenotypes among Han Chinese participants of the Genetic Epidemiology Network of Salt Sensitivity (GenSalt) study. Overnight fasting blood samples were drawn by venipuncture at a baseline examination among 1,881 GenSalt participants from 633 families. Total cholesterol (TC), high density lipoprotein cholesterol (HDL-C) and triglycerides (TG) were analyzed enzymatically using commercially available reagents. Low density lipoprotein cholesterol (LDL-C) levels were calculated using the Friedewald equation. Lymphocyte DNA samples were used for genotyping microsatellite markers and single nucleotide polymorphisms (SNPs). Multipoint quantitative trait linkage-analysis was performed for each of the TC, HDL-C, LDL-C, and log-transformed TG phenotypes using SOLAR software. Additive associations between each SNP in identified linkage regions (LOD>2) and the lipid phenotypes were assessed using a mixed linear regression model to account for family dependencies. The false discovery rate (FDR) method was used to adjust for multiple testing. Phenotypes were adjusted for age, gender, body mass index and field center in all analyses. Suggestive linkage signals were identified at 4p15.1-4p14 [maximum multipoint LOD score (MML)=2.45 at 4p15.1] for the LDL-C phenotype. Follow-up association analyses in this region revealed a significant association between novel marker rs2995976 and LDL-C after adjustment for multiple testing (P= 3.73×10^{-5}; FDR-Q=0.03). Mean LDL-C levels for rs2995976 genotypes G/G, A/G, and A/A were 88.36 (82.56 to 94.17), 93.06 (90.62 to 95.50), and 97.69 (95.97 to 99.41) mg/dL, respectively. SNP rs2995976 lies upstream from *PGM2*, a gene indirectly involved in pathways of fatty acid production and energy generation. Furthermore, we observed suggestive linkage at 2p11.2-2q12.3 (MML=2.22 at 2q11.2) and 11q25 (MML=2.21 at 11q25) for the log-transformed TG phenotype. Follow-up analyses did not reveal any significant association signals in this region. In summary, genomic regions on chromosomes 2, 4 and 11 may harbor important susceptibility loci for lipid phenotypes. In addition, a novel variant upstream of the *PGM2* gene was significantly associated with LDL-C. Further study is needed to confirm these findings.

2258F

Genome-wide linkage and regional association study of blood pressure response to cold pressor test in Chinese: the GenSalt study. *X. Yang[1], T.N. Kelly[2], X. Wu[1], J.E. Hixson[3], J. Chen[1], J. Cao[1], J. Li[1], L.C. Shimmin[3], J. Huang[1], D.C. Rao[4], J. He[2], D. Gu[1].* 1) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 2) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA; 3) Department of Epidemiology, University of Texas School of Public Health, Houston, TX, USA; 4) Division of Biostatistics, Washington University School of Medicine, St Louis, MO, USA.

The cold pressor test (CPT) has been documented to evaluate cardiovascular reactivity to stress. However, genetic mechanisms underlying blood pressure (BP) response to CPT remain unclear. We performed a genome-wide linkage scan and regional association analysis to identify genetic determinants of BP response to CPT. A total of 1,998 Han Chinese participants from the Genetic Epidemiology Network of Salt Sensitivity (GenSalt) study completed the CPT. BP measurements were obtained before and after the participants immersed their hands in ice-water for 1 minute. BP response to the CPT was defined as BP at time zero after ice-water immersion minus the pre-test BP. Multipoint quantitative trait linkage analysis was performed using SOLAR software. The additive associations between single SNPs in linkage regions (LOD>2) were assessed using a mixed linear regression model to account for familial correlations. In addition, gene-based associations under linkage peaks were also examined using the truncated product method. Age, gender, field center, and body mass index were adjusted in all analyses. The results showed that suggestive linkage signals were identified for BP responses to CPT at 20p13-20p12.3, with maximum multipoint LOD scores of 2.38 for systolic BP (SBP) response, 0.13 for diastolic BP (DBP) response and 1.30 for mean arterial pressure (MAP) response. Marker rs2326373, located in the 5' flanking region of the *SMOX* gene at 20p13, was significantly associated with DBP and MAP responses to CPT (P= 4.49×10^{-5} and 8.52×10^{-6}, respectively). DBP responses (95% CI) for genotypes G/G, G/A, and A/A were 6.84 (6.44 to 7.23), 4.11 (2.83 to 5.40), and 4.22 (3.13 to 5.31) mmHg, respectively, and MAP responses (95% CI) were 8.97 (8.56 to 9.39), 6.48 (5.27 to 7.69), and 3.91 (2.80 to 5.02) mmHg, respectively. A similar trend was also observed for SBP response. In addition, results of gene-based analyses showed that variants in genes *MCM8* and *SLC23A2* were jointly associated with SBP response to CPT (P= 4.0×10^{-5} and 2.2×10^{-4}, respectively), and variants in genes *MCM8* and *STK35* were jointly associated with MAP response to CPT (P=2.0×10^{-5} and 2.9×10^{-4}, respectively). In summary, under a suggestive linkage region on chromosome 20, we identified a novel variant near *SMOX* which showed strong and consistent associations with SBP, DBP and MAP responses to CPT. In addition, we identified gene-based associations of *MCM8*, *SLC23A2* and *STK35* in this region.

2259W

Whole-genome expression profile of calcified bicuspid and tricuspid aortic valves. *S. Guauque-Olarte[1], N. Gaudreault[1], P. Pibarot[1], P. Mathieu[1], Y. Bossé[1,2].* 1) Centre de recherche Institut universitaire de cardiologie et de pneumologie de Québec, Quebec, Canada; 2) Department of Molecular Medicine, Laval University, Quebec, Canada.

BACKGROUND: Calcific aortic valve stenosis (AS) is a fatal disease with no medical treatment other than surgery. Bicuspid aortic valve (BAV) is present in 1-2% of the population. Patients with BAV have increased risk of AS and develop symptoms 10-15 years younger than patients with a tricuspid valve (TAV). The objective of this study was to identify genes differentially expressed between BAV and TAV with and without calcification. METHODS: Twelve calcified BAV and 12 calcified TAV were explanted from white male patients who underwent aortic valve replacement surgery. All valves had the same degree of fibro-calcific remodeling. Non-stenotic TAV (controls) were taken from 12 white males who underwent heart transplantation. The gene expression profile of each valve was measured with the Illumina HumanHT-12 v4 Expression BeadChip. Normalization and quality controls were performed with the lumi package in R. Gene expression levels were compared between the three groups of valves (BAV, TAV, and controls) using the Significance Analysis of Microarrays program. A false discovery rate of 5% and a fold change cut-off of 2 were used. Pathway analysis was performed using Ingenuity Pathway Analysis. The microarray results will be compared with RNA-Sequencing data from the same valves performed on the Illumina HiSeq2000. RESULTS: Two up-regulated and 2 down-regulated genes were identified in BAV compared to calcified TAV. For the comparison BAV vs controls, 128 genes were differentially expressed including 80 genes up-regulated in BAV. Compared to controls, 42 genes were up-regulated in calcified TAV, while 21 genes were down-regulated. The genes differentially expressed between BAV and calcified TAV were significantly linked to pathways and functions related to AS development and progression including connective tissue disorders (p= 2.09E-4 - 6.07E-4), inflammatory disease (p=2.09E-4 - 2.69E-2), and cell death and survival (p=3.17E-4 - 4.53E-2). CONCLUSIONS: The gene expression profiles of calcified BAV and TAV are highly similar. In contrast, aortic stenosis induced substantial changes in aortic valves gene expression. The results of this study increased our understanding of the molecular basis of BAV and associated calcific stenosis. Expected outcomes are new therapeutic targets to prevent, slow the development or treat AS in patients with BAV and calcified TAV.

2260T

Peripheral blood microRNA profiles are associated with cardiometabolic disease. *L.C. Kwee[1], W.E. Kraus[1], U. Hidefumi[2], H. Toyoshiba[2], T. Andou[2], E.R. Hauser[1], S.G. Gregory[1], J. Bain[1], M. Muehlbauer[1], R. Urquhart[2], C.B. Newgard[1], S.H. Shah[1].* 1) Duke Institute of Molecular Physiology, Duke University Medical Center, Durham, NC; 2) Pharmaceutical Research Division, Takeda Pharmaceutical Company, Japan.

MicroRNAs (miRNAs) are short, noncoding RNA molecules that affect gene expression and have been implicated in cardiometabolic diseases. Small-molecule metabolites also are associated with these phenotypes; however, miRNA and metabolite profiles have not been integrated in cardiometabolic disease studies. We examined the relationship between circulating miRNA profiles with metabolites and cardiometabolic phenotypes to identify miRNAs that may mediate cardiometabolic outcomes. Fasting plasma samples were obtained from 712 subjects selected from the CATHGEN biorepository of patients referred for cardiac catheterization at Duke University. MiRNA levels were assayed using TaqMan RT-qPCR arrays, and the most variable quantile of mean-normalized miRNAs was used for analysis. Lipids, glucose, ketones, free fatty acids, 15 amino acids and 45 acylcarnitines were assayed using biochemical methods or targeted mass spectrometry. Outcome phenotypes were clinical cardiometabolic phenotypes (BMI, blood pressure, diabetes, and severity of coronary artery disease) and 14 orthogonal factors identified from principal components analysis of the metabolites. Regression models were used to test the association of each miRNA with each phenotype, adjusting for age, sex and race. Sixty-six miRNAs were significantly associated with at least one phenotype ($p<5\times10^{-4}$, Bonferroni adjusted for multiple testing of 101 miRNAs). Most of these were associated with a factor composed primarily of free fatty acids, serine and histidine (53 miRNAs, lowest $p=10^{-22}$) or with triglyceride levels (41 miRNAs, lowest $p=10^{-12}$); 36 miRNAs were associated with both phenotypes. The predicted gene target for the most significant miRNA, miR-21, is *YOD1*, which is involved in the biosynthesis of unsaturated fatty acids (KEGG). Additionally, multiple miRNAs were associated with long-chain acylcarnitines (10 miRNAs, lowest $p=3.1\times10^{-6}$), short-chain dicarboxylacylcarnitines (6 miRNAs, lowest $p=1.7\times10^{-6}$), a factor composed primarily of asparagine/aspartic acid and glutamine/glutamic acid (5 miRNAs, lowest $p=4.2\times10^{-5}$), and glucose (4 miRNAs, lowest $p=3.2\times10^{-5}$). In one of the largest studies of miRNAs in human cardiometabolic disease, we have identified biologically plausible and novel phenotypes associated with peripheral blood miRNA levels. These miRNAs and their predicted targets are excellent candidates for further analysis as potential regulators of cardiometabolic phenotypes.

2261T

DYNAMICS OF HSPC SUBTYPES IN NON-HUMAN PRIMATES REVEALED BY A DECADE-LONG CLONAL TRACKING STUDY. *S. Kim[1], N. Kim[6], A. Presson[2], M.E. Metzger[5], G.M. Crooks[3], D. An[4], R.E. Donahue[5], I.S.Y. Chen[1].* 1) Microbiology, Immunology, and Molecular Genetics; 2) Biostatistics; 3) Pathology and Laboratory Medicine; 4) School of Nursing, University of California Los Angeles, Los Angeles, CA 90095; 5) National Heart, Lung and Blood Institute, NIH, Bethesda, MD 20892; 6) Korean Bioinformatics Center, Daejeon, South Korea.

Hematopoietic stem cell and progenitor cell (HSPC)-based genetic therapy to treat HIV/AIDS and other previously incurable diseases is becoming increasingly realistic. To date, however, our understanding of the regenerative potentials of primate HSPCs lags far behind the level desired for current and future uses of these cells for therapeutic purposes. Here, we report detailed behavior patterns of repopulating HSPCs in four rhesus macaques transplanted for up to 12 years with autologous CD34+ HSPCs engineered with control lentivirus vectors expressing a fluorescence marker (EGFP) or therapeutic vectors expressing an shRNA against the C-C chemokine receptor type 5 (CCR5-shRNA), an essential co-receptor for R5-tropic strains of HIV-1. All animals showed normal hematopoietic recovery and maintained stable EGFP marking in all tested blood lineages. In order to study in vivo HSPC behaviors at the single cell level, vector marked clones in serially collected blood were analyzed by large-scale vector integration site (VIS) sequencing and bioinformatics analysis of genomic VIS sequences. Our long-term clonal tracking study revealed novel mechanistic insights into complex polyclonal hematopoietic reconstitution in primates, not seen in traditional population-based studies. Analysis revealed thousands of HSPC clones successfully engrafted and expanded sequentially over time in each animal, clustered into groups with different kinetics of repopulation, maintaining the total marked blood cells at a relatively stable level over the years. Consistent with recent discoveries in murine HSC studies, the long-term repopulating clones were distinctively grouped into 'myeloid-biased', 'lymphoid-biased', and 'balanced' subtypes based on their unique lineage output potentials. Interestingly, clones with more balanced lineage potentials, accounted for only 4 - 10% of the identified clones, yet predominated, contributing up to 25 - 71 % of total repopulating cells in test animals. Expression of a therapeutic gene for HIV/AIDS (CCR5-shRNA) also showed no notable effects on HSPC behaviors. This study is the first system level description of the detailed in vivo behavior patterns of primate HSPCs, providing new insights into primate repopulation. Our data also provide a potential frame of reference for future HSPC-based gene-therapies.

2262F

Long term outcome and success of liver transplantation in patients with progressive familial intrahepatic cholestasis: is there an association between genotype and outcome? *S.M. Herbst[1], J. Vermehren[2], M. Melter[2], U. Hehr[1].* 1) Center for Human Genetics, University of Regensburg, Regensburg, Germany; 2) Children's Hospital , University Hospital Regensburg, Regensburg, Germany.

Study and Objective: Infants with progressive familial intrahepatic cholestasis(PFIC) develop liver failure in the first or second decade and ultimately require liver transplantation. Three causal genes necessary for bile flow across the canalicular membrane have so far been identified: *ATP8B*1, *ABCB4* and *ABCB11*. Our aim was to further characterize the genotype phenotype relationship and to search for prognostic genetic parameters for long term outcome as well as the success of liver transplantation in order to improve individual treatment options for PFIC patients. **Method:** A total of 52 PFIC index patients were analyzed by direct sequencing between 2008 and 2012 for mutations in *ATP8B1, ABCB4* and *ABCB11*. Extensive long term follow-up data (follow-up range 2-35 years) of 8 patients with genetically confirmed PFIC was retrospectively analyzed to illustrate the specific problems in interdisciplinary medical care. **Results:** Most pathogenic mutations in PFIC patients were identified in *ABCB11* (56%), followed by *ATP8B1* (32%) and *ABCB4* (12%). Overall the most frequently detected type of mutations were missense mutations (73%), followed by truncating mutations (17%), splice mutations (7%) and small deletions (3%). 9 novel mutations were identified. Response to pharmacological therapy was insufficient in 87% of patients, thus, 62% of patients underwent biliary diversion. Subsequently, liver transplantation was performed in 5 out of 8 patients; one additional patient is currently listed for LTX at the age of 17 months. Liver transplantation was complicated by: bleeding and coagulation problems, severe rejection requiring retransplantation (1) and death (1). Interestingly, both patients with life threatening complications were carrier of a truncating *ABCB11* mutation. After successful liver transplantation the most severe long term complications were observed in *ATP8B1* mutations carriers and included: severe chronic diarrhea (>12x/d), renal failure (1 requiring renal transplant), missing catch-up growth, pancreatitis, deafness and polyneuropathy. **Conclusion:** Truncating mutations may be associated with a higher rate of complications in liver transplantation, possibly due to a stronger immune reaction towards the wild type protein. With an ongoing interdisciplinary study we address two major objectives (1) evaluating this hypothesis in a larger cohort and (2) using whole genome sequencing to search for further PFIC candidate genes in mutation negative patients.

2263T

First U.S. Orthotopic Liver Transplantation for Intractable Acute Intermittent Porphyria. *A. Ludtke[1], M. Balwani[1], L.U. Liu[2], H. Naik[1], M. Yasuda[1], A. Arvelakis[3], C. Yu[1], S.S. Florman[3], R.J. Desnick[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Division of Liver Diseases, Department of Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Recanati/ Miller Transplantation Institute, Icahn School of Medicine at Mount Sinai, New York, NY.

Acute Intermittent Porphyria (AIP) is a rare autosomal dominant inborn error of heme biosynthesis due to the half normal activity of hydroxymethylbilane synthase. The disease is characterized by life-threatening acute neurovisceral attacks which are triggered by certain drugs, dieting, and hormonal factors that markedly increase the expression of 5-aminolevulinate synthase 1, resulting in the increased hepatic production of the neurotoxic metabolites aminolevulinic acid (ALA) and porphobilinogen (PBG). Patients may be asymptomatic, experience sporadic attacks, or may have frequent attacks, which are treated with intravenous hemin; recurrent attacks may be difficult to control even with weekly hemin infusions. Since the liver is the major source of the overproduced ALA and PBG, orthotopic liver transplantation has led to complete clinical and biochemical correction in several patients who failed medical management in Europe. Here, we report the first US liver transplantation in a 36 year old Caucasian woman with intractable AIP, who was diagnosed at age 27 (HMBS mutation R173W). She had recurrent acute attacks, which became refractory to hemin and developed complications such as generalized paresis, chronic axonal polyneuropathy, multiple port infections and had prolonged hospitalizations. She became increasingly disabled, depressed and required very high doses of opiates for pain control. After 13 months of hospitalization, orthotopic liver transplantation was undertaken. Pre-transplant urinary ALA and PBG concentrations ranged from 4.0-17.6 mmol/mol creat (norm <3.0) and 14.2-75.6 mmol/mol creat (norm <1.1), respectively; plasma ALA and PBG levels were 2.0-2.7 uM (<0.2) and 4.7-5.9 uM (<0.05). Post transplant, urinary and plasma ALA and PBG concentrations both normalized by 5 and 23 hours, respectively. The chronic AIP-related pain resolved 4 days after surgery. Histological studies of the explanted liver revealed steatosis and cirrhosis due to iron overload secondary to the chronic hemin therapy. These results demonstrate that liver transplantation can correct the metabolic defect with dramatic biochemical and clinical improvement. We conclude that liver transplantation with its inherent morbidity and mortality should be considered in severely affected patients with acute hepatic porphyrias that are refractory to current medical management.

† Trainee Award Finalist

2264F

Creation of a new mouse model for the mRNA splicing disease Familial Dysautonomia.
E. Morini[1], P. Dietrich[2], I. Dragatsis[2], M. Salani[1], F. Urbina[1], S.A. Slaugenhaupt[1]. 1) Center for Human Genetic Research, Massachusetts General Hospital/Harvard Medical School, Boston, MA; 2) Department of Physiology, The University of Tennessee, Health Science Center, Memphis, TN.

Recent studies emphasize the importance of mRNA splicing mutations in genetic disease, with some estimates suggesting that around 30% of point mutations disrupt mRNA splicing. Familial dysautonomia (FD) is an excellent disease model for studying new strategies to correct splicing defects. FD is a recessive neurodegenerative disease caused by a splice mutation in the IKBKAP gene which leads to variable skipping of exon 20. We found that the small molecule kinetin can correct the IKBKAP splicing defect and increase the amount of normal mRNA and protein in FD cell lines. We have also shown that kinetin can increase the level of functional IKAP protein in mice following oral dosing in all tissues tested, including brain. Despite these remarkable advances, including our demonstration of in vivo efficacy to increase normal IKBKAP mRNA in the blood of FD patients, we lacked animal model in which to test the effect of targeting mRNA splicing to increase IKAP protein on FD phenotype. In order to create a phenotypic model of FD in which we could also manipulate mRNA splicing we introduced an FD transgene (TgFD9), which contains the human IKBKAP gene with the major FD splice mutation, into the Ikbkap Δ20/flox mouse model by sequential mating. The introduction of the human IKBKAP transgene attenuates the severe FD phenotype that we observed in the Ikbkap Δ20/flox mouse and recreates the same tissue-specific mis-splicing defect. TgFD9/IkbkapΔ20/flox mice show a reduced growth rate but, unlike Ikbkap Δ20/flox mice which die perinatally, most FD9/Ikbkap Δ20/flox mice survive postnatally. Methylene blue staining of tongues shows that FD9/Ikbkap Δ20/flox mice have a reduction of fungiform papillae. Since several FD-like features are present at birth, we analyzed body weight, superior sympathetic ganglia (SCG), stellate ganglia (SG) and cervical DRGs in E18.5 FD9/Ikbkap Δ20/flox embryos. Already at this stage the FD9/Ikbkap Δ20/flox embryos accurately model many features of the human disease. Furthermore, we characterized the mechanism underlying the low levels of endogenous mouse Ikbkap expression in FD9/Ikbkap Δ20/flox mice. Our results demonstrate that the new TgFD9 Ikbkap Δ20/flox mouse accurately models both the disease phenotype and the tissue-specific mRNA mis-splicing defect seen in FD patients. The creation of this new model has allowed us to initiate a detailed clinical trial of kinetin and will permit testing of other strategies for targeting mRNA splicing.

2265T

Phosphatidylserine: A potential gene modifying therapy for Familial Dysautonomia?
M. Salani[1], L. Norcliffe-Kaufmann[2], J. Martinez[2], E. Morini[1], F. Axelrod[2], S. Slaugenhaupt[1]. 1) Center for Human Genetic Research,-Massachusetts General Hospital, Harvard Medical School, Boston, MA; 2) Dysautonomia Center, New York University School of Medicine, New York, NY.

Familial dysautonomia (FD) is caused by a splicing error in the IKBKAP gene that encodes human Elongator protein 1. In these patients, exon 20 is frequently skipped during mRNA splicing, but cells retain the ability to produce a low level of normal (wild-type) IKBKAP mRNA and normal IKAP protein. Phosphatidylserine (PS, Sharp-thought®), an acidic phospholipid, has been shown to raise IKAP levels in fibroblast cell lines derived from FD patients by increasing IKBKAP transcription(1).Given that PS is safe and available over the counter we conducted a preliminary study to determine if PS increases IKBKAP expression in patients with FD. We enrolled 6 patients with FD, 16 to 23 years old, in an open-label titration protocol. Patients were examined at baseline (visit 1), after 2 months of taking 300 mg/day (visit 2) and again after 2 months of taking 600 mg/day of PS (visit 3). Blood was taken at each visit and de-identified and investigators blinded to sample identity. Blood was treated with Tri-Reagent, and RNA extracted according to manufacturer's specifications. Quantitative PCR was performed to measure the level of normal IKBKAP mRNA. PS was well tolerated and there were no adverse events or unexpected clinical abnormalities. No increase in IKBKAP was observed after two months of treatment with 300 mg/day, however, we did see an increase in four of six patients after an additional two months of 600mg/day PS. Our preliminary results indicate that PS might safely raise normal IKBKAP mRNA levels in blood from patients with FD, opening an exciting potential therapeutic path for the treatment of FD. A more extensive analysis using a larger number of patients and an increased dose of PS is underway. 1 Keren et al., PLoS One. 2010 Dec 29,5 (12): e15884.

2266F

Inhibition of retinoic acid signaling rescues inner-ear defects in a mouse model of CHARGE syndrome.
J.M. Skidmore[1], E.A. Hurd[2], A. Saiakhova[3], D.L. Swiderski[4], E.D. Sperry[5], P.S. Scacheri[3], Y. Raphael[5], D.M. Martin[1,5,6]. 1) Department of Pediatrics, University of Michigan, Ann Arbor, MI; 2) Centre for Neuroregeneration, University of Edinburgh, Edinburgh, Scotland; 3) Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH; 4) Department of Otolaryngology, University of Michigan, Ann Arbor, MI; 5) Medical Scientist Training Program, University of Michigan, Ann Arbor, MI; 6) Department of of Human Genetics, University of Michigan, Ann Arbor, MI.

CHARGE syndrome, a multiple congenital anomaly disorder, and *in utero* vitamin A deficiency or toxicity share many common developmental features including craniofacial defects, vision and hearing deficits, cardiac anomalies, and growth delays. CHARGE syndrome is caused by mutations in *CHD7*, encoding an ATP-dependent chromatin remodeling protein that regulates downstream target gene expression. Here we tested, using heterozygous loss of function *Chd7* mutant mice ($Chd7^{Gt/+}$), whether *Chd7* deficiency affects retinoic acid signaling and if altered retinoic acid signaling in utero influences *Chd7* mutant phenotypes. We focused specifically on the inner ear, due to the highly penetrant semicircular canal defects observed in both humans and mice with CHD7 deficiency. RNA-seq of microdissected e10.5 inner ear tissues revealed 50% or greater upregulation of 425 genes and downregulation of 146 genes in $Chd7^{Gt/+}$ compared to wild type mice. Several genes emerged as potential targets of CHD7, including genes encoding retinoic acid synthetic enzymes (*Aldh*), degradation enzymes (*Cyp26*), and receptors (*Rar, Rxr*). To test whether retinoic acid signaling during pregnancy influences *Chd7* deficiency phenotypes, we administered retinoic acid or citral (a retinoic acid inhibitor) to pregnant mice at e7.5. $Chd7^{Gt/+}$ and wild type embryos were dissected at e14.5 and inner ear structures analyzed after paint-filling. Retinoic acid treatment worsened, or had no effect on, $Chd7^{Gt/+}$ inner ear defects, whereas citral treatment partially corrected the typically fully penetrant semicircular canal dysgenesis. Levels of retinoic acid were monitored in *Chd7* mutant vs. wild type embryos using RARE (retinoic acid response element) reporter mice, which express β-galactosidase in response to retinoic acid. $Chd7^{Gt/+}$ embryos exhibited mildly increased β-galactosidase staining in the developing cochleovestibular ganglion, consistent with enhanced RA signaling. Aberrant retinoic acid signaling and gene expression with loss of *Chd7*, together with partial rescue of *Chd7* deficient phenotypes upon retinoic acid inhibition support an essential role for vitamin A metabolism in *Chd7* deficiency states. These studies have implications for understanding the pathogenesis of phenotypes in CHARGE syndrome, including variable expressivity and reduced penetrance of organ specific defects, and for enhancing ongoing efforts to develop effective therapies for CHARGE individuals.

2267T

Myhre syndrome-causing SMAD4 mutations result in disorganization of extracellular matrix that is corrected by losartan treatment.
P. Piccolo[1], P. Mithbaokar[1], V. Sabatino[1], J. John Tolmie[2], D. Melis[3], M.C. Schiaffino[4], M. Filocamo[5], G. Andria[3], N. Brunetti-Pierri[1,3]. 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Ferguson-Smith Department of Clinical Genetics, Yorkhill Hospital, Glasgow, UK; 3) Department of Translational Medicine, Federico II University of Naples, Naples, Italy; 4) Clinica Pediatrica, Istituto G. Gaslini, Genova, Italy; 5) Centro di Diagnostica Genetica e Biochimica delle Malattie Metaboliche, Istituto G. Gaslini, Genova, Italy.

Myhre syndrome (MIM 139210) is an autosomal dominant connective tissue disorder that presents with short stature, short hands and feet, facial dysmorphic features, muscle hypertrophy, thickened skin, and deafness. Myhre syndrome is caused by recurrent missense mutations affecting codon Ile500 of SMAD4 encoding a transducer mediating transforming growth factor β (TGF-β) signaling. We investigated the functional consequences of SMAD4 mutations in Myhre syndrome fibroblasts and observed altered expression of genes encoding matrix metalloproteinases and related inhibitors and a defect of extracellular matrix deposition. Disruption of microfibril network results in increased TGF-β bioavailability. Increased TGF-β signaling and progression of aortic root dilation in Marfan syndrome, a common connective tissue disorder, can be prevented by losartan, a TGF-β antagonists and angiotensin II type 1 receptor (AT1) blocker. In this study, we investigated whether losartan is effective at improving the extracellular matrix deposition defect of Myhre syndrome cells. We showed that losartan normalizes SMAD2 phosphorylation, restores balance of metalloproteinases and related inhibitors, and improves the extracellular matrix deposition in fibroblasts from Myhre syndrome patients. The results of this study may pave the way towards therapeutic applications of losartan in Myhre syndrome.

2268F

Tumor associated macrophages in neurofibromatosis. *C. Prada[1,3], E. Jousma[2], T. Rizvi[2], J. Wu[2], S. Dunn[2], N. Ratner[2].* 1) Dept Pediatrics Genetics, Cincinnati Children's Hosp Med Ctr, Cincinnati, OH; 2) Divisions of Experimental Hematology and Cancer Biology, Cincinnati Children's Hosp Med Ctr, Cincinnati, OH; 3) Center for Genomic Medicine and Metabolism, Cardiovascular Foundation of Colombia, Floridablanca, Colombia.

Background: Plexiform neurofibromas (PNFs) are one of the most common and debilitating complications of Neurofibromatosis type 1 (NF1). In animal models, Nf1-/- Schwann cells secrete large amounts of chemoattractants (RANTES, SCF, M-CSF, and VEGF), enhancing migration of mast cells toward the peripheral nerve. Macrophages can be recruited by several cytokines including RANTES and VEGF. We previously identifed macrophages infiltrating mouse PNFs. Macrophage can have anti-tumor (M1) or pro-tumor (M2) role at different cancer stages. We hypothesized that macrophages infiltration correlates with tumor progression and macrophage depletion induces cell death in Schwann cells. Methods: We tested this hypothesis in the Dhh-Cre; Nf1flox/flox mouse model of neurofibroma and human PNFs and MPNSTs. Immunohistochemistry was performed to quantify macrophage infiltration and macrophage subtype (M1 and M2 markers) in peripheral nerves and tumors at different disease stages to evaluate if increasing numbers of Iba1+CD11b+ cells correlates with progression (n= 20). Then, we treated 7 Dhh-Cre; Nf1flox/flox mice with 50mg/kg of minocycline, a broad-spectrum antibiotic that partially depletes macrophages, by i.p. injection and 7 Dhh-Cre; Nf1flox/flox mice with vehicle control, both for 1 month, to evaluate if macrophage depletion leads to cell death in tumors. Results: Large numbers of macrophages stained with Iba1 were seen in the peripheral nerve and PNFs from Dhh-Cre; Nf1flox/flox mouse model and PNFs from human samples. Macrophage accumulation was highest in the MPNSTs of mice and humans (p<0.001). These findings show that macrophage recruitment correlates with presence of tumor and correlates with disease progression (MPNSTs>PNFs>Nf1 mutant nerve>normal nerve). Staining with anti-CD163 an M2-marker showed that MPNST have 15% of Iba1+CD163+ cells. The percentage of macrophages were similar in the cell-rich and cell-poor regions of PNFs. Flow sorting from human neurofibromas showed an average of 8% (range 1.5 to 14%) of CD11b+ cells in the PNFs. Minocycline treated tumors have decreased Iba1+ macrophages and increased Schwann cell death supporting a role for macrophages on Schwann cell survival in PNFs. Conclusion: Macrophages are a major component of NF1 peripheral nerve tumors in human and mouse with a direct effect on tumor growth and survival. Accumulation of macrophages correlates with tumor progression and depletion may become a therapeutic strategy for PNFs.

2269T

OLIGOTHERAPEUTIC STRATEGIES FOR THE TREATMENT OF FRIEDREICH'S ATAXIA. *F. Ozsolak, D. Jun Li, D. Parekh, D. Knowlton, M. Wysk, R. Subramanian, J. Barsoum.* RaNA Therapeutics, Cambridge, MA.

Friedreich's ataxia (FRDA) is a recessively inherited disorder that arises due to cellular depletion of frataxin (FXN) protein and resulting defects in mitochondrial functions. FRDA is a progressive neuromuscular disease which lacks any FDA-approved therapy. The protein coding sequence of FXN is normal in the majority of FRDA patients, and the causative basis of this disease is the under-expression of the FXN gene, suggesting that upregulation of endogenous FXN expression could be an effective therapy. The most common molecular cause of this currently incurable disease is the expansion of GAA/TTC triplet repeats in the first intron of FXN gene. Repeat expansion beyond a certain threshold causes transcriptional defects which reduce FXN mRNA and protein levels. Despite long-standing research in the pathogenesis of FRDA, the means by which GAA-repeat number elevation leads to transcriptional silencing is not clear. Unusual DNA-DNA and DNA-RNA interactions formed in the long triplet repeat stretches, defects and alterations in splicing patterns and the formation of a heterochromatin-like structure are among the hypotheses being considered. In order to gain clues into the mechanisms responsible for the FXN deficit in FRDA, we undertook genome-wide analyses to examine the global and local RNA structure and chromatin structure and composition changes in FRDA patient cells. Epigenetic screens identified two chromatin modifying complexes as being important in establishing and/or maintaining repeat expansion-induced transcriptional repression at the FXN locus. We identified a novel non-coding RNA (ncRNA) potentially responsible for directing the localized epigenetic silencing of the FXN gene. Degrading this ncRNA led to at least partial heterochromatin reversal and FXN mRNA and protein level upregulation to therapeutically significant levels. The oligonucleotide-based therapeutic approaches developed here pave the way towards the design of multiple strategies for the treatment of FRDA and may have applications for the treatment of other human diseases. You may contact the first author (during and after the meeting) at fozsolak@ranarx.com.

2270F

Positive Effects of Short Course Androgen Therapy on the Neurodevelopmental Outcome In Boys with 47, XXY Syndrome at 9 Years of Age. *C.A. Samango-Sprouse[1,3], E. Stapleton[3], C. Sprouse[1,2], T. Sadeghin[3], F.L. Mitchell[3], A.L. Gropman[2].* 1) George Washington University, Washington, DC; 2) Department of Neurology, Children's National Medical Center, Washington, DC; 3) Neurodevelopmental Diagnostic Center for Young Children, Davidsonville, MD.

Background: Positive effects of early androgen treatment on neurodevelopmental performance in prepubertal males with 47, XXY have been documented at 36 and 72 months, giving support to the link between neurobiological treatment and neurodevelopmental outcome. Two previous studies have supported improvement in specific domains of neurodevelopmental performance with early androgen treatment.

Purpose: The aim is to determine if an early course of androgen treatment (3 injections of testosterone enanthate, 25mg, each) could have a positive impact on XXY boys at 9 years of age.

Methods: 59 prenatally diagnosed males with karyotypes of 47, XXY participated with one group (n=22) receiving androgen treatment in infancy and the second group untreated (n=37).

Results: There was a significant positive treatment effect in multiple visual motor domains (VP p=0.0012, MC p=0.0129, VMI p=0.0151). A positive treatment effect was observed on the BOT (Manual Coordination p=0.0003, Bilateral Coordination p=0.0001, Body Coordination p=0.0038, Speed/Agility p=0.0328, Strength p=0.0161 Upper Limb Coordination p=0.0004, and Strength/Agility p=0.0134).

Discussion: Long-term improved function has been observed in neurodevelopmental performance in XXY males at 36 and 72 months and now at 108 months, when treated with a short course of androgen in infancy. An early course of hormonal replacement may have a permanent positive effect on the neurodevelopment outcome in XXY in brain regions with known androgen receptors. These findings suggest early hormonal replacement in XXY males alters selected and vulnerable areas of neurodevelopmental performance. This is the third study demonstrating the positive effect of androgen on developmental performance further supporting the neurobiological link androgens, brain function and neurodevelopmental performance in XXY. Our findings provide additional support for the importance of early detection and treatment of boys with XXY.

2271T

Prenatal therapy in developmental disorders: drug targeting via intra-amniotic injection to treat X-linked hypohidrotic ectodermal dysplasia. *K. Hermes[1], P. Schneider[2], P. Kreig[3], A. Dang[2], K. Huttner[4], H. Schneider[1].* 1) German Competence Centre for Children with Ectodermal Dysplasias, Department of Pediatrics, University of Erlangen-Nürnberg, Germany; 2) Department of Biochemistry, University of Lausanne, Switzerland; 3) German Cancer Research Center, Heidelberg, Germany; 4) Edimer Pharmaceuticals, Inc., Cambridge, USA.

Background: X-linked hypohidrotic ectodermal dysplasia (XLHED), the most common inherited disorder of ectoderm development, is caused by a lack of the signaling molecule EDA-A1. In the Tabby XLHED mouse model, prenatal EDA-A1 replacement via maternal injection corrected the developmental abnormalities to a far greater extent than postnatal administration to newborn pups. This approach, however, may not be optimal for achieving therapeutic levels of corrective protein in the human fetus, and additionally exposes the mother to high serum levels of the exogenous molecule. We hypothesized that direct injection of EDA-A1 replacement protein into the amniotic fluid (AF) could result in fetal uptake via lung and gut, leading to sustained drug exposure at levels sufficient for successful treatment of XLHED. **Methods**: EDI200, a human IgG1:EDA-A1 fusion protein, was tested for stability in AF using a binding ELISA. Subsequently, EDI200 was injected into amniotic sacs of pregnant wild-type mice to evaluate drug uptake and pharmacokinetics. Fetal and maternal serum levels were monitored. Based on these results, EDI200 at a dose of 100 mg/kg fetal weight was administered intra-amniotically to E15 Tabby mouse fetuses and phenotypic correction was assessed. **Results**: EDI200 was demonstrated to be stable in AF at 37° for one week without detectable loss of activity. Intra-amniotic administration to E15 wild-type mice (35 mcg EDI200/fetus) resulted in substantial fetal uptake with mean serum levels of 9.0 mcg/ml and 1.2 mcg/ml at 6 hours and 96 hours, respectively. Maternal serum levels remained <0.1mcg/ml. In Tabby mice, a single maternal injection of EDI200 at E15 proved to correct the XLHED phenotype in offspring only partially. The complete spectrum of hair, sweat gland and dentition response following E15 intra-amniotic administration of EDI200 is being evaluated through adulthood and will be presented along with statistics on fetal loss related to this therapeutic approach. **Conclusions**: Intra-amniotic protein application to mice may lead to rapid fetal uptake, sustained substantial serum levels, and efficacy comparable with intravenous injection. As amniocentesis is generally a low-risk procedure in humans, this route of delivery may be suitable for drug targeting with reduced drug exposure of the mother and longer bioavailability due to a reservoir function of AF. It may, thus, represent a novel paradigm for treatment of disorders in early human development.

2272F

A Novel, Selective and Orally-available Glucosylceramide Synthase Inhibitor for Substrate Reduction Therapy of Fabry Disease. *J. Marshall[1], K. Ashe[1], E. Budman[1], D. Bangari[1], J. Nietupski[1], R.J. Desnick[2], R.K. Scheule[1], J.P. Leonard[1], S.H. Cheng[1].* 1) Genzyme, Framingham, MA; 2) Mount Sinai School of Medicine, New York, NY.

Fabry disease, an X-linked glycosphingolipid storage disorder, is caused by a deficiency of acid α-galactosidase A (αgal). Resulting progressive accumulation of globotriaosylceramide (GL-3) and lyso-GL-3 leads to kidney, heart, and cerebrovascular disease. Presently, Fabry disease is managed by periodic infusions of recombinant αgal (enzyme-replacement therapy; ERT). However, the inability of ERT to completely address disease manifestations in the heart and kidney has encouraged the development of alternative therapies. We had previously demonstrated that substrate reduction therapy (SRT) through antagonism of glucosylceramide synthase (GCS) can delay the accumulation of GL-3 in a mouse model of Fabry disease. Here, we describe the merits of a novel GCS inhibitor (Genz-682452) with favorable pharmacological properties and safety profile in Fabry mice. Treatment of Fabry mice with Genz-682452 starting at 3 months of age resulted in greater correction of a variety of disease biomarkers than when treatment was initiated in 12 months-old mice. Mice administered Genz-682452 exhibited significantly lower tissue levels of GL-3 and lyso-GL-3 and a delayed progression of a thermal nociceptive response than their untreated counterparts. SRT with Genz-682452, perhaps because it has a different biodistribution profile, was more effective than ERT at reducing the levels of the glycosphingolipids in the kidney, heart and CNS, organs that were not well-served by ERT. Importantly, mice treated by both ERT and SRT showed the greatest response suggesting the therapies are both complementary and additive. These results affirm the potential of SRT as an alternative and potentially adjuvant therapy for Fabry disease.

2273T

Successful pregnancy and lactation in a woman with mucopolysaccharidosis type I treated with laronidase. *Y. Xue[1], M. Castorina[2], D. Antuzzi[3], C. Sung[1], S. Richards[1], G. Cox[1].* 1) Genzyme, a Sanofi company, Cambridge, MA, USA; 2) Dipartimento di Tutela della Donna e della Vita Nascente, Pediatria; Universita Cattolica del Sacro Cuore, Rome, Italy; 3) Laboratorio di Neonatologia, Dipartimento di Tutela della Donna e della Vita Nascente, Pediatria; Universita Cattolica del Sacro Cuore, Rome, Italy.

Background: Few pregnancies have been reported in women with MPS I, a lysosomal storage disorder caused by α-L-iduronidase deficiency and subsequent accumulation of glycosaminglycans (GAG) throughout the body. With the availability of laronidase enzyme replacement therapy (Aldurazyme®, Genzyme, a Sanofi company, Cambridge MA, USA) more women with MPS I may become pregnant and want to breastfeed their infants. Although in animals, laronidase has had no effect on fertility or pregnancy, the lack of human data has led to the recommendation that laronidase be used in pregnancy only if needed and to be used with caution during breastfeeding. **Methods**: We describe an ongoing, prospective, open-label trial funded by Genzyme (ALID 01803, NCT00418821) of women with MPS I who plan to receive laronidase during pregnancy and while breastfeeding. Following local ethics committee approval and signed informed consent, the first patient has completed the study and the results are presented below. The effects of laronidase on the mother during pregnancy and on her infant for 12 months after delivery were assessed periodically through physical examinations, IgG and IgM anti-laronidase antibody titers, urinary GAG, laronidase activity in breast milk, and developmental testing in the infant. Adverse events and concomitant medications were evaluated continuously. **Results**: The patient, diagnosed with MPS I Scheie, had been treated with laronidase for 3 years, enrolled in the trial during her second pregnancy and continued to receive laronidase throughout her pregnancy and while breastfeeding. A healthy 2.5 kg boy was delivered by elective cesarean section at 37 weeks gestation. The baby was breastfed for 3 months. Laronidase was not detected in breast milk. Anti-laronidase IgG antibodies were present in the mother and in the umbilical cord blood at birth. However, the infant had no detectable anti-laronidase IgM antibodies, and the IgG titer declined over time, consistent with passive maternal transmission of antibodies. Urinary GAG levels in the infant were always normal. The infant was healthy and had normal development through 12 months. No drug-related adverse events occurred. **Conclusions**: We report the first known case of a successful pregnancy outcome and normal development in an infant whose mother continued laronidase treatment during pregnancy and lactation. Further data are necessary to confirm the safety of laronidase during pregnancy and breastfeeding.

2274F

ENCORE: a randomized, controlled, open-label non-inferiority study comparing eliglustat to imiglucerase in Gaucher disease type 1 patients on enzyme replacement therapy who have reached therapeutic goals. *M. Balwani[1], T.M. Cox[2], G. Drelichman[3], R. Cravo[4], T. Burrow[5], A.M. Martins[6], E. Lukina[7], B. Rosenbloom[8], L. Ross[9], J. Angell[9], A.C. Puga[9].* 1) Dept Human Gen, Mt Sinai Med Ctr-New York, New York, NY; 2) University of Cambridge, Department of Medicine, Addenbrooke's Hospital, Cambridge, UK; 3) Hospital de Niños Ricardo Gutiérrez, Buenos Aires, Argentina; 4) HEMORIO, Rio de Janeiro, RJ, Brasil; 5) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 6) Universidade Federal de São Paulo, São Paulo, SP, Brasil; 7) Hematology Research Center, Moscow, Russia; 8) Tower Hematology Oncology Medical Group, Beverly Hills, CA, USA; 9) Genzyme, a Sanofi company, Cambridge MA, USA.

Background: Gaucher disease type 1 is a multi-systemic lysosomal storage disorder caused by acid β-glucosidase deficiency. Intravenously administered enzyme replacement therapy has been the standard of care, however this significantly impacts quality of life. Eliglustat is a novel oral substrate-reduction therapy in development for adults with Gaucher disease type 1. This Phase 3 trial (ENCORE, NCT00943111, sponsored by Genzyme, a Sanofi company) compares eliglustat with imiglucerase in patients who had reached therapeutic goals on enzyme replacement therapy. Methods: Randomized (2:1 eliglustat: imiglucerase), controlled, open-label non-inferiority trial of 160 patients previously receiving enzyme replacement therapy for ≥ 3 years who had met therapeutic goals. The primary efficacy endpoint was percent of patients remaining stable after 52 weeks (using a pre-specified composite of spleen, liver, hemoglobin, and platelet parameters). Results: Eliglustat was statistically non-inferior to imiglucerase: 84% of eliglustat and 94% of imiglucerase patients maintained goals for all four parameters (lower bound of 95%CI of difference [-18.6%] within pre-specified non-inferiority thresholds). Individually, 94% of eliglustat patients maintained stability criteria for spleen, 95% for hemoglobin, 93% for liver, and 96% for platelets. Of the 16/99 eliglustat patients who did not meet stability criteria based on change from baseline, most patients (13/16) nevertheless had absolute values for the individual endpoints that remained within therapeutic goals. Baseline bone mineral density scores were normal in most patients and remained normal on eliglustat. Two eliglustat patients and one imiglucerase patient (2% in each arm) discontinued because of an adverse event. Over the course of 52 weeks, four adverse events were observed with ≥ 10% incidence compared with imiglucerase: fatigue (14% vs. 2%) headache (13% vs. 2%), nausea (12% vs. 0%), and upper abdominal pain (10% vs. 0%). The majority of adverse events were mild or moderate in severity for both groups. There were no treatment-related serious adverse events. Conclusions: In the Phase 3 ENCORE study, eliglustat was well tolerated and was non-inferior to imiglucerase in maintaining stability for 52 weeks in previously treated patients with Gaucher disease type 1.

2275T

The adverse event profile of eliglustat for the treatment of Gaucher disease type 1: results from a pooled analysis of four trials. *R. Mankoski, M.J. Peterschmitt, A.C. Puga, G.F. Cox, S. Marulkar, J. Angell, B. Gaemers, L. Ross.* Genzyme, a Sanofi company, Cambridge, MA.

Background: Gaucher disease type 1 is a multi-systemic lysosomal storage disorder resulting from glucocerebrosidase deficiency. Eliglustat is a novel oral substrate-reduction therapy in development for adults with Gaucher type 1. To better understand the adverse event profile of eliglustat, we evaluated pooled data from all Gaucher type 1 patients receiving eliglustat in on-going Phase 2 and Phase 3 trials. Methods: Adverse event data from four Genzyme-sponsored trials were pooled: an open-label Phase 2 trial (NCT00358150), two Phase 3 trials for which the primary analysis period is complete (ENGAGE, NCT00891202, a randomized, double-blind, placebo-controlled trial, and ENCORE, NCT00943111 , a randomized, controlled, open-label non-inferiority [to imiglucerase] trial of patients previously receiving ERT for ≥ 3 years who had met therapeutic goals), and an open-label lead-in period from the on-going phase 3 EDGE trial (NCT01074944) randomized, double-blind dosing trial). Results: A total of 393 patients received eliglustat with a mean (SD) duration of exposure of 1.4 (1.2) years and a total pooled duration of 535 patient-years. The majority of AEs had an onset within the first 6 months of treatment. Six adverse events occurred in more than 10% of patients: headache (17% overall; 12% mild, 3% moderate, 1% severe), arthralgia (14% overall; 8% mild, 5% moderate, 1% severe), nasopharyngitis (13% overall; 11% mild, 2% moderate, none severe), upper respiratory infection (11% overall; 9% mild, 2% moderate, none severe), diarrhea (10% overall; 9% mild, 1% moderate, none severe), and dizziness (10% overall; 8% mild, 2% moderate, none severe). Most AEs (81%) were considered to be unrelated to treatment by the treating physician. Across all studies, 42 serious adverse events (SAEs) were observed in 35 patients (9%). All SAEs were considered to be unrelated to eliglustat except for three that were possible related, two (in the same patient) considered to be probably related, and one considered to be definitely related. Thirteen patients (3%) had AEs that led to study withdrawal. Conclusions: The clinical development program for eliglustat is the largest for Gaucher disease type 1, and the pooled adverse event profile in this safety analysis demonstrates that eliglustat was well-tolerated and safe.

2276F

ENGAGE: A Phase 3, randomized, double-blind, placebo-controlled, multi-center study to investigate the efficacy and safety of eliglustat in adults with Gaucher disease type 1 (GD1): 9 month results. *S. Packman[1], D. Amato[2], M. Dasouki[3], G. Pastores[4], S. Assouline[5], M. Balwani[6], P.K. Mistry[7], S. Shankar[8], M.J. Peterschmitt[9].* 1) Dept Peds/Div Med Gen, Univ California Sch Med, San Francisco, CA; 2) Mount Sinai Hospital, Toronto, Canada; 3) University of Kansas Hospital, Kansas City, KS, USA; 4) New York University School of Medicine, New York, NY, USA; 5) Jewish General Hospital, Montreal, Quebec, Canada; 6) Mt. Sinai Hospital, New York, NY, USA; 7) Yale University School of Medicine, New Haven, CT, USA; 8) Emory University, Atlanta, GA, USA; 9) Genzyme, a Sanofi company Cambridge MA, USA.

Background: GD1 is one of the most common lysosomal storage disorders. Deficiency of the lysosomal enzyme glucocerebrosidase leads to accumulation of glucocerebroside in the spleen, liver, and bone marrow. Eliglustat, a novel oral substrate reduction therapy that selectively inhibits glucosylceramide synthase, is in development for the treatment of GD1. ENGAGE (NCT00891202) is a randomized, double-blind, placebo-controlled, Phase 3 trial sponsored by Genzyme, a Sanofi company investigating the efficacy and safety of eliglustat in untreated adults with GD1. Methods: Forty patients (mean age: 31.8 years; 20 males) with splenomegaly and thrombocytopenia and/or anemia were stratified by spleen volume and randomized 1:1 to receive eliglustat (50 or 100 mg BID depending on plasma levels) or placebo for 9 months. The primary efficacy endpoint was percent change in spleen volume (multiples of normal). Other efficacy measures included hemoglobin, liver volume, and platelets. Bone endpoints included bone marrow burden (BMB) scores and bone mineral density changes (DXA). Quality of life assessments included the Gaucher Disease Severity Scoring System (DS3). Safety monitoring included adverse event reporting, and lab and ECG evaluations. Results: In patients receiving eliglustat vs. placebo, mean spleen volume decreased (-28% vs. +2%, P<0.0001), mean hemoglobin increased (0.69 vs. -0.54 g/dL, P<0.0006), liver volume decreased (-5.20% vs. +1.44%, P<0.0072), and platelets increased (+32% vs. -9.06%, P<0.0001). Significant improvements (eliglustat vs. placebo) were observed for total (-1.1 vs. 0.0, p=0.002), spine (-0.6 vs. 0.1, p=0.002), and femur (-0.5 vs. 0.0, p= 0.026) BMB scores. Although patients with symptomatic bone disease were excluded, absolute change in total spine DXA Z-scores approached significance (LS mean treatment difference=0.2, p=0.06). Burden of disease, as measured by the Gaucher DS3, was significantly reduced following treatment with eliglustat vs. placebo (LS mean treatment difference=-0.3, p= 0.0452). No patient discontinued due to an AE, all of which were classified as mild to moderate; 39/40 patients transitioned into the ongoing open-label trial. Arthralgia and nasopharyngitis occurred in >10% of eliglustat vs. placebo patients. Conclusion: ENGAGE met its primary and secondary efficacy endpoints. Significant effects on bone marrow and a trend toward BMD improvement in spine were observed. Eliglustat was generally safe and well-tolerated.

2277T

Translating ManNAc into a novel therapeutic agent for patients with GNE Myopathy. *F.V. Celeste[1], L. Latham[1], J. DeDios[2], C. Ciccone[2], C. Robinson[1], J. McKew[1], M. Huizing[2], W.A. Gahl[2,3], N. Carrillo-Carrasco[1].* 1) Therapeutics for Rare and Neglected Diseases, National Center for Advancing Translational Sciences, Bethesda, MD; 2) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; 3) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892.

Background: GNE Myopathy is a rare muscular disease caused by mutations in GNE, which encodes the key enzyme in the sialic acid biosynthetic pathway. Patients have progressive muscle weakness leading to marked disability. Hyposialylation of muscle glycoproteins likely contributes to the pathophysiology. ManNAc, a pathway intermediate, prevents muscle weakness in the mouse model of GNE myopathy. Objectives: Translate ManNAc into a therapy for patients with GNE myopathy. Methods: A first-in-human Phase 1a, randomized, placebo-controlled, double-blind, single-dose study (ClinicalTrials.gov NCT01634750; IND No.78,091) was conducted to evaluate the safety, pharmacokinetics (PK), and pharmacodynamics (PD) of ManNAc. GNE myopathy subjects (n=22) were enrolled into 3 cohorts (Cohorts A: 3,000mg, B: 6,000mg, C: 10,000mg) with a 3:1 ManNAc-to-placebo ratio. Results: Grade I and II (CTCAE) adverse events were encountered, mostly unrelated to study drug. PK data are currently been evaluated. We utilized potential plasma biomarkers to evaluate the onset and duration of the PD response. Conclusions: It is critical to perform high-quality clinical trials to develop approved therapies for rare disorders. ManNAc, a promising therapy for GNE myopathy, is safe and well-tolerated. PK and PD data will guide the selection of dose and frequency for upcoming Phase Ib and IIa trials.

2278F

Evaluation of computed tomography *in vivo* to assess the therapeutic potential of cystine analogs as inhibitors of cystine stone formation in *Slc3a1* knockout mice. *A. Sahota[1], J. Parihar[2], M. Yang[1], D. Adler[3], W. Kim[4], D. Gordon[1], J.A. Tischfield[1].* 1) Dept Gen, Rutgers Univ, Piscataway, NJ; 2) Dept Surgery, UMDNJ-RWJMS, New Brunswick, NJ 08901; 3) Molec Imaging Ctr, Rutgers Univ, Piscataway, NJ; 4) Dept Math and Statistics, Univ South Florida, Tampa, FL 33620.

Background: *Slc3a1* knockout mice excrete cystine in the urine and males from age 3 months have cystine stones in the bladder. Cystine crystals are present in the urine in young mice, but very little is known about stone growth *in vivo*. Imaging technologies allow detailed investigation of biological processes in live animals. Here we evaluate the utility of computed tomography (CT) for identifying the spatial and temporal patterns of cystine stone deposition in *Slc3a1* knockout male mice and for assessing the efficacy of cystine dimethyl ester (CDME), a cystine analog, in ameliorating cystine stone formation. **Methods:** Mice (2-3 months old) were screened for the presence of stones using the Albira CT scanner; approximately 60% of these mice had stones. Initial bladder volume (as an indicator of stone volume) in the stone-positive mice was calculated using InviCRO VivoQuant software. Mice were then randomly assigned to treatment (CDME, 11 mice) or control (water, 7 mice) groups. Treatment consisted of 200 µl of 10 mg/ml CDME or water alone given daily by gavage for up to 10 weeks. Volume measurements were repeated at 1-2 week intervals. Mice were then sacrificed and bladder stones weighed, counted, and measured. Data on stone size and number were analyzed using chi-square. **Results:** Initial volumes in the control and treated groups were similar (12-96 mm^3 versus 11-86 mm^3, respectively). In both cases, there was a linear increase in volume that followed the equation $y=mx+c$. The equation for the water group ranged from $y=0.39x+11.0$ to $y=1.43x+84.3$. The equation for the CDME group ranged from $y=0.99x+18.4$ to $y=2.67x+73.3$. Thus, the two groups could not be distinguished by CT analysis. There was a small difference in the percentage of stone sizes in the two groups at the 5% level. The treatment group had 9.4% more stones and the control group had 19.6% fewer stones in the 1.1-2.0 mm range. The treatment group had 26.1% fewer stones and the control group had 54.6% more stones in the 3.1-4.0 mm range. Thus, CDME administration led to cystine stones that were smaller in size but larger in number compared with untreated mice. **Conclusions:** CT is a powerful tool for differentiating cystinuria mice with stones from those without stones, but longitudinal measurement of stone volume is not suitable for evaluating CDME efficacy *in vivo*. CDME administration leads to smaller stones which, based on volume measurements, cannot be distinguished from stones in untreated mice.

2279T

Pharmacologic inhibition of cytosolic translation offers a novel therapeutic approach for mitochondrial respiratory chain disease. *M.J. Falk[1], M. Peng[2], M. Tsukikawa[1], J. Baur[3], Z. Zhang[4], D. Gasser[2].* 1) Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Department of Physiology and Institute for Diabetes, Obesity, and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 4) Center for Biomedical Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Effective therapies do not currently exist for mitochondrial disease, a heterogeneous group of genetic disorders that cause multi-organ system dysfunction across all ages and ethnicities. Pharmacologically targeting the central alterations in the nutrient-sensing signaling network that contribute to mitochondrial disease symptoms is a novel treatment approach for which we have substantial data demonstrating benefit in cell and animal models of primary respiratory chain (RC) disease. mTORC1 is a key node in the nutrient-sensing signaling network whose regulation of cytosolic protein translation is highly relevant to RC disease. Cytosolic translation is significantly upregulated in both genetic and pharmacologic models of primary RC disease. Remarkably, cellular and animal health in RC deficient models can be restored by drugs that inhibit mTORC1. In particular, B6.*Pdss2*$^{kd/kd}$ mice that have RC disease caused by a coenzyme Q biosynthetic defect show increased mTORC1 activity prior to disease onset. The otherwise lethal renal glomerular disease in these animals is prevented by feeding either probucol or rapamycin, both of which are pharmacologic agents that our data demonstrate to effectively inhibit mTORC1. Further, we have discovered that more complete inhibition of cytosolic translation by cycloheximide, a drug that blocks the elongation phase of eukaryotic translation by binding the ribosome and inhibiting eEF2-mediated translocation, effectively prevents cell death from RC inhibition with rotenone, antimycin A, or oligomycin. Cycloheximide treatment also restores total cellular respiratory capacity in the setting of rotenone-based complex I inhibition. Collectively, these data suggest that dysregulated cytosolic protein translation is a significant cause of morbidity that can be therapeutically targeted to improve health in primary RC disease.

2280F

Drug screening using the dystroglycan null zebrafish. *G. Kawahara[1, 2], J. Widrick[1], V. Gupta[1, 2], J. Myer[1], M. Gasperini[1], A. Beggs[1, 2], L. Kunkel[1, 2, 3].* 1) Division of Genetics, Program in Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA, USA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA, USA; 3) Department of Genetics, Harvard Medical School, Boston, MA, USA.

Zebrafish are an ideal model for biomedical research, as zebrafish genetic models mimic much of the pathology of human diseases. Dystroglycan null fish (patchytail, dag1cl500) are an excellent model of the human dystroglycanopathies as they show disturbed muscle structure and a severe reduction of birefringence at 3-7 days post fertilization (dpf). The birefringence is a result of dorsal skeletal muscle deterioration and weakness that ultimately results in lethality of most dystroglycan null mutants at 20 dpf. We have performed a preliminary drug screen of 1120-chemicals from a small, commercial molecular library, Prestwick collection using homozygous patchytail zebrafish to identify potential therapeutic chemicals for treating the dystroglycanopathies, Embryos from heterozygous dag1cl500/wt zebrafish matings were cultured in normal fish water containing pools of up to 8 compounds from the Prestwick collection, and were incubated from 1 dpf to 4 dpf. At 4 dpf the treated fish were examined by birefringence to determine if there was any reduction from the expected 25% fry showing abnormal birefringence. Fish treated with some of the pooled chemicals were genotypically confirmed as dystroglycan null zebrafish but showed no defects in birefringence compared to the untreated control. This suggests that we may have successfully identified efficacious compounds for prevention of the phenotype caused by the dystroglycan gene mutation. In this short-term screen of 1120 compounds, we found eleven candidate drugs that prevent muscle pathology by significantly reducing the proportion of fry with abnormal birefringence at 4 days. Interestingly, five of the candidate drugs are involved in gamma-aminobutyric acid regulation. Dystroglycan null embryos were treated with each of these candidate drugs for a long-term period, from 5 dpf to 20 dpf. This treatment indicated that one of the candidate compounds greatly increased survival of dystroglycan null fish. Analysis of these treated surviving dystroglycan null fish showed they had normal muscle structure and normal neuromuscular junctions, whereas these structures are observed as abnormal in the untreated mutant fish. The candidate drugs from these screens will likely be informative for development of drugs to treat human dystroglycanopathies. The discovery of a small molecule and specific therapeutic pathway that might mitigate progression of this disease is highly relevant and significant.

2281T

Genetic and pharmacological reduction of ER stress rescues glaucoma in a murine model of glucocorticoid-induced glaucoma. *G. Zode[1], A. Sharma[1], X. Lin[1], C. Searby[1,2], K. Bugge[1,2], A. Clark[3], V. Sheffield[1,2].* 1) Department of Pediatrics, University of Iowa, Iowa, IA 52242; 2) Howard Hughes Medical Institute; 3) Department of Cell Biology & Anatomy and the North Texas Eye Research Institute, University of North Texas Health Science Center at Fort Worth, TX 76107.

Topical or systemic administration of glucocorticoids is known to induce ocular hypertension in a subset of the population. If untreated, these patients can develop glaucoma with features that closely resemble primary open angle glaucoma (POAG). The mechanisms underlying glucocorticoid-induced glaucoma are not fully understood partly due to the lack of an appropriate mouse model, as well as a lack of understanding of the genetic susceptibility of this disorder. Here, we report the development of a murine model of glucocorticoid-induced glaucoma that closely resembles glaucoma features observed in patients. We demonstrate that treatment with topical ocular 0.1% dexamethasone in WT mice leads to elevation of intraocular pressure (IOP), structural and functional loss of retinal ganglion cells, and axonal degeneration similar to glucocorticoid-induced glaucoma in human patients. We further demonstrate that dexamethasone-induced ocular hypertension is associated with chronic endoplasmic reticulum (ER) stress of the trabecular meshwork (TM). Dexamethasone activates the unfolded protein response and also induces pro-apoptotic transcriptional factor Chop in the TM. Deletion of Chop in mice reduces ER stress in the TM and prevents dexamethasone-induced ocular hypertension. Furthermore, a chemical chaperone, sodium 4-phenylbutyrate (PBA) prevents dexamethasone-induced ocular hypertension by reducing ER stress in the TM. ER stress thus plays a critical role in glucocorticoid-induced ocular hypertension and the reduction of ER stress via PBA can be used as a therapeutic strategy for treatment of glucocorticoid-induced glaucoma. This study also suggests that certain genes related to the ER stress pathway such as Chop may contribute towards increased susceptibility to steroid-induced glaucoma in humans.

2282F

Sildenafil citrate results in upregulation of heme oxygenase 1 and alleviation of symptoms in the *mdx^{5cv}* mouse model of Duchenne muscular dystrophy. *M. Gasperini[1], G. Kawahara[1,2], J. Widrick[1], M. Alexander[1,2], L. Kunkel[1,2, 3].* 1) Division of Genetics, Program in Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA, USA; 2) Department of Pediatrics. Harvard Medical School, Boston, MA, USA; 3) Department of Genetics, Harvard Medical School, Boston, MA, USA.

Duchenne muscular dystrophy (DMD), the most common form of muscular dystrophy, is caused by lack of the dystrophin protein, with no effective treatment at present. Zebrafish are a powerful tool for high throughput drug screening in vivo. From therapeutic drug screens using dystrophin-deficient *sapje* zebrafish, we have identified four drugs that increased affected fish survival rates at 20 days post fertilization. These drugs induced upregulation of heme oxygenase 1 (Hmox1) protein expression in surviving dystrophin null fish. One of these compounds, sildenafil citrate, is reported as a known effective mammalian drug for DMD. To confirm its therapeutic effects and its upregulation of Hmox1 in a DMD mouse model, sildenafil was administered to dystrophin null *mdx^{5cv}* mice. Sildenafil was dissolved to a 400 mg/L concentration in sterile-filtered, acidified drinking water. Two male and two female 4 week-old *mdx^{5cv}* mice were given ad libitum access to treated water for 6-8 weeks. During treatment, mice were tested for exercise tolerance by monitoring activity for 6 minutes immediately before and after 15 minutes of treadmill running. After the drug treatment period, skeletal muscle was histologically and physiologically examined to check for improvement of symptoms, as well as changes in Hmox1 expression. Whereas exercise tolerance at 2 and 4 weeks of drug administration showed no significant improvement, sildenafil treated mice displayed alleviated exercise fatigue symptoms after 6 weeks of drug administration. After treadmill running, their total seconds spent resting, centimeters traveled, and number of rearing events were significantly improved over the extremely fatigued vehicle-treated dystrophic mice. The sildenafil receiving mice also had a reduced number of centralized nuclei in skeletal muscle fibers in comparison to dystrophic controls, as counted in Haematoxylin and Eosin stains of diaphragm and tibialis anterior muscle slices. Immunoblot revealed that sildenafil treatment significantly increased Hmox1 protein expression in skeletal muscle when compared to vehicle-treated wildtype and *mdx^{5cv}* mice. These results point to Hmox1 as an effective therapeutic pathway in fish and mouse models of DMD. Hmox1 and drugs that impact it will likely be informative for development of therapeutics to treat humans with DMD.

2283T

Developing Small Molecule Inhibitors of p97/VCP Disease Mutants for Neurodegenerative Diseases. *T.-F. Chou[1], C.C Weihl[2], R.J. Deshaies[3].* 1) Division of Medical Genetics, Department of Pediatrics, Harbor-UCLA Medical Center and Los Angeles Biomedical Research Institute, Torrance, CA 90502, USA; 2) Department of Neurology, Washington University School of Medicine, 660 South Euclid Avenue, Box 8111, St. Louis, MO 63110, USA; 3) Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

p97 AAA (ATPase Associated with diverse cellular Activities; also called VCP [Valosin-Containing Protein) ATPase -- also known as p97 -- participates in key steps in ubiquitin-dependent protein quality control, autophagy, and numerous fundamentally important cellular functions. p97 targets proteins to 2 major degradation systems, the proteasome and autophagy pathways. This key role of p97 underscores its importance in protein homeostasis and strongly implicates its role in neurodegenerative diseases. Dominantly inherited missense mutations in p97 ATPase cause IBMPFD/ALS (inclusion body myopathy with Paget's disease of bone and frontotemporal dementia; amyotrophic lateral sclerosis). Some 20 single amino acid mutations in p97 have been identified. Current data suggest that all tested disease mutants form stable hexamers and have enhanced ATPase activity. We carried out a high-throughput screen to identify the first reversible p97 inhibitor, DBeQ. Following a structure-activity relationship study of the DBeQ scaffold, we constructed a library of p97 inhibitors. We have determined IC50 values of the top 3 inhibitors against 8 p97 disease mutants. Interestingly, some mutants were less sensitive to the inhibitors, with 5-fold higher IC50 values than for the wild type. We also performed enzyme kinetic analyses on wild type p97 and on the R155H and A232E disease mutants. Overall, our current data indicate that disease mutants are regulated by cofactors differently, compared to the wild type enzyme. Further study is required to develop mutant-selective inhibitors of p97 ATPase activity.

2284F

Rho Kinase Inhibition Therapy for Cerebral Cavernous Malformations.
D.A. Marchuk[1], D.A. McDonald[1], R. Shenkar[2], C. Shi[2], C.J. Gallione[1], C. Austin[2], A.G. Mikati[2], I.A. Awad[2]. 1) Molec Gen & Microbiol, Duke Univ Med Ctr, Durham, NC; 2) Neurovascular Surgery Program, University of Chicago Medicine and Biological Sciences, Chicago, IL.

Cerebral cavernous malformations (CCMs) are focal vascular malformations in the brain characterized by grossly dilated capillaries and vascular leak and hemorrhage, leading to stroke, seizure and other neurological complications. The disease occurs in sporadic or inherited (autosomal dominant) forms, the latter caused by mutation in one of three genes (CCM 1, 2 or 3). The vascular lesions in the familial cases follow a two-hit mutational mechanism of pathogenesis due to somatic mutation of the wild-type copy of the relevant CCM gene. Based on this two-hit model, we crossed heterozygous murine Ccm gene knockout alleles into mutant backgrounds with elevated rates of somatic mutation (p53 or Msh2 null alleles), and have generated robust murine models of CCM disease that fully and faithfully recapitulate the clinical phenotype. In parallel, in vitro studies have shown that the CCM proteins localize to endothelial cell-cell junctions and their loss leads to junctional instability associated with activation of RhoA and its effector, Rho kinase (ROCK). As a proof of principle, we recently demonstrated that ROCK inhibition using the drug fasudil reduces lesion burden in our murine Ccm1 model. In the current study we investigated the effect of ROCK inhibition in other Ccm murine genotypes, including a murine Ccm3 model, the most penetrant and morbid of the inherited forms in both the mouse and human. In all genotypes, fasudil treated mice exhibit decreased prevalence of CCM lesions compared to placebo controls. Lesions in treated animals exhibit fewer and smaller caverns (ie. 'immature' lesions) and are less likely associated with hemorrhage. Coming full circle, we have also now found that sporadic CCM lesions in human (the most common form of CCM) also exhibit somatic mutations in the same genes found mutated in the inherited cases. These combined data suggests mutation (somatic or inherited) of the CCM genes leads ROCK activation in all forms of CCM. Thus, ROCK inhibition may be a viable therapeutic option for all forms of CCM disease, including the most severe inherited form, CCM3, and possibly the more common, sporadic cases.

2285T

Development of a high-throughput screen for mRNA splicing modulators of IKBKAP.
F. Urbina, M. Nilbratt, E. Morini, M. Salani, S. Slaugenhaupt. Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA.

Familial Dysautonomia (FD) is an autosomal recessive sensory and autonomic neuropathy that primarily affects the central and peripheral nervous system. All patients with FD have an intronic splice site mutation in the IKBKAP gene (IVS20+6T$\rightarrow$C) that leads to aberrant splicing, with partial skipping of exon 20. This exon skipping is tissue specific and leads to a reduction of IKAP protein that is most pronounced in the nervous system. This reduction of IKAP leads to a number of symptoms including absence of overflow tears, gastroesophageal reflux, scoliosis, and orthostatic hypotension. Previously, we showed that the plant cytokinin kinetin dramatically improves aberrant splicing in FD patient lymphoblast and fibroblast cells. Further, kinetin works to modulate splicing in vivo improving exon 20 inclusion and increasing the amount of IKAP protein in all tissues, including brain, in a mouse model. In order to work towards optimizing kinetin and identifying new compounds that might be useful as human therapeutics, we established both primary and secondary screening assays. Using a previously created FD minigene construct, a dual luciferase-based primary assay was developed and optimized for high-throughput screening of compounds. A luciferase inhibition counterscreen was also developed. A secondary qPCR assay to evaluate the level of normal IKBKAP mRNA was established to both evaluate luciferase inhibitors and 'hit' compounds in both the primary assay and in patient cell lines. This screening funnel will allow for a large number of compounds to be evaluated in both a primary and patient-based secondary screen for splicing modulation. Several kinetin-like compounds have been evaluated to date and prove that our system will permit the identification of potential new therapies for this mRNA splicing disorder.

2286F

Paradigm Shift for the treatment of cognitive deficits in Down syndrome: A novel drug treatment rescues hippocampal deficits in the Ts65Dn mouse.
J.R. Korenberg[1], P. West[2], K. Wilcox[2], G. Smith[3], R. Kesner[4]. 1) Center for Integrated Neuroscience and Human Behavior, Brain Institute, Department of Pediatrics, Univ. of Utah, Salt Lake City, UT; 2) Department of Pharmacology and Toxicology, Univ. of Utah, Salt Lake City, UT; 3) Center for Integrated Neuroscience and Human Behavior, Brain Institute, Univ. of Utah, Salt Lake City, UT; 4) Department of Psychology, Univ. of Utah, Salt Lake City, UT.

It is now time to shift the paradigm from irrevocable wiring deficits to plasticity of neural circuitry solutions for therapeutics of neurodevelopmental disorders such as Down syndrome (DS). We report the use of the Ts65Dn mouse model of DS to establish a novel approach to the neural circuit definition and rescue of hippocampal deficits involving visual-spatial learning and memory. The identification and development of treatments for cognitive limitations in people with DS has needed greater committment by the pharmaceutical industry, a broader consideration of multiple neurotransmitter systems disturbed in DS, deeper integration with multi-target candidate compounds, parallel cognitive and behavioral tests in humans and animal models, and tests that specifically query the function of a specific brain circuit of known function. We have addressed these needs with a cross-disciplinary team to implicate the serotonin 5HT6 receptor antagonists as compounds whose effects on target neurotransmitter systems is inverse and therefore corrective, to the profile of serotonergic, cholinergic and GABAergic perturbations of Down syndrome. We report a novel battery of tests that dissect hippocampal functions and are compatible in mice and humans and further, that this class of compounds improves cognitive performance without increasing the inherent risk of seizures in the Ts65Dn mouse model of DS. We report the effects of 6 compounds representing a broad spectrum of 5HT6R antagonists, each of which rescues HC deficits of object identification and spatial orientation related to the function of CA 3 and the dentate gyrus. A 7th compound did not recognize the mouse receptor and did not rescue HC deficits. The effects of acute and/or chronic administration suggests that the 5HT6R antagonists may help to both acutely increase cognitive performance and chronically, to begin to repair neural circuitry in those with Down syndrome.

2287T

Sorbent Therapy (Rx) of Erythropoietic Protoporphyria (EPP).
P. Tishler. Brigham & Women's Hosp, Channing Division of Network Medicine, Boston, MA.

Sorbent Therapy (Rx) of Erythropoietic Protoporphyria (EPP) Peter V. Tishler PURPOSE: In EPP subjects, to determine the potential of a porphyrin adsorbant colestipol to reduce or eliminate cutaneous discomfort, hepatic dysfunction/failure, and body protoporphyrin content. BACKGROUND: EPP is a cutaneous porphyria caused by enzyme deficiency (ferrochelatase) or overproduction (δ-aminolevulinic acid synthase) in the heme biosynthetic pathway. Protoporphyrin accumulates, causing cutaneous photosensitivity and rarely hepatic dysfunction/failure. Prior In vitro,studies demonstrated that sorbents, such as cholestyramine, irreversibly bind protoporphyrin, and thus may therapeutically remove protoporphyrin secreted into the GI tract. METHODS: Three adult females were studied. After a no-Rx period of baseline blood studies (48-98 days), subjects received colestipol 2 gm twice daily (optimal dosage) from July-November. The time to sun-induced onset and nature of cutaneous discomfort were determined by stopwatch and answers to a questionnaire. Erythrocyte (RBC) and plasma free protoporphyrin concentrations were determined throughout. Changes were analyzed by the Cox Proportional Hazards model for skin, and mixed effects regression methods for protoporphyrin. RESULTS: Colestipol Rx improved skin sensitivity to sun 4.5-fold to infinity; all statistically significant. No hepatic dysfunction occurred before or during Rx. Changes in mean RBC & plasma protoporphyrin concentrations (μg/dl), during vs. before Rx (RBC - 1934 + 714 [SD] vs. 1712 + 455; plasma - 9.9 + 13.4 vs. 5.8 + 8.3) were not statistically significant. CONCLUSION: Colestipol ameliorated the protoporphyrin-induced painful skin response to sun. We could not assess its effect on liver function. Since blood protoporphyrin concentrations did not change, the mechanism is unclear. Colestipol Rx should be further evaluated in a controlled trial that includes youngsters, who are more liable for hepatic failure.

2288F

Fusion with Angiopep-2 to create proteins that cross the blood-brain barrier and are taken up into cells. *J.E. Lachowicz, M. Demeule, A. Regina, D. Boivin, A. Larocque, J.-P. Castaigne.* Angiochem, Montreal, Quebec, Canada.

Enzyme replacement therapy has been used successfully to treat peripheral symptoms of lysosomal storage disorders. However, treatment of the severe CNS symptoms associated with many LSDs has been impeded by the inability of enzymes to cross the blood-brain barrier (BBB). Receptor-mediated transcytosis is one strategy for allowing large molecules to cross the BBB. The low density lipoprotein-like 1 (LRP-1) receptor binds over 40 natural ligands for BBB transcytosis. Incorporation of Angiopep-2 (An2), a peptide that binds LRP-1, into small molecules, peptides, and biologics such as mAbs has been shown to increase brain permeability. The most advanced example is a conjugate of An2 and paclitaxel (ANG1005). In contrast to paclitaxel, ANG1005 is brain-penetrant and is not recognized by the P-gp efflux pump at the BBB. In addition, reduction in tumor size has been shown with ANG1005 both in preclinical models and in patients with primary and metastatic brain tumors. A visual demonstration of the An2 conjugate technology for proteins has been achieved by creating a recombinant human An2-GFP fusion protein. In vitro, native GFP is not readily taken up by cells. In contrast to native GFP, An2-GFP incubation results in observable green fluorescence within cultured fibroblasts. Mice that are systemically treated with native GFP via carotid artery or tail vein show similar brain fluorescence to vehicle-treated mice, which is limited to the capillary basement membranes. However, An2-GFP treatment results in abundant intracellular green fluorescence in brain. Ten minutes post injection, fluorescence is largely concentrated in astocytic end feet around capillaries, while 60 minutes post injection, a widespread fluorescence distribution is observed in neurons in all brain regions examined. At the subcellular level, punctate fluorescence is localized to presumptive endosomes, indicating active cellular uptake of An2-GFP. This study demonstrates that an An2-protein conjugate crosses the BBB and is taken up by cells in the brain, both attributes being critical for enzyme replacement therapy for CNS indications. We have applied this approach to an enzyme associated with mucopolysaccharidosis to create a brain-penetrant enzyme for enzyme replacement therapy.

2289T

Novel method for oral administration of ERT for the treatment of Gaucher disease. *M. Golembo, S. Velitzki, R. Chertkoff, E. Brill-Almon, Y. Shaaltiel.* Protalix Biotherapeutics, Carmiel, Israel.

Gaucher Disease (GD) is the most prevalent lysosomal storage disorder, caused by the deficiency of the lysosomal enzyme glucocerebrosidase (GCD). The accumulation of excessive glucocerebrosides in lysosomal compartments of macrophages causes hepatosplenomegaly, anemia and thrombocytopenia as well as bone pain and fractures. Enzyme replacement therapy (ERT) with GCD has been successfully used for the clinical treatment of GD for several years. All three currently approved ERT drugs are administered by intravenous infusion every 2 weeks. Although all enzymes are safe and efficient, the intravenous administration remains a limitation which affects patients' quality of life. Thus, orally administered ERT would have the definite advantage of the well-established therapy mechanism, without the limitation of the intravenous administration and will provide continuous enzyme secretion. Oral administration of proteins is one of the challenges of the biotherapeutics industry particularly due to early degradation of the proteins in the digestive tract. One of the approved ERTs, taliglucerase alfa, is expressed in carrot cells. The use of carrot cells as natural vehicle of the expressed human recombinant GCD is intended for oral delivery of the enzyme, due to the protection from degradation provided by the composition of the plant cell wall. Enzyme expressed in the ProCellEx system, prGCD, has the optimal glycosylation profile, with exposed mannose for efficient activity. This delivery method is particularly well suited as it uses edible carrot plant cells, and provides 'ready to use' enzyme which requires no modifications to obtain optimal glycosylation suited for receptor mediated uptake in target cells. This study shows the feasibility of this approach. Non clinical studies in rats and pigs fed with carrot cells expressing human GCD at either daily or repeat administrations, showed elevated levels of active GCD in both plasma and target organs (liver and spleen). Carrot cells expressing prGCD were shown to be safe and well tolerated by the animals and caused no major adverse symptoms. Clinical trial with oral administration of carrot cells expressing prGCD is on-going in GD patients. This study will evaluate enzyme delivery which allows daily intake and slow, continuous delivery of the drug. Once the enzyme is released in the blood stream it is expected to be similar to the approved ERT for which there is established clinical safety and efficacy data.

2290F

Preliminary Findings Evaluating Safety and Efficacy of Recombinant Human N-Acetlygalactosamine-6-Sulfatase in Pediatric Patients Less Than 5 Years of Age with Mucopolysaccharidosis IVA (Morquio A Syndrome, MPS IVA). *C. Haller[1], S.A. Jones[2], P. Harmatz[3], M. Bialer[4], R. Parini[5], K. Martin[1], P. Farmer[1], P. Slasor[1].* 1) Clinical Science, BioMarin Pharmaceutical Inc. , Novato, CA; 2) St. Mary's Hospital, CMFT, University of Manchester, MAHSC, UK; 3) Children's Hospital and Research Center, Oakland, CA,; 4) North Shore LIJ Health System, Manhasset, NY; 5) Az. Ospedaliera S. Gerardo, Monza, Italy.

Preliminary results after 26 weeks of treatment from an ongoing study evaluating safety and efficacy of recombinant human N-Acetylgalactosamine-6-sulfatase (rhGALNS) in 15 MPS IVA subjects <5 years of age are reported. The mean (range) age at baseline (BL) was 3.1 (0.8-4.9) years. Based on physical examination findings at BL, the majority of subjects had abnormal musculoskeletal features (14/15, 93.3%), abnormal general appearance (10/15, 66.7%), abnormal finding for HEENT (10/15, 66.7%) and corneal clouding (7/15, 46.7%). Most subjects had normal cardiovascular (12/15, 80.0%), respiratory (15/15, 100.00%), gastrointestinal (14/15, 93.3%), genitourinary (15/15, 100.0%), and neurologic systems (14/15, 93.3%) on exam. Standing height/length (n=15) was severely affected in many subjects; 7(46.7%) at <3rd, 3(20.0%) at ≥ 3rd-<10th, 2(13.3%) at >25th-<50th, and 3(20.0%) at ≥ 50th percentiles. Normalized urine keratan sulfate (uKS) was elevated 3 fold above the mean for age matched control population with a mean (range) of 35.9 (18.8-56.5) µg/mg creatinine (n= 15). After 26 weeks of weekly infusions with 2 mg/kg/day of rhGALNS, the most commonly reported adverse events (AEs) were vomiting in 12 (80.0%), pyrexia in 11 (73.3%), and cough in 8 (53.3%) subjects. The majority of AEs were mild-moderate. Four serious AEs were reported, tonsillar hypertrophy, skin infection, sepsis and hypersensitivity. No subjects discontinued treatment, and rhGALNS had a similar safety profile as seen in older children and adults. In 8 subjects with 26 weeks of data, rhGALNS led to a substantial decrease in mean (±SD) normalized uKS by -30.5 (±15.49)% after 2 weeks that was sustained at -35.2 (±15.57)% at 26 weeks. These results demonstrate the drug's pharmacodynamic effects. Mean height/length for age z-scores did not demonstrate significant change from BL (-1.8 SD) to Week 26 (-2.2 SD) for these 8 subjects. Although there was no control group for comparison, data in the literature indicate that these children experience growth failure in early childhood. The accumulation of KS begins early in the life of MPS IVA subjects, and is progressive and life limiting. Early intervention with rhGALNS treatment is well tolerated; produces decrease in keratan sulfate storage, and may show substantial benefit in young pediatric patients.

2291T

Survival Rates and Timing of Initiation of Treatment with Enzyme Replacement Therapy (ERT) Among Patients with Classic Infantile-Onset Pompe Disease Enrolled in the Pompe Registry. *P.S. Kishnani[1], S. Jones[2], A. van der Ploeg[3], E. Mengel[4], B. Byrne[5], A. Vellodi[6], N. Leslie[7], S. Shankar[8], P. Tanpaiboon[9], D.W. Stockton[10], J.B. Hennermann[11], Z. Devecseri[12], J. Kempf[12], J. Keutzer[12], Y-H. Chien[13].* 1) Division of Medical Genetics, Department of Pediatrics, Duke University Medical Center, Durham, NC, USA; 2) Manchester Center for Genomic Medicine, CMFT, University of Manchester, UK; 3) Department of Pediatrics & Center for Lysosomal and Metabolic Diseases, Erasmus MC University Medical Center, Rotterdam, the Netherlands; 4) University of Mainz, Mainz, Germany; 5) Department of Pediatrics, University of Florida, Gainesville, FL, USA; 6) Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; 7) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 8) Emory University School of Medicine, Atlanta, GA, USA; 9) Children's National Medical Center, Washington, DC, USA; 10) Wayne State University School of Medicine, Detroit, MI, USA; 11) Charite Universitaetsmedizin Berlin, Berlin, Germany; 12) Genzyme, a Sanofi company, Cambridge, MA, USA; 13) National Taiwan University Hospital, Taipei, Taiwan.

Background: Pompe disease presents as a clinical spectrum with variable severity, progression, and muscle involvement. Early clinical diagnosis and initiation of enzyme replacement therapy (ERT), before the development of severe, irreversible symptoms, are important, especially in very young patients with the most severe form of the disease. Aim: To report the results of a descriptive analysis of the chances of survival and ventilator-free survival and its relation to timing of initiation of ERT among patients with symptom onset ≤12 months with cardiomyopathy (classic infantile-onset Pompe disease) in the Pompe Registry, sponsored by Genzyme. Methods: All patients in the Registry with symptom onset ≤12 months with evidence of cardiomyopathy and a record of treatment with ERT were eligible for analysis. Kaplan-Meier curves were fitted, stratified by patients with a recorded first ERT infusion <3 months or ≥3 months of age. Events were defined as death and use of invasive ventilation therapy or death. The time to the event was derived as time from birth. Only patients at risk for an event were included in the analyses. Because patients from Taiwan may have been identified by newborn screening (NBS) and not clinical diagnosis, analyses were performed with and without patients with a reporting physician from Taiwan. Results: A total of 140 patients with symptom onset ≤12 months with evidence of cardiomyopathy were eligible for analysis. Patients who received their first treatment with ERT at <3 months of age (n=36) had better survival rates than patients who started ERT at ≥3 months of age (n=104) at 12 months of follow-up (94% for ERT initiation at <3 months of age vs 91% for ERT initiation at ≥3 months of age) and more noticeably at 24 months (85% vs 73%) and 36 months (81% vs 64%) of follow-up. Invasive ventilator-free survival rates also were better for patients with initiation of ERT <3 months of age (n=24) compared with patients who started ERT at ≥3 months of age (n=69) at 12 months (91% vs 90%) and more noticeably at 24 months (82% vs 66%) and 36 months (76% vs 56%) of follow-up. Results were similar at all timepoints when patients from Taiwan were excluded. Conclusions: Earlier initiation of treatment in classic-infantile onset Pompe disease patients may improve the chances of survival and lead to better clinical outcomes overall and improvement of symptoms for these young, very ill patients.

2292F

PRX-102: an efficient plant expressed ERT for Fabry disease. *A. Paz, T. Kizhner, A. Schulman, M. Golembo, E. Brill-Almon, Y. Shaaltiel.* Protalix Biotherapeutics, Carmiel, Israel.

PRX-102, is a chemically modified, plant-cell expressed recombinant human alpha-GAL-A enzyme (prh-alpha-GAL-A) developed as enzyme replacement therapy (ERT) for the treatment of Fabry disease. There are currently two ERTs of recombinant human alpha-GAL-A commercially available: agalsidase-alfa and agalsidase-beta. The plant cell expressed recombinant human alpha-GAL-A has been modified by cross-linking and the addition of PEG molecules resulting in an enzyme which exhibits higher stability and prolonged activity under physiologically relevant conditions when compared to the commercially available drugs. We demonstrated that PRX-102 is active and localizes to the intracellular target, the lysosome, in Fabry patients' cells. PRX-102 presents higher stability under physiologically relevant conditions, and extended circulation residence time as compared to available commercial ERT. Furthermore, pharmacodynamic studies using a mouse model of Fabry disease show higher uptake in target organs than available commercial ERT and significant reduction of Gb3 accumulation. The extended circulation time, together with the increased uptake of PRX-102 as compared to commercial ERT, has the potential for a more efficient ERT for the treatment of Fabry disease. Currently, the PRX 102 enzyme is under evaluation in Fabry patients as part of a clinical study.

2293T

The Influence of a Polymorphism in the Gene Encoding Angiotensin Converting Enzyme (ACE) on Treatment Outcomes in Late-Onset Pompe Patients Receiving Alglucosidase Alfa. *R.C. Baek, R.E. Palmer, R.J. Pomponio, A.J. Mcvie-Wylie.* Genzyme, a Sanofi Company, Framingham, MA.

Angiotensin Converting Enzyme (ACE) catalyzes the conversion of angiotensin I to the vasoactive peptide angiotensin II, and degrades bradykinin, a potent vasodilator. One outcome of these two actions is vasoconstriction and an increase in blood pressure. Half of the variation in human ACE activity can be accounted for by an insertion/deletion (I/D) allele in intron 16 of the ACE gene. An insertion of an Alu repeat in intron 16 results in lower ACE activity (I allele), and conversely, the deletion (D allele) results in higher ACE activity. Furthermore, the I allele has been reported to be associated with the predominance of slow-twitch muscle fibers (Type I) and the D allele with fast-twitch muscle fibers (Type II). Two publications recently suggested that the ACE I/D polymorphism modifies the clinical presentation of Pompe patients and influences treatment outcomes following alglucosidase alfa enzyme replacement therapy (ERT). Specifically, patients with the D/D genotype presented with an earlier onset of disease, higher CK levels at diagnosis, increased pain, and more severe disease progression (Fillippi et al, 2010), and had poorer treatment outcomes on ERT (Ravaglia et al, 2012). We investigated these findings in a large cohort of late-onset Pompe patients included in a randomized, placebo-controlled trial of alglucosidase alfa. Our results also suggest that patients carrying two DD alleles demonstrate an attenuated response to treatment relative to the I/D and I/I genotypes. This result was not associated with the antibody response to alglucosidase alfa. An investigation into the disease status of patients at entry into the clinical trial is ongoing.

2294F

Polymeric Nanocarriers as Vehicles for the Treatment of Lysosomal Storage Diseases. *M. Latorre-Esteves, A. Román.* Chemical Engineering, University of Puerto Rico- Mayagüez Campus, Mayagüez, PR.

Lysosomal Storage Diseases (LSDs) are a group of inheritable genetic diseases caused by mutant lysosomal enzymes, leading to the accumulation of undigested macromolecules in the lysosome, thus causing increases in lysosome size and number, cellular dysfunction, clinical abnormalities, and premature death. Some LSDs can be treated with Enzyme Replacement Therapy (ERT) through intravenous administration of a recombinant enzyme in replacement of the defective enzyme. However, this is an expensive and inefficient method with adverse side effects associated with the high enzyme amounts required for the treatment, the need of post-translational modification of the enzyme and the host immune system response. We hypothesize that nanocarriers composed of Polyethylene glycol and Polycaprolactone (PEG-PCL) block copolymers can enhance ERT by eliminating the need of enzyme modification and protecting it from host immune system until lysosomal target is reached. We have designed these nanocarriers to remain stable at physiological pH and destabilize at acidic pH, allowing them to reach the cell intact and degrade once inside the acidic lysosome, only then releasing therapeutic cargo into affected cellular organelle. In order to obtain the appropriate nanocarriers for ERT application, we synthesized a group of PEG-PCL nanocarriers, varying the production process and surfactant concentration. The most suitable combination was found by performing dynamic light scattering analysis (DLS), gel permeation chromatography (GPC), Themogravimetric Analysis (TGA) and several spectroscopic techniques. We found the nanocarriers are not cytotoxic in cell lines tested. We are currently working on the reproducibility of syntheses and protein loading capacity into nanocarriers.

2295T

Augmenting glucocerebrosidase activity in the CNS as a therapeutic strategy for Gaucher-related synucleinopathies. *P. Sardi, J. Clarke, C. Viel, M. Chan, C. Treleaven, J. Bu, L. Sweet, M. Passini, J. Dodge, L. Stanek, S. Cheng, L. Shihabuddin.* Genzyme, a Sanofi Company, Framingham, MA.

Mutations of *GBA1*, the gene encoding glucocerebrosidase, represent a common genetic risk factor for developing the synucleinopathies Parkinson's disease (PD) and dementia with Lewy bodies (DLB). PD patients with or without *GBA1* mutations also exhibit lower enzymatic levels of glucocerebrosidase in the central nervous system (CNS), suggesting a possible link between the enzyme and the development of the disease. To probe this link further, we evaluated the efficacy of augmenting glucocerebrosidase activity in the CNS of a mouse model of Gaucher-related synucleinopathy ($Gba1^{D409V/D409V}$) and a transgenic mouse overexpressing A53T α-synuclein. Adeno-associated virus-mediated expression of glucocerebrosidase in the CNS of symptomatic $Gba1^{D409V/D409V}$ mice completely corrected the aberrant accumulation of the toxic lipid glucosylsphingosine and reduced the levels of ubiquitin and proteinase-K-resistant α-synuclein aggregates. Importantly, hippocampal expression of glucocerebrosidase in $Gba1^{D409V/D409V}$ mice (starting at 4 or 12 months old) also reversed their cognitive impairment when examined using the novel object recognition test. Overexpression of glucocerebrosidase in the CNS of A53T α-synuclein mice reduced the levels of soluble α-synuclein, suggesting that this glycosidase can modulate the development of α-synucleinopathies. Hence, increasing glucocerebrosidase activity in the CNS represents a potential therapeutic strategy for GBA1-related and non-GBA1-associated synucleinopathies.

2296F

Permanent genetic modification of dividing cells using episomally maintained S/MAR DNA vectors and the correction of a cancer phenotype in renal tumour cells. *R. Harbottle[1,2], SP. Wong[2].* 1) DNA Vector Research Group, German Cancer Research Centre (DKFZ), Heidelberg, Germany; 2) Gene Therapy Research Group, Imperial College London, UK.

The simple, stable and efficient application of episomal DNA vectors to genetically modify dividing cells without the risk of integration-mediated genotoxicity provides a valuable tool in cell biology research. Here, we demonstrate the utility of Scaffold/Matrix Attachment Region (S/MAR) DNA plasmid vectors to rapidly and simply generate novel genetically modified cell lines. In this study we utilize these vectors to model the restoration of a functional wild-type copy of the gene implicated in the renal cancer Birt-Hogg-Dube (BHD) in a cell-line (UOK257) derived from a BHD patient. Inactivation of the BHD gene, encoding a protein called folliculin (FLCN) has been shown to be involved in the development of sporadic renal neoplasia in BHD. Persistent genetic correction of UOK257 cells with an S/MAR-FLCN plasmid (UOK257-FS) restores FLCN expression and normalises downstream TGFβ signalling. We demonstrate that UOK257-FS cells show a reduced growth rate in vitro and suppressed xenograft tumour development in vivo, compared to the original FLCN-null UOK257 cell line. We also show that mTOR signaling in serum-starved FLCN-restored cells is differentially regulated compared to the FLCN-deficient cells indicating the complex role of FLCN in separate signalling pathways. The novel UOK257-FS cell line will be a useful tool for studying different feedback loops in signalling pathways affected in BHD pathogenesis and provides further insight into the role of the FLCN in BHD. More significantly, this study demonstrates the suitability of episomally maintained S/MAR DNA vectors to successfully model the persistent functional expression of a therapeutic gene in a cancer cell line. It should also establish this class of vectors as effective tools for investigating signalling components differentially regulated in different cancers and to aid the identification of novel cancer markers for diagnosis and therapy. Additionally, this study illustrates the utility of this class of vectors for the genetic modification of cells without the risk of genotoxicity.

2297T

Brain-directed recombinant AAVrh10-ATP7A plus subcutaneous copper histidine extends lifespan in the *mo-br* model of Menkes disease. *E.Y. Choi, M.R. Haddad, K. Patel, S.G. Kaler.* Molecular Medicine Program, NICHD/NIH, Bethesda, MD.

Menkes disease is a debilitating X-linked recessive neurometabolic disorder caused by mutations in a copper-transporting ATPase, ATP7A. No long-term, predictably effective treatment for affected individuals has been developed in the 50 years since the initial description of Menkes disease. While subcutaneous injections of copper may bypass the defect in intestinal copper absorption and normalize blood copper levels, at least a rudimentary level of ATP7A function is needed to restore proper intraneuronal copper transport, and attain normal neurodevelopment. Notably, early copper replacement in some affected patients whose mutations permitted as little as 5-10% residual ATP7A copper transport activity led to completely normal neurodevelopmental outcomes. We found previously that intracerebroventricular (ICV) administration of rAAV (serotype 5) harboring an expression cassette with a compact version of human *ATP7A* had a salutary, synergistic effect with ICV copper admininistration in a lethal mouse model of Menkes disease, *mo-br*. This combination therapy led to higher brain copper concentrations, reduced neuropathology, and enhanced survival when administered to the lateral ventricles of newborn *mo-br* mice (Donsante et al. *Mol Ther*, 2011). A recently identified AAV serotype, rh10, shows broader neuronal tropism than AAV5, demonstrating robust transduction of multiple neuronal cell types, as well as choroid plexus epithelia. Broad neuronal transduction would be especially desirable in instances of a complete loss-of-function ATP7A mutation. Brain-directed rAAVrh.10 represents a desirable route of administration from immunological and tissue target perspectives. Therefore, we have advanced our efforts in the *mo-br* model of Menkes disease by evaluating the efficacy of ICV administration of rAAVrh.10-ATP7A in combination with subcutaneous copper histidine (CuHis). We demonstrated extended lifespan and brain neurochemical correction with this regimen. The optimal timing of these treatments to date was low dose (1.6×10^9) rAAVrh10 on day 2 with subcutaneous CuHis (10 µg) on day 5 (single administration of each). These dose-ranging preclinical studies presage application of this approach in human subjects with Menkes disease associated with severe loss-of-function ATP7A mutations.

2298F

Genotoxicity following AAV Gene Therapy for Methylmalonic Acidemia (MMA) in Mice. *R.J. Chandler, A.A. Ashok, G.K. Varshney, M.C. LaFave, W. Wu, V.J. Hofffmann, A.G. Elkahloun, S.M. Burgess, C.P. Venditti.* Nat Human Genome Res Inst, National Institutes Hlth, Bethesda, MD.

Many studies have repeatedly demonstrated the safety of adeno-associated virus (AAV) vectors. However, one report has documented an increased incidence of hepatic cellular carcinoma (HCC) in neonatal mice after AAV gene therapy and further implicated insertional mutagenesis by the vector at the *RIAN* locus as a causative factor (Donsante *et al., 2007*). We have previously demonstrated the pre-clinical efficacy of AAV gene delivery using a neonatal lethal MMA murine model. Although the evaluation of genotoxicity was not the intended focus of our studies, the long-term surveillance of both AAV-treated MMA mice and control littermates has revealed that 75% (n=48) of AAV8-CBA-Mut treated mice developed HCCs between 12 and 21 months compared to a 2% (n=41) HCC rate in untreated littermates. In addition, 50% of the mice similarly treated with a control vector, AAV8-CBA-GFP, also developed HCCs, indicating that overexpression of the Mut transgene is not responsible for the development of HCC. We next performed a dose escalation study by increasing the AAV8 dose from $7\times10^{11-12}$ GC/kg (n=16) to 1×10^{14} GC/kg (n=19) and noted a corresponding increase in the incidence of HCC from 12% to 84%, respectively. As previously observed by Donsante *et al., 2007*, our preliminary data indicates that insertional mutagenesis of the vector at the *RIAN* locus, with subsequent dysregulation of local miRNA expression, occurs in AAV-associated HCCs. Determining why AAV gene delivery is infrequently associated with tumorigenesis in mice and whether such toxicity is relevant to humans remain as unresolved questions for those using AAV in human gene therapy applications.

2299T

Developing resources to alleviate muscle atrophy in FSHD by genome engineering. *S. Das, B.P. Chadwick.* Department of Biological Science, The Florida State University, Tallahassee, FL.

Facioscapulohumeral muscular dystrophy (FSHD), a debilitating disease with currently no cure or effective therapy, is the third most common inherited form of muscular dystrophy. FSHD is primarily characterized by progressive weakness and atrophy of skeletal muscle of the face, shoulders and upper arms. Both forms of this autosomal dominant disorder (FSHD1 and FSHD2) are likely caused by the reactivation of the double homeobox gene DUX4 that is epigenetically repressed in somatic cells. The more common form, FSHD1, is caused by a contraction in the number of repeat units of the macrosatellite D4Z4 at 4q35. Each repeat unit contains the DUX4 open reading frame. FSHD2 is a result of a digenic inheritance of mutant SMCHD1 gene and a permissive 4q allele but is contraction-independent. The inappropriate expression of DUX4 in somatic muscle inhibits muscle differentiation and replenishment. Importantly, both forms of FSHD have a strict requirement for specific allelic variants of chromosome 4q and at least one unit of D4Z4, as complete loss of the repeat does not result in disease. Taking advantage of this, we are focused on development of an effective therapeutic strategy for FSHD using genome engineering in human cells through custom-built Transcription Activator Like Effector Nucleases (TALENs). Currently, we are developing TALENs to target regions immediately proximal and distal to the D4Z4 array that can be used to remove the array or disrupt the polyadenylation signal on a permissive chromosome, in order to alleviate toxic gain of function. We are currently testing candidate nucleases on model cell lines. Successful editing of the diseased allele will permit use of our custom-built nucleases, to modify patient specific induced pluripotent stem cells (iPS), with the ultimate goal of using modified cells as a possible therapeutic strategy for muscle restoration. Our approach could be used in FSHD patient-specific cell transplantation therapy, in an attempt to alleviate this debilitating condition, and improve quality of life.

2300F

Gene Therapy for Mucopolysaccharidosis VII: Evaluation of Intrathecal rAAV Vectors in the Canine Model. *B.L. Gurda[1], P. Wang[2], P. Bell[1], J. Bagel[3], T. Sikora[2], P. O'Donnell[2], Y. Zhu[1], H. Yu[1], T. Ruane[2], R. Calcedo[1], M. Casal[2], C.H. Vite[3], K.P. Ponder[4], J.M. Wilson[1], M.E. Haskins[2].* 1) Pathology and Laboratory Medicine, School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Pathology and Medical Genetics, School of Veterinary Medicine, University of Pennsylvania Philadelphia, PA; 3) Section of Neurology & Neurosurgery, Department of Clinical Studies, School of Veterinary Medicine University of Pennsylvania, Philadelphia, PA; 4) Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO.

Mucopolysaccharidosis Type VII (MPS VII) is a lysosomal storage disease resulting from mutations in β-D-glucuronidase (GUSB) necessary for the catalysis of glycosaminoglycans (GAGs). The central nervous system (CNS) has been difficult to treat with systemic gene therapy due to inefficient transduction of the brain and poor diffusion of enzyme across the blood brain barrier. The goal of this study was to determine the expression of two adeno-associated viral (AAV) vectors in the CNS of MPS VII dogs, comparing intravenous (IV) injection with or without intrathecal (IT) administration. Four MPS VII dogs were treated IV at post-natal day 3 with 2E13 genome copies (GC)/kg of either an AAV9 (n=2) or AAVrh10 (n=2) vector encoding the canine GUSB cDNA under the control of a chicken beta-actin promoter. At 60 days post-IV injection one animal from each group was injected IT into the cisterna magna with the other vector (≈4E12GC/kg). Enzyme activity in the serum and CSF was monitored over 6 months, while brain and spinal cord collected at sacrifice were analyzed for GUSB activity, a secondary storage substrate (GM3 ganglioside), and vector copy number. Serum GUSB activity stabilized at ≈30% of normal in all four dogs. Activity in the CSF was near normal in the dogs that did not receive vector IT and 140- and 100-fold normal in the AAVrh10 and AAV9 IT-treated dogs, respectively. Mean biochemical GUSB activity of six brain regions was >4.5-fold normal in the IT-AAV9 dog and >2.5-fold normal for the IT-AAVrh10 dog, correlating to vector copy numbers. The activity in the cervical spinal cord was 4.7- and 2.5-fold normal for AAV9 and AAVrh10, respectively. GUSB activity in the CNS from the IV-only injected dogs was <4% of normal. Only after IT delivery were low levels of AAV-antibodies detected in the CSF. Histochemical staining for GUSB activity was widespread throughout the CNS in the IT-injected dogs: Purkinje cells of the cerebellum were particularly prominent. GUSB staining was seen in the choroid plexus of all treated dogs, likely explaining the normal GUSB activity in the CSF of the IV-only treated dogs. There was a reduction in GM3 ganglioside in the IT-treated dogs by immunohistochemistry. Somatic tissues were also analyzed. In conclusion, this study determined that IT injection of AAV vectors resulted in impressive levels of expression in the CNS of MPS VII dogs and has tremendous potential for treating the CNS manifestations of lysosomal storage disorders.

2301T

Translational Fidelity of Intrathecal Delivery of scAAV9-SMN1 for Spinal Muscular Atrophy. *M.A. Passini[1], J. Bu[1], A.M. Richards[1], C.M. Treleaven[1], C.R. O'Riordan[1], A. Scaria[1], A.P. Kells[2], L. Samaranch[2], W. San Sebastian[2], T. Federici[3], M.S. Fiandaca[4], N.M. Boulis[3], K.S. Bankiewicz[2], L.S. Shihabuddin[1], S.H. Cheng[1].* 1) Genzyme, a Sanofi Company, Framingham, MA; 2) University of California, San Francisco, CA 94103; 3) Emory University, Atlanta, GA 30322; 4) Georgetown University, Washington DC 20057.

Spinal muscular atrophy (SMA) is a pediatric neuromuscular disease caused by mutations in SMN1. Previously, we showed that delivery of an AAV vector encoding SMN1 into the CNS of a mouse model of SMA corrected several aspects of the disease and afforded a significant extension in their lifespan. Here, we determined (i) the minimal number of genetically modified motor neurons in the spinal cord of SMA mice necessary for clinically relevant efficacy, and (ii) the level of motor neuron transduction efficiency following intrathecal delivery in juvenile pigs and non-human primates. Injection of 5e10, 1e10, and 1e9 genome copies (gc) of scAAV9-hSMN1 into the CNS of SMA mice extended their median lifespans from 14 days to 153d, 70d, and 18d, respectively. Analysis of the spinal cords treated with 5e10, 1e10, or 1e9 gc showed that 30-60%, 10-30%, and <5% of the motor neurons had detectable levels of hSMN protein, respectively. Hence, achieving at least 10-30% motor neuron transduction was sufficient to significantly improve motor function and survival in SMA mice. To determine the results of gene transfer in larger animal models, a viral vector encoding GFP (scAAV9-GFP) was administered intrathecally into juvenile pigs and monkeys. Analysis showed that a motor neuron transduction rate as high as 25-75% throughout the spinal cord in these large animal studies. Thus, the number of motor neuron transduction shown necessary for clinically relevant efficacy in SMA mice could be obtained in large animal models, justifying the continual development of intrathecal gene therapy for SMA.

2302F

Treatment of MFRP (Membrane frizzled-related protein)-related degeneration in patient-specific stem cells and a preclinical mouse model. *Y. Li, Y.T. Tsai, C.W. Hsu, W.H. Wu, S.H. Tsang.* Edward Harkness Eye Institute, Columbia University, New York, NY.

In the current era of personalized medicine, a large number of genetic variants have been discovered in patients with various diseases using next-generation sequencing techniques. Traditionally, to prove that genetic variants cause diseases, we have had to rely on animal models. However, substantial differences exist between mice and humans, including but not limited to drastic differences in lifespans. For instance, dopaminergic neuron projections are shorter in relation to overall body length in mice than in humans; thus, alpha synuclein Snca knockout mice do not develop Parkinson's disease. Instead, as new reprogramming technologies have developed, it has become more feasible to generate patient-specific stem cell lines to validate sequence variants, elucidate pathophysiology, and perform targeted drug screening. We have used such reprogramming technologies to prove that a novel mutation in the *MFRP* (membrane-type frizzled-related protein) gene is pathogenic. Light microscopic images of human stem-cell-derived retinal cells taken from a patient displayed dysmorphic cells. Correction of the defect restored the patient's retinal cells' morphology and function. Similarly, correcting genetic defects in the $Mfrp^{rd6}/Mfrp^{rd6}$ mouse can restore visual function in this preclinical model of retinitis pigmentosa.

2303T

The combined use of cis-acting genetic elements to enhance non-viral plasmid stability in mammalian cells. *B. Abbas[1], M.AK Abdalla[2], M. El-Mogy[3], B. Lam[4], V. Misic[1], Y. Haj-Ahmad[1,4].* 1) Brock University, St. Catharines, Ontario, Canada; 2) Department of Biochemistry, Faculty of Science, Alexandria University, Egypt; 3) Molecular Biology Department, National Research Center, Dokki, Cairo, Egypt; 4) Norgen Biotek Corp., Thorold, Ontario, Canada.

Gene transfer and gene expression are attractive methods for the generation and the delivery of recombinant therapeutic proteins, which basically need an efficient and safe delivery tool. Viruses have been used for many years as gene transfer vehicle for the delivery and the transient expression of recombinant proteins. However, to date, there is no efficient non-viral delivery tool for expression of plasmids (pDNA) as compared to viruses. It is proposed that the unsuccessful development of non-viral vector may be due to many barriers such as the transfection efficiency and plasmid stability. It has been reported that viruses contain cis - acting RNA elements that facilitate the posttranscriptional processing and the exporting of mRNA. Furthermore, it was also reported that the DNA cis - acting matrix attachment region elements (MAR) promote homologous recombination to enhance the level of DNA stability in the cell. Moreover, BGH poly signal has been reported to facilitate mRNA stability and RNA export process. Our hypothesis is, by using the viral cis - acting elements, BGH poly signal and/or matrix attachment region (MAR) in a non-viral vector may prolong the stability and hence the expression of the delivered gene. Therefore, in this study, we constructed various plasmid vectors harboring different cis-acting elements that were evaluated for enhancing plasmid DNA as well as mRNA stability in two mammalian cell-lines, HEK293 and HeLa. Both cell lines were transfected with equal copy number and plasmid DNA and mRNA copy numbers per cell were measured during three weeks. Data showed that the combination of multiple cis-acting elements has significant, unexpected effect on both plasmid DNA and mRNA stability. In summary, these results provide valuable information to improve non-viral vectors stability and transient gene expression in mammalian cells; with implications in both, gene expression studies, as well as gene therapy.

2304F

Expression of human GNE through adeno-associated virus mediated therapy delays progression of myopathy in the GNE myopathy mouse model. *M.C. Malicdan[1,2], T. Okada[3], S. Takeda[3], F. Funato[1], M. Huizing[2], I. Nonaka[1], Y.K. Hayashi[1], Z. Argov[4], I. Nishino[1], S. Mitrani-Rosenbaum[4], S. Noguchi[1].* 1) MGB, NHGRI, National Institutes of Health, Bethesda, MD; 2) Department of Neuromuscular Research, National Institute of Neuroscience, NCNP, Tokyo, Japan; 3) Department of Molecular Therapy, National Institute of Neuroscience, NCNP, Tokyo, Japan; 4) Hadassah-Hebrew University Medical School, Jerusalem, Israel.

GNE myopathy, also known as distal myopathy with rimmed vacuoles (DMRV) or (hereditary inclusion body myopathy (hIBM) is an early adult, moderately progressive myopathy due to mutations in GNE. This gene encodes a bifunctional enzyme critical to synthesis of sialic acid. GNE myopathy so far has been shown to be a disorder due to reduced sialylation of certain glycoconjugates in tissues including the muscle, as replenishment of sialic acid in the murine model has prevented the onset of a muscle phenotype. The caveat in sialic acid supplement therapy lies in the pharmacokinetic properties of sialic acid; its relatively short half-life due to rapid excretion after oral administration implies the need to explore other modalities. This paper thus aims to demonstrate the utility of gene therapy in ameliorating the muscle phenotype of the DMRV/hIBM mouse model. In this study, we generated AAV expressing human wt-*GNE* cDNA and enhanced green fluorescent protein (eGFP) under the transcriptional control of cytomegalovirus promoter in AAV 2/8 capsids. AAV 2/8 constructs that carry either human wt GNE or luciferase (in control virus) were injected intravenously into adult and symptomatic GNE myopathy mice. Unaffected littermates were also injected as controls. At 10 weeks after injection, eGFP expression was seen in a remarkable number of cells in the skeletal muscle, liver, kidney, heart, and spleen. Measurement of mRNA with specific human versus mouse probes revealed an increase in virus-derived human GNE expression. More importantly, GNE myopathy mice injected with AAV2/8-wt h*GNE* at 47 weeks of age showed a significant improvement in survival, motor performance, muscle size and contractile properties, as compared to those mice injected with AAV2/8-luciferase. These results establish a proof of principle in using AAV-mediated gene therapy for GNE myopathy.

2305T

Antisense Therapy for Genetic Disorders. *T.R Grossman.* Antisense Drug Discoverey, ISIS PHARMACEUTICALS, Carlsbad, CA.

Advances in deciphering the complex roles RNA plays in normal health and disease have been substantial over the past decade and RNA is becoming an increasingly important target for therapeutic intervention; accordingly therapeutic strategies that modulate RNA function are becoming increasingly important. Antisense oligonucleotides (ASO) are perhaps the most direct therapeutic strategy to approach RNA and ASO technology has emerged as a powerful alternative to conventional small molecule approaches or gene replacement strategies for the treatment of genetic disorders. ASO are short, synthetic single-stranded DNA sequences designed to bind to target RNA by well-characterized Watson-Crick base pairing, and once bound to the target RNA, can modulate RNA function through a variety of post binding events. ASO-mediated gene silencing occurs either through degradative mechanisms, where the target RNA is cleaved by endogenous nucleases, or non-degradative mechanisms, where ASO binding sterically blocks or modulate translation, capping, or splicing. Genetic disorders in which a dominant mutation results in a toxic gain of function, such as liver diseases associated with aggregates in liver like Hereditary Transthyretin (TTR) Amyloidosis, are attractive to approach with ASO therapy. An ASO directed to TTR mRNA resulted in dramatic reductions in mutated TTR protein in mouse models and in normal volunteers. In addition, ASO targeting CTR1 is under evaluation as therapy for Wilson disease. ASO designed to CTR1 mRNA results in significant reduction in CTR1 levels resulting in reduction in copper levels, in the blood, brain and liver and improved liver pathology in a mouse model of Wilson disease. Furthermore, antisense oligonucleotides can be effective therapies using non-degradative mechanisms. For treatment of Spinal muscular atrophy (SMA) an ASO designed to correct SMN2 splicing, restores SMN expression and extended life span in SMN mouse model. Many additional genetic diseases are successfully being targeted with ASO technology for example; ASO inhibition of apolipoprotein B synthesis by Kynamro (mipomersen sodium) is an effective therapy to reduce LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia. This presentation will cover antisense technology platform and will include data from preclinical research and clinical trials with ASO drugs treating multiple genomic disorders.

2306F

Acute Hepatic Porphyrias: Inhibition of Hepatic ALAS1 with an RNAi Therapeutic Provides Effective Prevention and Treatment of Induced Acute Attacks in Acute Intermittent Porphyria Mice. *M. Yasuda[1], L. Gan[1], C. Yu[1], W. Querbes[2], A. Liebow[2], K. Fitzgerald[2], R.J. Desnick[1].* 1) Gen & Genomic Sci, Mount Sinai Sch Med, New York, NY; 2) Alnylam Pharmaceuticals, Cambridge, MA.

The acute hepatic porphyrias include four inherited disorders of heme biosynthesis that are characterized by life-threatening acute neurovisceral attacks. The episodic acute attacks are precipitated by various factors that increase the expression of hepatic 5-aminolevulinic acid synthase (ALAS1), the first and rate-limiting enzyme in the heme biosynthetic pathway, resulting in the marked accumulation of the neurotoxic porphyrin precursors, δ-aminolevulinic acid (ALA) and porphobilinogen (PBG). The acute attacks are currently treated with a series of intravenous hemin infusions, but a faster-acting and safer therapy is desirable, particularly for patients with frequent attacks. Therefore, we designed a liver-directed RNA interference (RNAi) therapeutic using small interfering RNAs (siRNAs) targeted specifically against *Alas1* (designated ALAS1-siRNAs) and evaluated its effectiveness following intravenous administration in a mouse model of acute intermittent porphyria (AIP), the most common acute hepatic porphyria. These mice have slightly elevated baseline urinary and plasma ALA and PBG levels which are markedly increased (~30- and >85-fold in plasma, respectively) by administration of the porphyrinogenic drug phenobarbital, thereby inducing a 'biochemical' acute attack. To identify a potent ALAS1-siRNA, >40 siRNAs targeting the *Alas1* gene were screened for their ability to reduce *Alas1* mRNA expression in cultured hepatic cells. The most active compounds were formulated into lipid nanoparticles for efficient hepatic delivery and evaluated for their ability to downregulate hepatic *Alas1* mRNA in wild-type mice. A single prophylactic administration of the optimal ALAS1-siRNA in the AIP mice completely prevented the phenobarbital-induced upregulation of hepatic *Alas1* mRNA/protein expression and the consequential increases of plasma and urinary ALA and PBG for approximately two weeks. When ALAS1-siRNA was infused during an induced acute attack, plasma ALA and PBG levels decreased within 8 to 12 hours, much more rapidly and effectively than a single hemin infusion. Subcutaneous injection of a hepatocyte *N*-acetylgalactosamine (GalNAc) receptor-targeted ALAS1-siRNA preparation was effective in preventing the phenobarbital-induced acute attacks, thereby facilitating convenient therapeutic delivery. These preclinical studies provide proof-of-concept for the development of siRNA-mediated therapy to prevent and treat the acute attacks in the acute hepatic porphyrias.

2307T

Silencing mutant Htt by AAV-mediated expression of RNAi ameliorates disease manifestations in the YAC128 mouse model of Huntington's disease. *L.M. Stanek, P.S. Sardi, B. Mastis, A. Richards, S.H. Cheng, L.S. Shihabuddin.* Neuroscience, Genzyme a Sanofi Company, Framingham, MA.

Huntington's disease (HD) is a fatal autosomal dominant neurodegenerative disease that results from an expansion of polyglutamine residues in the huntingtin (Htt) protein. With the identification of the underlying basis of HD, therapies are being developed that reduce the expression of the causative mutant Htt. RNA interference (RNAi) that seeks to selectively reduce the expression of such disease-causing agents is emerging as a potential therapeutic strategy for treating dominant disorders. This study examines the merits of administering a recombinant adeno-associated viral (AAV) vector designed to deliver a shRNA to target and lower Htt levels in the striatum, at correcting the biochemical and behavioral deficits associated with the YAC128 mouse model of HD. We showed that AAV-mediated shRNA expression partially reduced the levels of wild-type and mutant Htt in the striatum. Concomitant with these reductions were significant improvements in both motor and behavioral deficits in YAC128 mice. Lowering mutant Htt levels was also associated with a reduction in the number of Htt aggregates that accumulated in the striatum of YAC128 mice. These results are supportive of the development of AAV-mediated RNAi as a therapeutic strategy for the treatment of HD.

2308F

Hematopoeitic-stem-cell based therapy for HIV disease. *J. Bodor, J. Chalupnikova, K. Pavelcova, J. Jencik, R. Klubal, gp120 Consortium.* Czech Gene Bank, Prague, 4, Czech Republic.

Our aim is to develop a novel technology platform to cure HIV disease. The major innovation is to use hematopoetic stem cell (HSC) transplantation of the cells resistant to HIV (such as CCR5 delta 32 cells spontaneously occuring in 4-15% of EU population) while improving the outcome of engraftment of HIV resistant cells in HIV patients with AIDS lymphoma or leukemia by suppressing graft-versus-host disease using gp120. This is based on novel evidence for the cure of HIV infection by CCR5 delta 32 stem cell transplantation ('Berlin patient' reported in Allers et al. 2011, Blood 117: 2791) and by novel insights of the protection from graft-versus-host disease by HIV-1 envelope protein gp120-mediated in human regulatory T cells by elevated levels of cAMP (Becker et al. 2009, Blood 114: 1263; Vaeth et al. 2011, PNAS 108: 2480; Klein et al. 2012, J Immunol 188: 1091). The concept of allogeneic stem cell transplantation of the cells resistant to HIV using regimen improving function of regulatory CD4+ T cells favours suppressing graft-versus-host disease while retaining the benefits of graft-versus-leukemia effect. Therefore, the concept constitutes an outstanding example of cure rather than indefinite anti-retroviral treatment. The unprecedented power of the hematopoetic stem cell transplantation of the CCR5 delta 32 cells resistant to HIV has been proven to cure HIV infection in the case reported last year ('Berlin patient'). Successful reconstitution of CD4+ T cells at the systemic level as well as in the gut mucosal immune system was found while the patient remained without any sign of HIV infection. Furthermore, during the process of immune reconstitution, evidence for the replacement of long-lived host tissue cells with donor-derived cells indicates that the size of the viral reservoir has been reduced over the time. Next step is to create database of CCR5 delta 32 donors in Stem Cell Bank and further optimize treatment of graft-versus-host disease by gp120 in phase I clinical trials at Transplantation Unit. Clients are HIV patients with AIDS-related lymphoma or specific forms of leukemia. Side effects of HSC against HIV will be minor and restricted to graft-versus-host disease. HSC against HIV treatment costs are in the same range as the costs of HSC therapy, but considerably less expensive than lifelong anti-retroviral (ART) therapy.

2309T

Lipid storage and impaired function in iPSC and monocyte - derived Gaucher macrophages are reversed with a non-inhibitory chaperone. *E. Sidransky[1], E. Aflaki[1], N. Tayebi[1], B. Stubblefield[1], E. Maniwang[1], G. Lopez[1], E. Goldin[1], S. Patnaik[2], J. Marugan[2].* 1) Molec Neurogen, 1A213, MGB/NHGRI/NIH, Bethesda, MD; 2) NIH Chemical Genomics Center, NIH Center for Translational Therapeutics, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA.

Gaucher disease (GD), the inherited deficiency of glucocerebrosidase, manifests with lysosomal glycolipid storage, primarily in macrophages. Chemical chaperone therapy has been developed as a means to increase glucocerebrosidase levels in patients. Since improved cell models of GD exhibiting a storage phenotype were needed to facilitate drug development, monocyte and induced pluripotent stem cell- (iPSC) derived-macrophages from patients with GD were generated and characterized. Both patient (n= 20) and iPSC-derived macrophages (n=5 lines) share the same phenotype and exhibit reduced glucocerebrosidase activity. Storage, enhanced with tagged glycolipids and erythrocyte ghosts from patients with GD, was confirmed by lipid quantification and fluorescence microscopy. The Gaucher macrophages showed efficient phagocytosis, while efferocytosis of dead cells was impaired. Intracellular reactive oxygen species production was reduced and phagosome maturation altered, indicating that Gaucher cells have defective digestion of phagocytes. Using high-throughput screening with patient spleen as the source of mutant enzyme, several non-inhibitory, small molecule chaperones for glucocerebrosidase were identified. The lead compound reversed the disease phenotype, enhanced glucocerebrosidase specific activity, reduced lipid storage and normalized the efferocytic index. Our new capability to evaluate alterations in models closely mimicking the disease phenotype should greatly enhance our understanding of disease pathogenesis and facilitate the development and characterization of new treatments for Gaucher disease.

2310F

MicroRNA-486 overexpression delays the disease pathology of dystrophin-deficient muscle. *M.S. Alexander[1], J.C. Casar[1], N. Motohashi[1], N.M. Vieira[1], M.J. Gasperini[1], J.A. Myers[1], E.A. Estrella[1], P.B. Kang[1,5], F. Shapiro[6], G. Kawahara[1], E. Gussoni[1], L.M. Kunkel[1,2,3,4].* 1) Genetics, Boston Children's Hospital/Harvard Medical School, Boston, MA; 2) The Manton Center for Orphan Disease Research at Boston Children's Hospital, Boston, MA; 3) Department of Pediatrics and Genetics at Harvard Medical School; 4) Harvard Stem Cell Institute, Cambridge, MA 02138; 5) Department of Neurology, Boston Children's Hospital and Harvard Medical School; 6) Departments of Orthopedic Surgery at Boston Children's Hospital and Harvard Medical School, Boston MA 02115.

Duchenne Muscular Dystrophy (DMD) is caused by mutations in the dystrophin gene that result in the dysregulation of many signaling pathways that interact directly or indirectly with the dystrophin protein. Previously, we identified miR-486 as being strongly reduced in its expression levels in the dystrophin-deficient mouse and muscle biopsies of human DMD patients. Here we report that transgenic overexpression of the muscle-enriched microRNA, miR-486, in mdx5cv (dystrophin-mutant) mice resulted in improved serum biochemistry, reduced apoptosis, increased myofiber size, and improved muscle regeneration following injury. Transient overexpression of miR-486 also resulted in similar improved muscle physiology in the mdx5cv mouse. Using a bioinformatic approach, we identified DOCK3, dedicator-of-cytokinesis-3, as being a direct downstream target of miR-486 in skeletal muscle. Manipulation of DOCK3 expression in myoblast cell culture had strong effects on normal and DMD myoblast proliferation, apoptosis, and differentiation capabilities. Together, these studies demonstrate that stable overexpression of miR-486 delays many of the signs of the disease pathology of dystrophin-deficient muscle.

2311T

Epigenetic upregulation of Survival of Motor Neuron 2 (SMN2) protein by oligonucleotides targeting long non-coding RNA. *K.M. Hussey[1], R. Davey[1], E. Brand[1], D. Knowlton[1], B. Schwartz[1], R. Subramanian[1], J. McSwiggen[1], A. Krieg[1], J. Keil[2], H. Cardona[2], C. DiDonato[2], J. Barsoum[1].* 1) RaNA Therapeutics, Cambridge, MA; 2) Northwestern University & Lurie Children's Hospital of Chicago Research Center, Chicago, IL.

Spinal Muscular Atrophy (SMA) is a motor neuron disease affecting 1 in 6,000 children. This inherited disease leads to muscle weakness and atrophy and is mainly caused by a deleted or defective survival of motor neuron 1 (SMN1) gene. The SMN1 (telomeric) protein is a key player in cellular splicing events as a member of the snRNP complex in spinal cord neurons. A nearly identical gene, SMN2 (centromeric), is expressed in the same cells, albeit at very low levels due to a single point mutation that results in defective splicing. SMN2 is able to functionally compensate for SMN1 and support a milder phenotype when SMN2 copy number is increased, suggesting SMN2 as a primary therapeutic target. RaNA's goal is to specifically up-regulate the transcripts derived from the SMN2 gene, ultimately increasing functional SMN2 protein and improving SMA disease conditions.

Our platform is based on a recently discovered class of long noncoding RNAs (lncRNAs) that interact with the Polycomb Repressive Complex 2 (PRC2) and act in cis to repress gene transcription. RaNA's approach is to utilize short synthetic oligonucleotides (oligos) to interfere with this lncRNA-PRC2 interaction specifically at the SMN2 locus, relieve SMN2 repression, and drive transcription of the SMN2 gene. To this end, we have identified several oligos that up-regulate full length SMN2 mRNA 3-5 fold and protein by 3-7 fold above baseline in SMA patient-derived fibroblasts. These oligos target lncRNA transcripts in the SMN2 gene locus. Our results indicate that RaNA oligos enhance transcription of full length SMN2 mRNA and functional protein as well as correct the defective splicing of the transcript. Preliminary studies in a humanized SMA mouse model suggest that RaNA oligos can positively affect preclinical outcomes.

Current in vivo studies are focused on determining the effects of RaNA oligos on various preclinical phenotypes and survival in humanized SMA mouse models. Furthermore, we are investigating global- and locus-specific epigenetic and transcriptional changes to understand the mechanisms of RaNA oligo action. Collectively, our data suggests a novel epigenetic mechanism to increase SMN2 protein production that will likely provide a significant therapeutic benefit to SMA patients.

2312F

Towards the therapy for Angelman syndrome. *L. Meng[1], A. Award[2], F. Rigo[2], A. Beaudet[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Isis Pharmaceuticals, Carlsbad, CA.

Angelman syndrome (AS) is a severe neurodevelpmental disorder caused by maternal deficiency of the imprinted gene *UBE3A*. Although the molecular mechanisms of maternal deficiency are diverse, all AS patients carry at least one copy of paternal *UBE3A*, which is silenced but intact. By correcting the expression level of *UBE3A* via activating the silenced paternal allele, the disease might be treated. *UBE3A-ATS* is the antisense transcript of *UBE3A* that negatively regulates its expression. By truncating this antisense RNA in mice, we showed that depletion of *Ube3a-ATS* is sufficient to activate expression of paternal *Ube3a* and rescue phenotypic defects in the Angelman syndrome mouse model, including motor defects, cognitive deficit, and impaired long-term potentiation. Next, to develop specific therapy for Angelman syndrome, we sought to activate paternal *Ube3a* by knocking down *Ube3a-ATS* with antisense oligonucleotides (ASOs). With screens performed in cultured mouse neurons, we successfully identified several mouse specific ASOs that can achieve full activation of paternal *Ube3a*. Different from topoisomerase inhibitors, ASOs we identified are more site-specific, with little or minimal effect on nearby paternally expressed genes of *Snrpn*, *Snord115* and *Snord116*. Compared with PBS-treated animals, mice recieved intracerebroventricular injection of ASOs show significant knock-down of *Ube3a-ATS* and activation of paternal *Ube3a* in the cortex and other brain regions four weeks post injection, indicating that the ASOs function effectively *in vivo*. Future studies aim to achieve phenotypic improvement in the Angelman syndrome mouse model following CNS administration of ASOs and identifying human-specific ASOs for treating Angelman syndrome.

2313T

Decoding the biochemistry of Snyder-Robinson Syndrome: Using yeast to model the polyamine biosynthetic defect. *J.S. Albert[1], M.K. Chattopadhyay[2], W. Bone[1], L.A. Wolfe[1], W.A. Gahl[1], C.F. Boerkoel[1].* 1) NHGRI, National Institutes of Health, Bethesda, MD; 2) NIDDK, National Institutes of Health, Bethesda, MD.

Polyamines are simple, positively charged, ubiquitous molecules that interact with anionic compounds such as DNA, RNA, and ATP. Homeostasis of the polyamines putrescine, spermidine, and spermine is essential to cell growth and survival and is maintained primarily by the biosynthetic enzymes ornithine decarboxylase (ODC), S-adenosylmethionine decarboxylase (AdometDC), spermidine synthase (SRM) and spermine synthase (SMS). The balance of spermine and spermidine is crucial for proper chromatin structure, ion channel regulation, transcription and translation. Disruptions in polyamine homeostasis are associated with Snyder-Robinson syndrome (SRS), an X-linked intellectual disability syndrome first reported in 1969. SRS results from a mutation in SMS and produces a complex multi-systemic phenotype including intellectual disability, dysmorphic facies, muscle hypotonia, kyphoscoliosis, osteoporosis, seizures, and speech and gait abnormalities. The mechanism whereby dysregulation of polyamine homeostasis results in SRS is unknown. We are using an *S. cerevisiae* knockout model of *SMS* (*SPE4*) to study the polyamine biosynthetic pathway. Using HPLC analysis, we have found that SPE4 knockout yeast strains, compared to wild-type, have only a 1.3 fold increase in spermidine under physiologic conditions and only a 1.5 fold increase in spermidine in the presence of supra-physiologic (10mM) spermidine concentrations. This attenuated increase in spermidine is similar to that observed in individuals with SRS. Spermine is undetectable in the *SPE4* knockout strain. We also found elevated levels of methylthioadenosine (MTA), a product of the biosynthetic pathway and an inhibitor of the upstream enzyme, spermidine synthase, in the cell extract of the *SPE4* knockout, but not in the media. This suggests that increase in MTA could serve to protect the cells from accumulating deleterious amounts of spermidine. These results suggest that accumulation of spermidine and deficiency of spermine are not the cause of SRS. We discuss this as well as our ongoing studies on the possible biochemical etiology for SRS.

2314F

Secondary Coenzyme Q10 deficiency in patients with BRAF mutations. *E. Trevisson[1], M.A. Desbats[1], C. Cerqua[1], M. Doimo[1], A. Casarin[1], C. Santos-Ocaa[2], P. Navas Lloret[2], L. Salviati[1].* 1) Woman and Child Health, Clincal Genetics Unit, Univ Padova, Padova, Italy; 2) Centro Andaluz de Biologia del Desarrollo, Universidad Pablo de Olavide, Sevilla, Spain.

Coenzyme Q10 (CoQ) is a small lipophilic molecule that plays fundamental cellular functions: it acts as an electron transporter between complexes I and II of the mitochondrial respiratory chain and complex III, it functions as an antioxidant and as a cofactor of several mitochondrial dehydrogenases, among which dihydroorotate dehydrogenase, an essential enzyme involved in pyrimidine biosynthesis. Coenzyme Q10 is composed of a polar quinone group derived from tyrosine and a polyisoprene tail of different length in different species . In yeast the CoQ biosynthetic pathway requires at least 12 different proteins, encoded by COQ genes, that are organized into a multienzyme complex organized around Coq4p. In humans this process is still poorly understood. Coenzyme Q deficiency is a biochemical phenotype observed in a number of different clinical conditions. A genetic classification distinguishes primary deficiencies, in which the reduction in CoQ content is caused by mutations in genes directly involved in the CoQ biosynthetic process (PDSS1, PDSS2, COQ2, COQ4, COQ6, ADCK3, COQ9), and secondary forms, which are caused by defects in genes unrelated to the CoQ biosynthetic pathway (ETF, ETFDH, BRAF,APTX). Biochemically, patients present with reduction of CoQ in skeletal muscle and in skin fibroblasts. in addition to a reduction of the enzymatic activities of respiratory chain complex I+III and II+III (two CoQ-dependent reactions). BRAF encodes for a serine/threonine kinase acting downstream to Ras and regulating the MAP kinase/ERK signaling pathway. As stated above, mutations in this gene causes cardio-facio-cutaneous (CFC) syndrome and have been associated to Coenzyme Q10 deficiency in a single patient (Aeby et al 2007). Here we described two patients affected by CFC syndrome due to BRAF mutations who display myopathic features and responded to CoQ supplementation. Measurement of CoQ from skin fibroblasts shows a reduction in ubiquinone content; also CoQ biosynthetic rate was decreased in these cells. We report the biochemical characterization of CFC patients fibroblasts. Our data indicate that CFC can be complicated by secondary CoQ deficiency and have important implications for the treatment of this condition.

2315T

Fine-mapping of the *Atp7a* promoter deletion and biochemical characterization in the *mottled-dappled* mouse model of Menkes disease. *M.R. Haddad[1], K. Patel[1], P. Sullivan[2], D.S. Goldstein[2], J.A. Centeno[3], S.G. Kaler[1].* 1) Molecular Medicine Program; NICHD, NIH, Bethesda, MD; 2) Clinical Neurosciences Program, NINDS, NIH, Bethesda, MD; 3) Division of Biophysical Toxicology, Joint Center for Pathology, Malcolm Grow Medical Clinic, Andrews Air Force Base, Camp Springs, MD.

The *mottled-dappled* (*mo-dp*) is a mouse model of X-linked recessive Menkes disease caused by a large, previously uncharacterized deletion in the 5' region of *Atp7a*, the mouse ortholog of human *ATP7A*. Affected mutants die *in utero* at E17 and show bending and thickening of the ribs, and distortion of the pectoral and pelvic girdles and limb bones (Kaler, *Nat Rev Neurol* 2011). We have embarked on a prenatal approach (Haddad et al, *Mol Ther Nucleic Acids* 2013) for rescue of *mo-dp* to provide additional proof-of-concept in support of translation to viral gene therapy for Menkes disease. To develop a reliable genotyping assay for the pre-clinical gene therapy study, we designed a custom 4x180K microarray on the mouse X chromosome and tiled 60-mer probes around the region of interest near the *Atp7a* locus. The average resolution (probe spacing) of the microarray in the vicinity of the deletion was 675 bp. We performed comparative genomic hybridization (CGH) using extracted DNA from normal and carrier *mo-dp* females. CGH revealed a deletion of ≈9.4 kb. We fine-mapped the deletion breakpoints using PCR and successfully amplified a junction fragment of 630 bp using the following primers: F-TGTCAGCCTTTCTGGGAAGT and R-CAGCACTGGGGAGGTTAAGA. Sequencing the junction fragment disclosed the exact breakpoint positions and determined that the mo-dp deletion is precisely 8,245 bp and includes 2 kb in the promoter region, exon 1 and nearly 6 kb of intron 1 of *Atp7a*. Western blot analyses of *mo-dp* heterozygote brains showed diminished amounts of full length Atp7a protein compared to normal, consistent with reduced expression due to the promoter region deletion on one allele. In addition, brain neurochemicals from heterozygous females showed statistically significant higher DA:NE and DOPAC:DHPG ratios compared to normal ($p=0.0128$ and 0.0034, respectively), consistent with partial deficiency of dopamine-beta-hydroxylase, a copper-dependent brain enzyme that converts dopamine to norepinephrine. Brain copper measurements are pending. Our results delineate the molecular details of the *mo-dp* mutation, enable a convenient genotyping assay for experimental therapeutic efforts, and disclose novel biochemical findings in heterozygous female carriers of this allele. The characterization and potential rescue of this mouse model will improve correlation of genotype with biochemical phenotype, as well as understanding of treatment response in the context of a null *Atp7a* mutation.

2316F

Defective autophagy limits IL-1 beta production by eliminating active inflammasomes in human Gaucher macrophages. *E. Aflaki, N. Moaven, B. Stubblefield, G. Lopez, E. Sidransky.* Medical Genetics Branch, NHGRI, Bethesda, MD.

Gaucher disease (GD), the most prevalent lysosomal storage disorder, is caused by mutations in GBA1, the gene encoding glucocerebrosidase (GCase). The disease is characterized by the presence of glucosylceramide (GlcCer)-laden macrophages in the liver, spleen and bone marrow. Although in GD GCase is deficient in all cell types, the phenotype manifests primarily in macrophages. There is a dynamic relationship between phagocytosis and autophagy, the evolutionarily ancient process of lysosomal self-digestion of organelles, apoptotic corpses and cytosolic pathogens. We investigated autophagy in macrophages derived from peripheral monocytes from ten patients with type 1 GD (G-Macs) and two induced pluripotent stem cell (iPSC)-derived Gaucher macrophage lines, all sharing the common GD genotype N370S/N370S. Increased P62 expression in these macrophages indicated impaired autophagy. Levels of both LC3-II and P62 were increased in G-Macs during erythrophagocytosis and inflammation, as demonstrated by inflammasome activation with LPS and ATP. Moreover, expression of VPS34 and Rab7, proteins important in the fusion of early endosomes to late endosomes, decreased significantly. Autophagosomes deliver phagocytes to lysosomes for degradation, whereas inflammasomes are activated by infection, which in turn regulates the activity of caspase-1 and the maturation of interleukin (IL)-1 beta and IL-18. Macrophage polarization, evaluated using PCR arrays, indicates that G-Macs manifest the alternatively activated phenotypes in the presence of LPS with ATP. Importantly, G-Macs show increased expression of IL-18 and IL-10, but decreased IL-1 beta expression. Our data indicates that impaired secretion of some inflammatory cytokines results from the suppression of autophagy, resulting in increased Atg16L in G-Macs. This leads to inhibition of secretion and cleavage of IL-1β in these cells in the presence of LPS and ATP. This study demonstrates that lysosome dysfunction leads to the suppression of autophagy, followed by inhibition of inflammasome activity.

2317T

Resistin increases the expression of aggrecanases in human chondrocytes. *K.O. Yaykasli[1], E. Yaykasli[2], E. Kaya[3], M. Ozsahin[4], M. Uslu[5].* 1) Department of Medical Genetics, Medical Faculty, Duzce University, Duzce, Turkey; 2) Department of Medical Biology and Genetics, Institute of Health Science, Duzce University, Duzce, Turkey; 3) Department of Medical Pharmacology, Medical Faculty, Duzce University, Duzce, Turkey; 4) Department of Physical Medicine and Rehabilitation, Medical Faculty, Duzce University, Duzce, Turkey; 5) Department of Orthopedics and Traumatology, Medical Faculty, Duzce University, Duzce, Turkey.

Objective: The adipokine, resistin, was discovered in 2001 and expressed by not only adipose tissue but also by macrophages in high levels. First, resistin has been linked to insulin resistance, but further studies have linked it to other physiological processes, such as inflammation and immunity. Rheumatoid arthritis (RA) is an inflammatory and autoimmune disease that is characterized by irreversible destruction of the extracellular matrix (ECM). Aggrecanases are the primary enzymes responsible for the degradation of ECM. Several current investigations have found that elevated resistin levels are associated with aggrecanase expression. The aim of this study was to investigate the putative roles of resistin in aggrecanases upregulation. Material and Method: Human articular chondrocytes were cultured with resistin at certain concentrations (100, 250, and 500 ng/ml). After incubation for 12h and 24h, total RNA was extracted using TriPure reagent, and 2 µg of total RNA was reverse-transcribed by random primer. The ADAMTS1, ADAMTS4, ADAMTS5, ADAMTS9, and β-actin genes expression were analyzed by quantitative real-time polymerase chain reaction. Results and Conclusion: Resistin increased the expression of ADAMTS1, ADAMTS4, ADAMTS5, and ADAMTS9 genes at 100 ng/ml and 250 ng/ml doses. The irreversible destruction of ECM by aggrecanases is a key process in RA. An elucidation of the underlying causes for elevated levels of aggrecanases may be beneficial to understanding the etiology of RA. In conclusion, resistin with inflammatory properties is involved in RA by increasing the expression level of aggrecanases. However, further studies are needed to clarify the action mechanism of resistin.

2318F

Small molecule inhibition of glucosylceramide synthase affects bone remodeling in mice. *A.J. Leger, Z. Luo, I-H. Wu, N.S. Yew, S. Ryan, K. Malley, L. Sweet, S. Schiavi, T.K. Sampath, S.H. Cheng.* Genzyme, a Sanofi Company, 49 New York Avenue, Framingham, MA 01701-9322, U.S.A.

Patients with type I Gaucher disease (GD1) present not only with hematological and visceral manifestations, but multiple bone abnormalities including osteopenia, Erlenmeyer flask deformity, bone crisis, and avascular osteonecrosis. A recent mouse model of GD1 generated via conditional knockout of glucocerebrosidase (Gba1) in hematopoietic and mesenchymal cells displayed GD1-like bone pathophysiologies including osteopenia, reductions in bone formation rate, and osteonecrosis. The skeletal phenotype present in this transgenic model was attributed to defects in both osteoblast proliferation and differentiation. Given the potential relationship between glucosylceramide (GL-1) and osteoblast activity, we sought to determine the impact of inhibiting GL-1 synthesis upon bone physiology in mice. We therefore administered the glucosylceramide synthase (GCS) inhibitor Genz-667161 (60 mg/kg/day) in chow to C57Bl/6 mice for a period of 6 weeks and subsequently examined trabecular bone in femurs and lumbar vertebrae via histomorphometric analysis. In distal femur we observed an increase in bone volume fraction with a concomitant decrease in trabecular spacing in mice treated with Genz-667161. In L4 lumbar vertebrae we also noted a decrease in trabecular spacing with an associated increase in trabecular number. These effects in lumbar vertebrae would be expected to provide an increase in bone strength. This preliminary data suggests that GL-1 and perhaps other glycosphingolipids may play a role in beneficial bone remodeling and provide the rationale for further studies that examine the effect of GL-1 upon bone formation and resorption.

2319T

Fat mass and obesity-associated (FTO) Protein interacts with CamKII and modulates the activity of CREB signaling pathway. *L. Lin, P. Jin.* Human Genetics, Emory University School of Medicine, Atlanta, GA.

Polymorphisms in the fat mass and obesity-associated (FTO) gene have been associated with obesity in humans. Alterations in FTO expression in transgenic animals affect body weight, energy expenditure and food intake. FTO is a nuclear protein and its physiological function of FTO remains largely unknown. To understand the molecular functions of FTO, we performed yeast two-hybrid screen to identify the protein(s) that could directly interact with human FTO protein. Using multiple assays, we demonstrated that FTO interacts CaMKII. FTO interacted with three isoforms of CaMKII: α, β and γ, which are protein kinases that phosphorylate a broad range of substrate. We found that the overexpression of FTO could delay the dephosphorylation of CREB in human neuroblastoma (SK-N-SH) cells. Expression levels of NPYY1 and BDNF, which are targets of CREB, were dramatically increased due to the prolongation of CREB phosphorylation in the presence of FTO. Both NPYY1 and BDNF have been shown to regulate food intake and energy homeostasis. Thus, our results suggested that FTO could potentially modulate obesity by regulating the activity of CREB signaling pathway.

2320F

An infantile case of hepatomegaly, lactic acidosis, hypoglycemia, ketosis, and hyperlipidemia of unknown etiology. *Y. Watanabe[1,2], Y. Seki[1], T. Yanagi[1], T. Mizuochi[1], T. Takeuchi[1], J. Iwamoto[1], M. Yoshino[1], S. Watanabe[3], T. Inokuchi[2], S. Yano[4], K. Yoshiura[3], T. Matsuishi[1,2].* 1) Department of Pediatrics and Child Health, Kurume University, Kurume, Japan; 2) Research Institute of Medical Mass Spectrometry, Kurume University, Kurume, Japan; 3) Department of Human Genetics, Nagasaki University, Nagasaki, Japan; 4) Genetics Division, Department of Pediatrics, LAC+USC Medical Center, University of Southern California, Los Angeles, CA, USA.

Background: Inborn errors of energy metabolism including amino acid and fatty acid metabolism defects have recently become disorders screened by tandem mass spectrometry based newborn screening. The majority of patients with these disorders show positive results with characteristic profiles suggestive of a specific diagnosis. On the contrary, patients with carbohydrate metabolism and mitochondrial respiratory chain complex (RCC) defects often do not show positive newborn screening results with characteristic profiles. Hepatomegaly, hypoglycemia, ketosis, and lactic acidosis can be seen in patients with carbohydrate metabolism and RCC defects. We had a patient with infantile onset hypotonia, chronic massive hepatomegaly, hypercholesterolemia, chronic ketosis, mild developmental delay and episodic lactic acidosis: No diagnosis could be made on the enzyme assays. Objective: To recruit and perform exome studies on patients with similar clinical findings. Case report: The patient was a 6-month-old Japanese male with hepatomegaly (fatty liver), lactic acidosis, hyperlipidemia, hypoglycemia, ketonuria, and mild developmental delay. The patient was born to a 26-year-old Japanese mother and a consanguineous Japanese father (half sibling through their mother). Pregnancy was reportedly unremarkable. The patient was delivered at 37 weeks and weighed 2940g. Apgar scores were 8^1 and 9^5. The patient was fed with breast milk. At age 5 months, generalized mild hypotonia, failure to thrive and hepatomegaly (the edge was at 14 cm below RCM) were noted by his primary pediatrician. The following laboratory studies were performed: AST 267, ALT 57, Plasma amino acids-unremarkable, urine organic acid screening-mild~ moderate elevation of xanthine, glycerol, 3-mehtylglutaconic acid, plasma acylcarnitine- C2 (40.9 uM) and C6 (0.5 uM), Ca 12mg/dl, PO4 2.9mg/dl, FFA 1791uM, total ketones 2849uM, 3OHB 2211uM/AA 638uM, blood lactate 118mg/dl, uric acid 1.14mg/dl. Enzyme assays for glycogen storage diseases including phosphorylase, debranching enzyme, and phosphorylase b kinase were unremarkable. Fructose 1,6 Biphosphatase was unremarkable. Liver biopsy at age 7 months showed severe fatty liver of unknown etiology. Mitochondrial respiratory chain assays with liver tissues were unremarkable. Discussion: The patient's presentation was suggestive of glycolytic pathway defects including fructose metabolism defects. Exome studies may identify the gene(s) responsible for the findings.

2321T

Delineation of renal-independent disease mechanism underlying Hyperphosphatemic Familial Tumoral Calcinosis caused by GALNT3 mutations. *S.G. Ziegler[1], Y. Yang[2], D. Malhotra[2], M.T. Collins[3], N. Bhattacharyya[3], R.I. Gafni[3], W.A. Gahl[4,5], H.C. Dietz[1,6].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Genetic Disease Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, Bethesda, MD; 4) Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD; 5) Undiagnosed Diseases Program, National Institutes of Health, Bethesda, MD; 6) Howard Hughes Medical Institute, Chevy Chase, MD.

Hyperphosphatemic Familial Tumoral Calcinosis (HFTC) is characterized by progressive deposition of calcium phosphate complexes in soft connective tissues and periarticular spaces; serum phosphate levels are markedly increased. Loss-of-function mutations in three genes (*FGF23*, *GALNT3*, and *KLOTHO*) involved in renal phosphate excretion have been implicated in the pathogenesis of HFTC. FGF23, which is stabilized via O-glycosylation by GALNT3, is a circulating hormone that, with its co-factor klotho, decreases the expression of inorganic phosphate (Pi) transporter genes, *Npt2A* and *Npt2C*, in renal proximal tubule cells thus reducing Pi reabsorption. It has long been assumed that the calcification seen in HFTC is due to high levels of circulating Pi, which can then complex with calcium and precipitate. Here, we demonstrate that this calcification process can occur in an isolated, *in vitro* model system of ectopic calcification. Primary fibroblasts derived from two unrelated HFTC patients with biallelic *GALNT3* mutations can calcify *in vitro* when simulated with osteogenic media. These patients' cells also have increased tissue non-specific alkaline phosphatase (TNAP) and ENPP1 gene expression and enzyme activity; TNAP converts pyrophosphate (PPi) to inorganic phosphate (Pi), a positive regulator of calcification, while ENPP1 hydrolyzes ATP to AMP and PPi, a negative regulator of calcification. We have previously found that regulation of the PPi/Pi ratio may influence TNAP and ENPP1 levels, contributing to homeostasis. We found that HFTC cells have increased Gli-1 expression, a transcription factor that binds upstream of TNAP and increases target gene expression. To further explore the role of Gli-1 in ectopic calcification, we treated patient cells with a Gli-1 inhibitor (GANT58). GANT58 prevented calcification in our HFTC cell lines. In the literature, there is suggestive evidence that FGF23 can modulate Gli-1 expression via Gli-3. While the mechanism relating FGF23 and Gli-1 remains to be elucidated, we have demonstrated a renal-independent mechanism of ectopic calcification in HFTC and implicated the importance of Gli-1 signaling and TNAP upregulation. Further delineation of this pathway will help us better understand the basic mechanisms of ectopic calcification and potentially reveal new treatment strategies for HFTC and perhaps related disorders.

2322F

The mitochondrial F-box protein FBXL4 is necessary for efficient mitochondrial respiration. *C.A. Biagosch[1,2], X. Gai[3], D. Ghezzi[4], M.A. Johnson[5], H.E. Shamseldin[6], T.B. Haack[1,2], A. Reyes[5], M. Tsukikawa[7], C.A. Sheldon[7], S. Srinivasan[8], M. Gorza[1,2], L.S. Kremer[1,2], T.M. Strom[1,2], E. Place[7,9], S. Vidoni[5], L. Wong[10], M. Salih[11], E. Al-Jishi[12], C.P. Raab[13], F. Furlan[14], J.A. Mayr[15], V. Konstantopoulou[16], M. Huemer[17], T. Meitinger[1,2], P. Freisinger[18], W. Sperl[15], F.S. Alkuraya[6, 19], M.J. Falk[7], M. Zeviani[4,5], H. Prokisch[1,2].* 1) Institute of Human Genetics, Technische Universität München, Munich, 81675 Germany; 2) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, 85764 Germany; 3) Department of Molecular Pharmacology and Therapeutics, Loyola University Stritch School of Medicine Maywood, IL 60153, USA; 4) Department of Molecular Neurogenetics, Institute of Neurology Besta, 23888 Milan, Italy; 5) MRC Mitochondrial Biology Unit, CB2 0XY, Cambridge, UK; 6) Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Sauri Arabia; 7) Division of Human Genetics and Metabolic Disease, Department of Pediatrics, Children's Hospital of Philadelphia and University of Pennsylvania Perleman School of Medicine, Philadelphia, PA 19104, USA; 8) Department of Animal Biology, University of Pennsylvania School of Veterinary Medicine, Philadelphia, PA 19104, USA; 9) Ocular Genomics Institute, Massachusetts Eye and Ear Infirmary (MEEI), Harvard Medical School, Boston, MA 02114, USA;; 10) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA; 11) Department of Pediatrics, King Khalid University Hospital and College of Medicine, King Saud University, Riyadh, Saudi Arabia; 12) Salmaniya Medical Complex, Arabian Gulf University, Bahrain; 13) Department of Pediatrics, Nemours/ AI DuPont Hospital for Children, Thomas Jefferson University, Wilmington, DE 19803, USA; 14) Unit of Metabolic Disorders, Department of Pediatrics, Foundation MBBM/ San Gerardo University Hospital, 20900 Monza, Italy; 15) Department of Paediatrics, Paracelsus Medical University Salzburg, Salzburg 5020, Austria; 16) Department of Pediatrics, Medical University of Vienna, Vienna 1090, Austria; 17) Department of Pediatrics, LKH Bregenz, Bregenz 6900, Austria; 18) Department of Pediatrics, Klinikum Reutlingen, Reutlingen, 72764 Germany; 19) Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia.

Mitochondrial diseases are amongst the most common causes of childhood neurometabolic deficiencies. They comprise a group of conditions characterized by their extensive clinical and genetic heterogeneity, making identification of the underlying disease gene difficult. Poor phenotype genotype correlations result in a wealth of candidate genes and many more have yet to be discovered. The lack of obvious candidate genes hampers gene prioritization among more than 200 known mitochondrial disease genes and provides a model situation for large scale mutation screening. In this study, whole exome sequencing and autozygosity mapping in subjects with combined OXPHOS-deficiencies revealed 9 disease-segregating mutations in *FBXL4*. This gene encodes a member of the F-box family of proteins, some of which are involved in phosphorylation-dependent ubiquitination and/or G-protein receptor coupling, thought to be localized in the nucleus and cytosol. Surprisingly, we could localize FBXL4 to mitochondria using immunoflourescence and subcellular fractionation. Muscle and fribroblast tissue of patients displayed reduced activities of all respiratory chain complexes as well as mtDNA depletion. Biochemical rescue experiments based on the lentiviral-mediated expression of wildtype *FBXL4* was used to restore reduced oxygen consumption rate of OXPHOS-defective fibroblast cell lines. All individuals manifested early-onset lactic acidemia, hypotonia and developmental delay caused by severe encephalomyopathy, whereas some presented with dysmorphism, skeletal abnormalities, poor growth, gastrointestinal dysmotility, renal tubular acidosis, seizures and episodic metabolic failure. So far, the results suggest a role of FBXL4 in organell maintenance, which is currently under investigation.

2323T

Mitochondrial dysfunction in a patient with a mutation in the *MRPP1* gene encoding a subunit of the mitochondrial RNAseP complex. *M.D. Metodiev[1], Z. Assouline[3], M. Rio[3], N. Bahi-Buisson[4], A. Munnich[1,2,3], A. Rötig[1,2,3].* 1) INSERM U781 and Foundation Imagine, Paris, France; 2) Université Paris Descartes-Sorbonne Paris Cité, Paris, France; 3) Hôpital Necker-Enfants Malades, Paris, France; 4) Departament de pediatrie, Hôpital Necker-Enfants Malades, Paris, France.

Mitochondria synthesize ATP through oxidative phosphorylation in all eukaryotic cells. Unlike other organelles, mitochondria contain their own genome, mtDNA, which in humans encodes 11 messenger-RNAs, 2 ribosomal-RNAs and 22 transfer-RNAs. The messenger RNAs are translated into 13 polypeptides, which are essential for the assembly of OXPHOS complexes (C) I, III, IV and V. Transcription of mtDNA generates long, polycistronic transcripts which are processed by mitochondrial RNAseP (mt-RNAseP) and RNAseZ at the 5'- and 3'-ends of tRNAs, respectively. RNAseP is composed of three subunits: MRPP1, MRPP2 and MRP3. Knockdown of either subunit results in inactivation of RNAseP and a decrease in the levels of mature tRNAs, mRNAs and rRNAs. Interestingly, MRPP1 and MRPP2 also function to establish N1-methylation of purines at position 9 in several tRNAs. Notably, mutations in the methyltransferase domain of MRPP1, though eliminating its methylase activity, do not affect the processing activity of RNAseP indicating that both functions are independent.

We used whole-exome sequencing in order to identify mutations in novel nuclear genes associated with mitochondrial dysfunction. We focused on a series of patients with multiple respiratory chain deficiency with no qualitative or quantitative mtDNA anomaly. This approach enabled us to identify a patient carrying a homozygous mutation in the MRPP1 gene. This patient is the only girl born to consanguineous parents of Turkish origin. She presented feeding difficulties at one month of age. She then progressively developed trunk hypotonia, nystagmus and hypertrophic cardiomyopathy. She had hyperlactatemia. She died at 5 months of age. CI and CIV deficiency was found in muscle, liver and heart. Blue-native PAGE analysis revealed a quantitative defect of CI and CIV in cultured skin fibroblasts. Northern blot analysis of RNA isolated from mutant fibroblasts shows that steady-state levels of mitochondrial transcripts are unchanged indicating that the mutation does not affect the activity of RNAseP. Analyses of the G9- and A9-methylation in tRNAs by primer extension are ongoing. In conclusion, we report here the first example of abnormal RNAseP as a cause of respiratory chain deficiency. Moreover, our results support the view that exome sequencing is a powerful tool for the identification of disease gene mutations even in sporadic cases.

2324F

Gaucher disease: Clinical, biochemical and histopathological effects of 11 years of enzyme therapy in a chronic neuronopathic variant. *Y. Sun[1,3], T.A. Burrow[1,3], W. Zhang[2], B. Quinn[1], D.P. Witte[2,3], L. Bailey[1], C.E. Prada[1,4], G.A. Grabowski[1,3].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Division of Pathology and Laboratory Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH; 4) Centro de Medicina Genómica y Metabolismo, Fundación Cardiovascular de Colombia, Colombia.

Gaucher disease is caused by inherited mutations of *GBA1*, the resultant defective acid β-glucosidase (GCase) activity, and insufficient degradation glucosylceramide (GluCer) and glucosylsphingosine (GluS). Enzyme replacement therapy (ET) with GCase has major visceral effects, but limited efficacy in brain and lung. Here, the effects of 11 years of ET on brain and visceral tissues were examined from autopsy of a chronic neuronopathic Gaucher disease patient. The nearly complete correction of the hepatic, splenic, and bone marrow disease contrasted with the large clusters of CD68 positive Gaucher cells in the interstitia/alveolae of lungs, and, ubiquitously, within the periadventitial spaces of brain blood vessels, i.e., macrophages. In the brain, a few such macrophages were within the cerebellar, basal ganglial and hippocampus parenchyma. Only few perivascular macrophages were positive for the macrophage mannose receptor (MMR). In addition to large losses of neurons and absence of Purkinje cells, the above brain regions were positive for GFAP (astrogliosis), phosphorylated Tau, and synuclein, indicating proinflammation and neurodegeneration. GluCer and GluS accumulations were in massive mesenteric lymph nodes and lungs, but not in the liver. In lung and lymph nodes, major GluCer species were GC16:0, GC24:0 and GC24:1. In brain, cortical gray matter, basal ganglia, and hippocampus had large GluCer and GluS accumulations. In the white matter enriched regions, brainstem and spinal cord, GluCer and GluS levels were very low. This may reflect cell type (neuron and microglia vs. oligodendrocyte) specific GluCer metabolism. In basal ganglia and hippocampus, GluCer species contained the long acyl chain GC24:0 and GC24:1, indicative of their visceral origin and presence in the predominant perivascular Gaucher cells that were unaffected by ET. These unique data show that long-term ET with mannose-terminated GCase was corrective for liver, spleen, and bone marrow. In comparison, ET had limited access to lung, lymph nodes, and brain. Surprisingly, the perivascular storage cells, which are on the vascular side of the blood brain barrier, were negative for the MMR, thereby blunting the effects of MMR-targeted ET, which has significant impact for therapies on either side of the blood brain barrier. In particular, the extra- and intra-parenchymal CNS Gaucher cells indicates the need for non-ET based therapies for global correction in the brain.

2325T

Intracellular itinerary of normal and mutant ATP7A in polarized motor neurons. *L. Yi, S.G. Kaler.* Molecular Medicine Program, NICHD/NIH, Bethesda, MD.

The P-type ATPase ATP7A regulates cellular copper homeostasis by its activity at the trans-Golgi network and plasma membrane, with location normally governed by intracellular copper concentration. In addition to causing Menkes disease, defects in ATP7A may lead to the disease variants occipital horn syndrome, and ATP7A-related distal motor neuropathy, a newly discovered adult-onset condition that resembles Charcot-Marie-Tooth type 2 disease and for which the precise pathophysiology remains obscure. We previously characterized the two ATP7A motor neuropathy mutations (T994I, P1386S) and identified molecular mechanisms for abnormal intracellular trafficking of ATP7A in affected patients' cells (Yi et al., *Hum Mol Genet* 2012). Our subsequent experiments have demonstrated that knockdown of the adaptor protein complex 1 (AP-1) leads to preferential plasma membrane localization of ATP7A under normal copper conditions. Both an ATP7A truncation mutant with deletion of the carboxyl-terminal 14 amino acid residues (including the di-leucine motif), and a di-leucine ATP7A mutant (LL to AA at residues 1487/1488) recapitulated the PM localization seen with AP-1 knockdown. In contrast, adaptor protein complex-2 (AP-2) knockdown resulted in failure to retrieve ATP7A from the plasma membrane after cellular copper depletion. Immunoprecipitation experiments confirmed the physical interaction of AP-1 and AP-2 with wild-type ATP7A. These interactions were each abrogated by the P1386S mutation, consistent with our prior finding of destabilized 8th transmembrane helix insertion associated with this allele. Taken together, our data implicate AP-1 and AP-2 in normal ATP7A trafficking by binding the carboxyl-terminus dileucine motif. Motor neurons are polarized cells with somatodendritic and axonal segments. We found that wild type ATP7A localizes in the somatodendrite region of differentiated NSC-34 motor neurons whereas the P1386S allele mainly localizes in axons. Taken together, our results suggest that AP-1 normally retains ATP7A at the trans-Golgi network in the somatodendritic segment of motor neurons, that disruption of this interaction results in release of ATP7A to the axons or axonal membranes, and that the latter effect is intensified by the loss of interaction with AP-2, thus disturbing normal directional polarity and leading to motor neuron disease.

2326F

Characterization of an ACAD10 Deficient Mouse Model: Pathological and Biochemical Analyses. *K. Kormanik[1], D. El Demellawy[2], A-W. Mohsen[2], A. Karunanidhi[2], M. Reyes-Mugica[2], J. Vockley[1,2].* 1) University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA 15224; 2) University of Pittsburgh School of Medicine, Pittsburgh, PA 15224.

Acyl-CoA dehydrogenase 10 transcript (ACAD10) codes for 1059-aa peptide with a ~400-aa C-terminus domain homologous to members of the ACAD flavoenzymes family involved in oxidation of fatty acids and intermediates of aa metabolism. A polymorphism in the ACAD10 gene has been linked to obesity in Pima Indians, but its physiological role remains unknown. The objective is to identify the role of ACAD10 in physiology and its link to diabetes. We have generated an ACAD10 deficient mouse and an anti-ACAD10 antibody to conduct pathological, biochemical, and molecular studies. Deficient animals are viable and fertile, but accumulate excess abdominal adipose tissue. Pathological studies reveal that ACAD10 deficient mice manifest an early inflammatory liver process and secondary splenic extramedullary hematopoiesis. Although skeletal muscle is histologically normal, deficient mice have elevated creatine kinase when fasted indicative of rhabdomyolysis. Metabolomics analysis identifies elevated levels of malonylcarnitine, succinylcarnitine, methylmalonylcarnitine, glutaroylcarnitine, adipoylcarnitine in deficient mouse urine samples. Immunological studies reveals antigen detected in pancreas, lung and brain and localized to peroxisomes. RNAseq analysis reveals a broad spectrum of changes in gene expression including genes involved in the response to oxidative stress. We've made significant progress towards identifying the role of ACAD10 in physiology.

2327T

Mutations in *HADHB*, which encodes the β-subunit of mitochondrial trifunctional protein, cause infantile onset hypoparathyroidism and peripheral polyneuropathy. *K. Yamada[1], M. Naiki[1,2], N. Ochi[3], Y. Kato[4], J. Purevsuren[5], R. Kimura[1], D. Fukushi[1], S. Hara[6], Y. Yamada[1], T. Kumagai[7], S. Yamaguchi[5], N. Wakamatsu[1].* 1) Dept Gen, Inst Dev Res, Aichi Human Serv Ctr, Kasugai, Aichi, Japan; 2) Dept Pediatr, Nagoya Univ Grad Sch Med, Aichi, Japan; 3) Dept Pediatr, Daini-Aoitori Gakuen, Aichi Pref Hosp Habilit Ctr, Aichi, Japan; 4) Inst Health Sci, Tokushima Bunri Univ, Tokushima, Japan; 5) Dept Pediatr, Med, Shimane Univ, Shimane, Japan; 6) Dept Pediatr, Toyota Memor Hosp, Aichi, Japan; 7) Dept Pediatr Neurol, Kobato Gakuen, Aichi Human Serv Ctr, Aichi, Japan.

The mitochondrial trifunctional protein (MTP) is a hetero-octamer composed of the four α- and four β-subunits, which catalyzes the final three steps of mitochondrial β-oxidation of long chain fatty acids. To date, only two cases with MTP deficiency have been reported to be associated with hypoparathyroidism and peripheral polyneuropathy. We report that sibling cases born to consanguineous parents have MTP deficiency associated with infantile onset hypoparathyroidism and peripheral polyneuropathy. Genome-wide linkage analysis and array CGH analysis eliminated known diseases associated with autosomal recessive forms of isolated hypoparathyroidism or hereditary sensorimotor neuropathy (Charcot-Marie-Tooth disease). Yet, we identified a homozygous mutation (c.1175C>T, [p.A392V]) in exon 14 of *HADHB* located at the determined disease locus in the subjects. Biochemical analysis revealed that the presented subjects had MTP deficiency. Structural analysis indicated that the identified A392V and the reported N389D mutations associated with hypoparathyroidism are both specifically located near the active site and affected the conformation of the β-subunit. The present study demonstrated that missense *HADHB* mutations could cause infantile onset hypoparathyroidism. Since MTP deficiency is a treatable disease, MTP deficiency should be considered when subjects have hypoparathyroidism as the initial presenting feature at infancy.

2328F

Thiamine pyrophosphate deficiency secondary to TPK-1 mutation presenting as Leigh's disease: diagnosis and management within a sibling pair. *J.I. Fraser[1,2], S. Yang[3], A. Vanderver[4], T. Chang[4], L. Cramp[4], K. Chapman[3], G. Vezina[5], P. Smpokou[3], U. Lichter-Konecki[3], D.J. Zand[3].* 1) Medical Genetics Branch/NHGRI/NIH, Bethesda, MD; 2) Department of Pediatrics, CNMC, Washington, D.C; 3) Division of Genetics and Metabolism, CNMC, Washington, D.C; 4) .Department of Neurology, CNMC, Washington, D.C; 5) Department of Radiology, CNMC, Washington, D.C.

Background: Thiamine pyrophosphate (TPP) is essential for pyruvate dehydrogenase complex, α-ketoglutarate dehydrogenase, and branched-chain ketoacid dehydrogenase function. Five patients have been reported with thiamine pyrophosphokinase-1 (TPK1) deficiency, with encephalopathy, extrapyramidal, bulbar, and cerebellar signs, and metabolic decompensation leading to death. Some patients received high dose thiamine therapy. Patients: We present a sibling pair with Leigh's disease, progressive hypotonia, regression, and chronic encephalopathy. Whole exome sequencing in the younger sibling demonstrated homozygous TPK1 mutations. The older sibling had died from progressive neurologic disease with metabolic strokes, and the younger sibling had progressive neurologic decline. MRI demonstrated putamen and thalamic abnormalities in the younger, with similar and progressive evolution in the older sibling. The patient was started on high dose thiamine, niacin, biotin, and α-lipoic acid; ketogenic diet was initiated to reduce metabolic burden. He subsequently demonstrated improvement in neurologic function, with reattainment of lost milestones. Conclusions: TPK1 deficiency should be considered in children with Leigh's disease without identifiable mutations, and this gene should be evaluated as a candidate for testing in commercially available sequencing panels. Reducing metabolic burden in these children by addition of supplements and transition to ketogenic diet may reverse some of the neurologic progression and improve outcome.

2329T

Deficiency of the mitochondrial phosphate carrier presenting with myopathy or cardiomyopathy: two new cases. *L.C. Pyle, E.J. Bhoj, M. Li, R.C. Ahrens-Nicklas, C. Ficicioglu, N. Sondheimer, M. Yudkoff.* Section of Metabolic Disease, Children's Hospital of Philadelphia, Philadelphia, PA.

The SLC25A3 mitochondrial phosphate transporter is required to facilitate transport of inorganic phosphate across the inner mitochondrial membrane, for production of ATP via oxidative phosphorylation. Isoform A is expressed in heart and muscle, while B is the predominant isoform expressed in other tissues. Deficiencies in this transporter have been reported in five individuals to-date, across two sibships. Here we report two new cases of pathologic SLC25A3 mutations, in two different families, both presenting in the neonatal period. Case 1 presented as neonatal primary lactic acidosis, with subsequent improvement when intubated, but recurrent failure to extubate. Mitochondrial next generation sequencing revealed a homozygous IVS-9 mutation in SLC25A3, which has been previously shown in previous family 1 to cause a novel splice site, affecting an isoform specific to heart and skeletal muscle. This patient also has hypertrophic cardiomyopathy. Case 2 presented as neonatal severe hypertrophic cardiomyopathy. Mitochondrial next generation sequencing revealed two novel predicted pathogenic variants in SLC25A3. These cases expand our understanding of the SLC25A3 presentation. They also highlight the relevance of mitochondrial studies in the setting of neonatal cardiomyopathy and neonatal primary lactic acidosis.

2330F

Use of next generation sequencing (NGS) in mitochondrial (mt) disorders: whole mitochondrial genome analysis. *S. Seneca[1,2], K. Vancampenhout[2], G. Van Dyck[2], J. Smets[3], A. Van Lander[3], R. Van Coster[3], D. Daneels[2], W. Lissens[1,2], L. De Meirleir[4].* 1) Center for medical genetics, UZ Brussel, Brussels, Belgium; 2) Vrije Universiteit Brussel (VUB, REGE), Brussels, Belgium; 3) University Hospital Ghent, Division of Paediatric Neurology and Metabolism, Ghent, Belgium; 4) UZ Brussel, Division of Paediatric Neurology, Brussels, Belgium.

Mitochondrial (mt) diseases are a heterogeneous group of disorders, caused by both nuclear and mt genome mutations. A correct diagnosis is challenging, mainly because of the absence of clear phenotype-genotype correlations, the existence of heteroplasmy (presence of at least two different mt genotypes in the same cell) and the very large number of genes involved. Current traditional molecular diagnoses for disorders caused by a mt DNA defect, relies on the identification of (common) point mutations and large deletions with sequencing and Southern blot procedures. Complete mt genome sequencing, using Sanger nucleotide sequencing techniques, is reserved only for few well selected patients. The method is laborious, and not very sensitive to detect nucleotide variations below 15-20% heteroplasmy. Next Generation Sequencing (NGS) is a booming technology, promising to be an accurate and cost effective method to investigate considerable amounts of DNA, including the complete mt genome of patients. DNA samples of thirty patients with a (suspected) mt disorder were sequenced using the Ion PGM™ sequencer. Their mt genome has previously been characterized. A 100% coverage of the whole mt genome was obtained, 99.6% of the variants were concordant with Sanger sequencing. In depth sequencing allowed a sensitive detection of (pathogenic) heteroplasmic variations in a range of 4 to 79%, both in patients and controls. A sensitivity down to 4% is a major advantage in comparison with dideoxy nucleotide analysis. In addition, large multiple and single deletions, with visualization of their breakpoints, were identified. This study shows that NGS will be playing a major role in the molecular diagnostics of mt DNA disorders.

2331T

Variability in the level of erythrocyte glucose uptake in two patients of Glut1 deficiency syndrome with the same SLC2A1 mutation. *N. Ishihara[1], J. Natsume[1], K. Yanagihara[2], Y. Fukuhara[3], K. Yamada[3], Y. Yamada[3], Y. Azuma[1], T. Negoro[1], N. Wakamatsu[3], K. Watanabe[1,4].* 1) Dept Pediatrics, Nagoya Univ Grad Sch Medicine, Nagoya, Japan; 2) Dept Dev Med, Osaka Med Ctr and Res Inst for Maternal and Child Health, Osaka, Japan; 3) Dept Genet, Inst Dev Res, Aichi Human Service Ctr, Aichi, Japan; 4) Dept Med Sci, Faculty Health and Med Sci, Aichi Shukutoku Univ, Aichi, Japan.

Purpose: Glucose transporter 1 deficiency syndrome (Glut1-DS) is a congenital metabolic disorder characterized by intractable seizures with early infantile onset, developmental delay, movement disorders and acquired microcephaly. The diagnosis is based on hypoglycorrhachia, impaired in vitro glucose uptake into erythrocytes, and heterozygous mutations in *SLC2A1* gene. Here we report two patients with the same *SLC2A1* mutation and different results in the erythrocyte glucose uptake to reveal clinical heterogeneity in Glut1-DS. Subjects and Methods: Subject 1 is a 15-year-old girl, born to healthy non-consanguineous parents. Atypical absence seizures and dystonic posturing were seen since 2 years of age, and astatic episodes after a long walk has seen since 10 years of age. Hypoglycorrhachia and decreased erythrocyte glucose uptake were observed. She was diagnosed with Glut1-DS and started on the ketogenic diet. Subject 2 is a 17-year-old girl, born to healthy non-consanguineous parents. Since 4 years of age, she showed myoclonus, ataxia, and loss of consciousness with drooling. Such events were mostly seen in the late afternoon, and recovered with eating. Hypoglycorrhachia was observed while erythrocyte glucose uptake was normal. We did not start ketogenic diet, because of undetermined Glut1-DS diagnosis. Results: We performed direct sequencing of DNA extracted from white blood cells of the two subjects, and identified the same mutation in *SLC2A1* (R330X). To ensure the function of R330X, we constructed wild type and R330X FLAG tagged *SLC2A1* expression vectors and transfected to HEK293 cells. Wild type *SLC2A1* expressed at plasma membrane while R330X stayed in cytoplasm. Thus *SLC2A1* harbouring R330X is non-functional. Discussion: This is a first report of a Glut1-DS patient with normal erythrocyte glucose uptake caused by R330X mutation. There is variability in the level of erythrocyte glucose uptake even in patients with the same *SLC2A1* mutation, and it might be misleading. As Glut1-DS is a treatable epileptic encephalopathy, early diagnosis and treatment should be necessary. Conclusion: Glut1-DS caused by heterozygous mutations may show various level of erythrocyte glucose uptake, thus mutation analysis of *SLC2A1* should be performed for correct diagnosis.

2332F

Suspected mitochondrial myopathies in the era of genomic medicine. *A.M. Atherton[1,2], S.F. Kingsmore[2], B.A. Heese[1,3], S.E. Soden[2,4], C.J. Saunders[2,3], E.G. Farrow[2], L. Willig[2,5], N.A. Miller[2], L.D. Smith[1,2,3].* 1) Division of Genetics, Children's Mercy Hospitals and Clinics, Kansas City, MO; 2) The Center for Pediatric Genomic Medicine, Children's Mercy Hospitals and Clinics, Kansas City, MO; 3) Pathology and Laboratory Medicine, Children's Mercy Hospitals and Clinics, Kansas City, MO; 4) Division of Developmental and Behavioral Sciences, Children's Mercy Hospitals and Clinics, Kansas City, MO; 5) Division of Pediatric Nephrology, Children's Mercy Hospitals and Clinics, Kansas City, MO.

Mitochondrial myopathies (MM), a heterogeneous group of inherited disorders with multisystem involvement, can be caused by genetic alterations in the nuclear or mitochondrial genome. Historically, individuals with a suspected MM were diagnosed based on blood and/or cerebral spinal fluid biochemical testing, neuroimaging or invasive muscle biopsy results. Nonspecific results often lead to the diagnosis of a "suspected mitochondrial myopathy". Extensive, generally expensive, molecular testing panels for genes causing MM are available but may also be uninformative. We explored the utility of exome sequencing for molecular diagnosis by analyzing whole exome sequences of 19 patients from 15 families with clinical symptoms suggestive of a mitochondrial disorder without a definitive diagnosis. Five of the patients (26%) were confirmed to have a primary MM, five patients (26%) were identified as having an unrelated disorder with secondary mitochondrial dysfunction, one patient (5%) has likely pathogenic variants in novel MM candidate genes and eight patients (42%) had negative testing results. Overall, 58% of patients received a molecular diagnosis utilizing whole exome sequencing. These results support the initial use of whole exome sequencing in patients suspected of having a MM as this testing modality is more practical, efficient, specific and cost effective.

2333T

Mucopolysaccharidosis IVA (Morquio A) Molecular Analysis: A Review of the Advantages and Limitations of Molecular Testing of GALNS in the Diagnosis of Morquio A. *M. Hegde[1], M. Al Sayed[2], A. Brusius-Facchin[3], H. Church[4], M.J. Coll[5], M. Fietz[6], L. Gort[5], F. Kubaski[3], D.K. Jin[7], L. Lacerda[8], S. Leistner-Segal[3], A. Morrone[9], S. Pajares[5], L. Pollard[10], I. Ribeiro[8], F.J. Tsai[11], K.L. Tylee[4], N. Miller[12].* 1) Dept Human Gen, Emory Univ Sch Med, Atlanta, GA; 2) Department of Medical Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 3) Laboratório de Genética Molecular, Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil; 4) Willink Biochemical Genetics, Central Manchester University Hospitals NHS Foundation Trust, Saint Mary's Hospital Oxford Road, Manchester, United Kingdom; 5) Sección de Errores Congénitos del Metabolismo-IBC, Servicio de Bioquímica y Genética Molecular, Hospital Clínic, CIBERER, IDIBAPS, Barcelona, Spain; 6) SA Pathology, Women's and Children's Hospital, North Adelaide SA, Australia; 7) Samsung Medical Center, Sunkyunkwan, University School of Medicine, Seoul, Korea; 8) Unidade de Bioquímica Genética, Centro de Genética Médica Jacinto Magalhães (CGMJM) do INSA, Porto, Portugal; 9) Molecular and Cell Biology Laboratory, Dept of Neurosciences, Psychology, Pharmacology and Child Health, University of Florence and Paediatric Neurology Unit and Laboratories, Meyer Children's Hospital, Florence, Italy; 10) Biochemical Genetics Laboratory, Greenwood Genetic Center, Greenwood, South Carolina, United States; 11) China Medical University Hospital, Taiwan; 12) BioMarin Pharmaceutical Inc., Novato, California, United States.

Mucopolysaccharidosis IVA (MPS IVA, Morquio A) is an autosomal recessive lysosomal storage disease. Mutations in the gene lead to deficient N-acetylgalactosamine-6-sulfatase (GALNS) activity. Laboratory diagnosis is commonly based on urinary excretion of keratan sulfate, decreased GALNS enzyme activity in vitro, which is the gold standard for diagnosis, and molecular analysis of the GALNS gene. Accurate and efficient diagnosis of Morquio A can be challenging. We focus our review here specifically on the use of molecular testing/mutational analysis, as performed by authors' institutions, to elucidate the challenges of molecular testing for GALNS and we recommend potential solutions. Situations in which mutations were or can be missed are explored. Recommendations for reducing the rate of missed mutations include: parental testing to confirm biallelic inheritance, analysis to identify cases of partial/whole GALNS gene deletions or uniparental disomy, and use of advanced methods other than DNA sequencing to find otherwise undetectable mutations. To support accurate diagnoses we recommend providing clear, interpretable molecular testing reports and suggest caution when using mutation prediction programs. A global, open-access, locus-specific database of GALNS mutations is also proposed to facilitate improved dissemination of information and to help reduce the number of mutations with unknown significance identified by GALNS sequencing.

2334F

Practical strategies for the identification of common mutations in Mucopolysaccharidosis IVA patients. *F. Kubaski[1], A.C. Brusius-Facchin[1-2], P.F.V Medeiros[3], R. Giugliani[1-2], S. Leistner-Segal[1-2].* 1) Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil; 2) 2-Universidade Federal do Rio Grande do Sul/ Post Graduation Program in Medicine: Medical Sciences; 3) 3-Universidade Federal de Campina Grande/ Hospital Universitário Alcides Carneiro, Campina Grande, Brazil.

Mucopolysaccharidosis IVA or Morquio A syndrome, is an autosomal recessive disorder caused by deficiency of lysosomal enzyme N-acetylgalactosamine-6-sulfatase (GALNS), which results in lysosomal storage of the glycosaminoglicans keratan sulfate and chondroitin-6-sulfate in tissues, causing clinical manifestations. The phenotypes vary from the classical form with severe bone dysplasia, corneal opacity, short trunk dwarfism, heart involvement with a life span of 20 to 30 years, to an attenuated form with mild bone involvement and normal life span. Both forms of Morquio A patients have normal intelligence. The prevalence of MPS IVA ranges from 1/76.000 to 1/640.000 live births. Objectives: To analyze and characterize the genotype of Brazilian patients with MPS IVA, through molecular study of recurrent mutations in GALNS gene, enabling the estimative of frequency and the establishment of a protocol for routine screening for these mutations. Methods: Molecular analysis of GALNS gene was performed in 48 Brazilian patients by ARMS-PCR to detect six recurrent mutations (p.G116S/ p.G139S/ p.N164T/ p.L307P/ p.S341R/ p.R386C) followed by amplification of coding regions by PCR of all exons and sequencing. Results: These mutations were found in 47.9% of our sample, which were present in 40.6% of the alleles. The most frequent mutation was p.S341R (16.7%), followed by p.R386C (11.5%), p.G116S (7.3%) and p.N164T (3.1%). The mutations p.G139S and p.L307P were found only once in two different patients. After screening for recurrent mutations, we started sequencing all exons of GALNS. Until the present moment we found 2 described mutations in homozygosis (p.V239F and p.R386H). Furthermore, three new mutations were identified: p.V16E, p.E51K and p.C165Y (all homozygous and born from consanguineous parents). Conclusions: In Brazil, mutation p.S341R was found only in patients from the Northeast region. The identification of heterozygous individuals within these families will be important for genetic counseling and for estimating the disease prevalence in this region. Further studies to identify the origin of this mutation, including haplotype and segregation analyses are in progress, and will be evaluated in conjunction with epidemiological data. Sequencing of the whole coding region of GALNS will be important to define the genotype in the remaining patients and to allow correlations regarding the patients' phenotype.

2335T

Oligosaccharides As Biomarkers For Mucopolysaccharidosis IVA. *G. Asif, L. Arthur, E. Woolley, X. Li, M. He.* Human Genetics, Emory University, Atlanta, GA.

The mucopolysaccharidoses (MPS) are a group of lysosomal storage disorders with deficiencies in lysosomal hydrolases that are needed for degradation of glycosylaminoglycans (GAGs). In this study, we describe a characteristic group of urinary oligosaccharides found in patients with various types of MPS, such as mono antennary monosialo oligosaccharide and bi-antennary disialo oligosaccharide as well as certain galactosylated oligosaccharides. We found that these groups of oligosaccharides could be used as first tier diagnostic screening in urine for majority of the known MPS subtypes, including MPS I, II, III, IVA, VI. It was reported previously that lysosomal hydrolases involved in GAG degradation, such as N-Acetylgalactosamine-6-sulfate sulfatase (GALNS), forms a lysosomal multi-enzyme complex with neuraminidiase, galactosidase and cathepsin A. Thus the elevation of these sialylated and galactosylated oligosaccharides in MPS patients could reflect the disruption of supramolecular organization of the lysosomal multi-enzyme complex. To explore whether these biomarkers have significantly different profile from normal profile, 125 control urine samples were analyzed and 13 known MPS IVA patients' free oligosaccharide profiles were compared. Our analysis reveals that the mono-antennary monosialo oligosaccharide and bi-antennary disialo oligosaccharide at m/z 1532.76 and 2547.6, respectively, were significantly higher in MPS IVA patients than their levels in controls. The concentration of these two oligosaccharide biomarkers in urine, once normalized with urine creatinine level, does not significantly differ between urine collected at different time points of the same day or at different time points through 6 month of the study in MPS IVA patients. Therefore, these urinary oligosaccharides may serve as reliable biomarkers for the screening and potentially disease monitoring in patients with MPS IVA.

2336F

Biochemical, Molecular, and Clinical Presentations of 3 Patients with X-linked Epsilon-Trimethyllysine Hydroxylase (TMLHE) Deficiency, A Disorder of Carnitine Biosynthesis Associated with Autism Spectrum Disorder. *M. Comeaux, C. Schaaf, A. Petel, Y. Yang, F. Scaglia, A. Beaudet, S. Elsea, Q. Sun.* Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Previous research showed that deletion of exon 2 in the gene encoding epsilon-trimethyllysine hydroxylase (TMLHE) may be a risk factor for non-dysmorphic autism. Biochemical determination of trymethyllysine (TML) and γ-butyrobetaine (GBB) in plasma and urine of affected individuals can be used as biochemical markers for diagnosis of this disorder. We present the biochemical, molecular, and clinical phenotypes of three new patients with TMLHE deficiency. Patient 1 is a 4 year-old boy with a diagnosis of autism spectrum disorder who presented to genetics due to a recent episode of regression. Chromosome microarray (CMA) analysis was normal in this patient; however, whole exome sequencing found a hemizygous c.961_962del (p.I321fs) mutation in the *TMLHE* gene. Furthermore, biochemical analysis showed a low level of plasma carnitine, and analyte testing showed normal plasma and grossly abnormal urinary TML levels. The patient was subsequently placed on L-carnitine supplementation (200 mg/kg/day, divided in BID dosing), and his symtomatology has shown marked improvement in joint attention, eye contact, and social interaction. Patients 2 and 3 are twins born at 37 weeks of gestation who presented at birth with congenital heart defects. CMA analysis found a 4.5 Mb deletion of 8p23.1, a region that contains over 20 genes and has previously been associated with congenital heart defects. Additionally, the CMA found a hemizygous deletion of Xq28 that encompassed exon 2 of the *TMLHE* gene with a minimum deletion size of 3 kb in both sibs. Subsequent biochemical testing of plasma and urine revealed extremely high TML, low GBB, and normal carnitine values. These findings were discussed with the family and close monitoring of the twins' development was recommended in light of the Xq28 finding. These cases highlight the benefits of biochemical testing in combination with CMA and whole exome studies toward early diagnoses and prevention of symptoms associated with autism. Furthermore, they emphasize the importance of including disorders of carnitine biosynthesis in the differential diagnosis of patients with autism spectrum disorder especially given that some of the associated phenotypes may be amenable to carnitine supplementation.

2337T

Detection of 4-hydroxy-2-oxoglutarate aldolase metabolites in urine for the diagnosis and monitoring of primary hyperoxaluria type III. *L. Hasadsri, P. Loken, D. Gavrilov, D. Matern, K. Raymond, P. Rinaldo, S. Tortorelli, D. Oglesbee.* Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, MN.

Background: Mutations in the HOGA1 gene, encoding the mitochondrial enzyme 4-hydroxy-2-oxoglutarate aldolase (HOGA), were recently discovered to cause what is now known as primary hyperoxaluria type III (PH3). PH3 is a rare, autosomal recessive disorder characterized by the onset of recurrent kidney stone formation and progressive nephrocalcinosis during childhood. Molecular analysis of HOGA1 is currently available on a research basis, but the functional consequences of many variants found in this gene are still poorly understood. Timely diagnosis of patients with PH3 thus remains a significant challenge. HOGA catalyzes the conversion of 4-hydroxy-2-oxoglutarate (HOG) into glyoxylate and pyruvate, the final step in the metabolism of hydroxyproline. Patients with PH3 have been reported to have increased excretion of HOG and other HOG-related metabolites in their urine. Here, we describe the development of a novel method for the detection of urinary metabolites in PH3 via gas chromatography-mass spectrometry (GC-MS). **Methods:** Urine samples were treated with hydroxylamine hydrochloride at pH 9 and extracted with 5:1 (v/v) ethyl acetate/ propan-2-ol, then evaporated to dryness under nitrogen. Residues were derivatized with BSTFA+TMCS and pyridine and analyzed on an Agilent 6890/5975 operating in scanning mode, with either helium or hydrogen as the carrier gas. **Results:** HOG was undetectable in urine by routine organic acid analysis, even when present at high concentrations and after oximation. When oximated in an alkaline solution and extracted under more polar conditions, however, trimethylsilyl derivatives of HOG were easily identified, as were oxalic, glycolic, glyceric, glyoxylic, and pyruvic acid. Utilizing hydrogen for the mobile phase reduced sample run time by approximately 50%, and resulted in significantly improved resolution of the oxalate and glyoxylate peaks. Sample preparation time was <5 hours, and analytical run time was 8 minutes. **Conclusions:** GC-MS analysis of urine can provide a rapid, sensitive, and non-invasive means for the diagnosis and biochemical characterization of primary hyperoxaluria type III. Next steps include optimizing the assay's extraction efficiency, and characterizing urine samples from individuals with a molecular diagnosis of PH3, asymptomatic heterozygous carriers, and patients with a strongly suspected diagnosis of PH3.

2338F

Unraveling mitochondriopathies by exome sequencing. *L.S. Kremer[1], T.B. Haack[1], R. Kopajtich[1], B. Haberberger[1], C.A. Biagosch[1], T. Wieland[1], T. Schwarzmayr[1], A. Walther[1], T.M. Strom[1], J.A. Mayr[4], W. Sperl[4], M. Zeviani[2], P. Freisinger[3], T. Klopstock[5], R.W. Taylor[10], A. Rötig[9], A. Munnich[9], U. Ahting[7], M.B. Hartig[1], J.A.M. Smeitink[8], G.F. Hoffmann[6], A. Lombes[11], T. Meitinger[1], H. Prokisch[1].* 1) Institute of Human Genetics, Technische Universität München and Helmholtz Zentrum München, Munich, Germany; 2) MRC Mitochondrial Biology Unit, Cambridge, United Kingdom; 3) Department of Pediatrics, Klinikum Reutlingen, Reutlingen, Germany; 4) Department of Pediatrics, Paracelsus Medical University Salzburg, Salzburg, Austria; 5) Department of Neurology, Friedrich-Baur-Institute, Ludwig-Maximilians-Universität München, Munich, Germany; 6) Department of Pediatrics, University Hospital, Heidelberg, Germany; 7) Department of Clinical Chemistry, Städtisches Klinikum München, Munich, Germany; 8) Nijmegen Centre for Mitochondrial Disorders, Department of Pediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 9) INSERM U393 and Service de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 10) Mitochondrial Research Group, School of Neurology, Neurobiology and Psychiatry, The Medical School, University of Newcastle upon Tyne, United Kingdom; 11) INSERM, UMRS 1016, Institut Cochin and AP-HP, Hôpital de La Salpêtrière, Service de Biochimie Métabolique et Centre de Génétique moléculaire et chromosomique, Paris, France.

Mitochondrial disorders are a genetically and clinically highly heterogeneous group of diseases characterized by faulty oxidative phosphorylation. Despite good progress in the field, most disease causing mutations still have to be identified. We applied whole exome sequencing in 250 unrelated individuals with juvenile-onset mitochondrial disorder. In a quarter of patients, we detected mutations in known disease genes. In another quarter of patients, we identified mutations in genes previously not associated with mitochondrial disorders. While mutations in some genes are rare like MGME1, the first exonuclease involved in mitochondrial replication (Kornblum et al., Nat. Genet. 2013), mutations in other genes are more frequent, with ACAD9 being the most common finding with 13 cases. New examples are ELAC2, a tRNA modifying enzyme involved in the translation of mitochondrial proteins, and FBXL4, a protein with yet unknown function associated with reduced mitochondrial protein content. Exome sequencing allows for comprehensive detection of disease causing mutations, most notably, it allows for the detection of mutations in genes not predicted to encode mitochondrial proteins. It therefore represents the method of choice for the molecular diagnosis of the heterogeneous group of mitochondriopathies for which it has permitted rapid identification of new genes. Further improvement of the sequencing technology holds promise for a further increase in yield by increasing coverage and detecting indels and copy number variants. However, several issues remain to be considered including how to tackle diseases caused by di- or oligogenic mutations with synergistic effect, how to annotate variants in non-coding regions and how to identify mutations with dominant effect. Nevertheless, in addition to shedding light on mitochondrial physiology, each newly identified gene augurs the possibility for new treatment options, as for example riboflavin supplementation in the case of mutations in the riboflavin transporter SLC52A2 or the flavoprotein ACAD9.

2339T

Diagnosis of adrenoleukodystrophy using liquid chromatography-mass spectrometry. *M. Kosuga[1], H. Nakajima[2], K. Kida[1], J. Fujimoto[2], T. Okuyama[1].* 1) Department of Clinical Laboratory Medicine, National Center for Child Health and Development, Tokyo, Japan; 2) Clinical Research Center,National Center for Child Health and Development, Tokyo, Japan.

Adrenoleukodystrophy (ALD) is caused by a mutation in the ABCD1 gene and results in the apparent defect in peroxisomal beta oxidation and the accumulation of the saturated very long chain fatty acids (VLCFA). Hematopoietic stem cell(HSCT) transplantation is the only treatment that can stop the demyelination that is the hallmark of the cerebral forms of the disease. In order to be effective, HSCT should be done at an early stage of the disease. Thus, development of a simple screening method is important to enable the identification of pre-symptomatic ALD patients before showing symptoms of brain involvement. ALD is associated with an accumulation of VLCFA, particularly cerotic acid (26:0) and its metabolite lysophosphatidylcholine (Lyso PC26:0), which means that these molecules could be biomarker targets in the detection of ALD. We have developed a mass screening method for ALD through the quantification of Lyso PC or saturated fatty acid with different lengths of acyl chain using a liquid chromatography-mass spectrometry (LC-MS) selective reaction monitoring system. Those lipids could be extracted from dried blood spots. The analysis of each of the lipids with LC-MS was completed within 5 minutes per sample, thus it would be an ideal method for screening of ALD.

2340F

Creatine Deficiency Due To Targeted Disruption Of Alanine:Glycine Amidinotransferase Leads To Learning And Memory Deficits. *W. Craigen[1, 2], Y. Lai[2], M. Costa-Mattioli[3], L. Stoica[3], P.J. Zhu[3], L. Mbye[1].* 1) Molecular & Human Genetics; 2) Pediatrics; 3) Neuroscience, Baylor College of Medicine, Houston, TX.

Phosphocreatine serves as an important reservoir of high-energy phosphate for ATP synthesis in tissues that have fluctuating demands for energy. Human disorders of creatine biosynthesis and transport exist, leading to intellectual disabilities, epilepsy, and poor growth. Creatine biosynthesis requires two enzymes, the first, L-arginine:glycine amidinotransferase, the product of the Gatm gene, catalyzes the rate-limiting transfer of the amidino group from arginine to glycine, yielding guanidinoacetate and ornithine. We generated a mouse strain lacking Gatm. Homozygous Gatm deficient mice have an almost complete absence of creatine and guanidinoacetate, and exhibit poor growth and muscle weakness. A consequence of creatine deficiency is the global reduction in skeletal muscle and brain mitochondrial respiratory chain activities, and a compensatory increase in mtDNA content and citrate synthase activity. Gene expression profiling reveals the activation of the integrated stress response. When challenged with the seizure-inducing drug kainate, creatine deficient mice are resistant to the induction of status epilepticus and epilepsy-associated mortality. When tested for associative learning by fear conditioning paradigms, creatine deficient mice have a marked deficit in learning, and are similarly compromised in the Morris water maze test. This animal model of creatine deficiency demonstrates the central role creatine metabolism plays in normal neurologic functions.

2341T

Somatic Mosaicism with Reversion to Wild Type in Congenital Disorder of Glycosylation, Type IIb. *H. Vega[1], L. Wolfe[1], DR. Adams[1], M. He[2], S. Rosenzweig[3], WA. Gahl[1], C. Boerkoel[1].* 1) National Human Genome Research Institute, National Institute of Health, Bethesda, MD; 2) Department of Human Genetics, Emory University, Atlanta, GA; 3) National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD.

Somatic mosaicism with reversion to wild type resulting in correction of inherited mutations is a rare phenomenon that results in the proliferative advantage of cells containing the wild type allele. We report on a family with two affected sibs with Congenital Disorder of Glycosylation (CDG-IIb) caused by mutations of mannosyl oligosaccharide glucosidase (MOGS). CDG-IIb, which is characterized by severe developmental delay, generalized hypotonia, dysmorphic features, hygammaglobulinemia and accumulation of tetrasaccharide (Glu3Man1) in urine, is reported as a single case of lethality in the neonatal period. Our patients survived the neonatal period and were 11 and 6 years old when evaluated. These two CDGIIb patients bear three different mutations: a c.370C>T nonsense mutation of exon 2 inherited from the mother and two missense mutations in exon 1, c.65C>A and c.329G>A, inherited from the father. We hypothesized that mutation reversion might account for the survival of these children. To test this, we cloned and sequenced PCR products from blood-derived genomic DNA and found six different alleles in addition to the maternal and paternal inherited alleles. These new alleles,which are absent in both parents, range from alleles free of mutations to alleles with the all three mutations in cis. We propose that this wide range of alleles was generated as a result of somatic recombination, and that somatic mosaicism for reversion to wild type allele explains the survival of these CDG-IIb patients.

2342F

The clinical and mutation spectrum of Korean patients with urea cycle disorders. *H.W. Yoo[1,2], B.H. Lee[1,2], G.H. Kim[2], J.M. Kim[2], Y.M. Kim[1], J.H. Kim[1].* 1) Department of Pediatrics, Asan Medical Ctr, University of Ulsan College of Medicine, Seoul, Korea; 2) Medical Genetics Center, Asan Medical Ctr, Korea.

Urea cycle disorder (UCD) is caused by the deficiencies of the six enzymes, carbamoyl phosphate synthetase (CPS), N-acetylglutamate synthetase (NAGS), ornithine transcarbamoylase (OTC), argininosuccinate synthetase (ASS), argininosuccinate lyase (ASL) and arginase (ARG) or mitochondrial ornithine transporter (ORNT1). This study was undertaken to review the clinical and molecular genetic characteristics of Korean UCD patients diagnosed in one institution. A total of 68 patients, 30 males and 38 females, from 66 unrelated families were diagnosed based on genetic and biochemical profiles during past 15 yrs. UCD manifested in neonatal period in 38 patients (64.4%), whereas 27 patients (45.8%) were late onset UCD. The remaining two patients (3.4%) were identified in presymptomatic periods. Thirty patients (44.1%) were diagnosed as OTC deficiency, 22 (32.4%) as ASS deficiency, 10 (14.7%) as CPS deficiency, 3 (4.4%) as ASL deficiency, 1 as argininemia, 1 as Hyperonithinemia-Hyperammonemia-Homocitrullinuria (HHH) syndrome, and 1 as NAGS deficiency. Mutation detection rate was 97.7 % (43/44 alleles) in the ASS1 gene, 90.0 % (24/27 alleles) in the OTC gene, and 100 % (6/6 alleles) in the ASL gene, whereas it was 55.6 % (10/18 alleles) in the CPS1 gene. Total 25 novel mutations were identified. Most patients presented as symptomatic hyperammonemia, whereas an argininemia patient as spastic diplegia without pronounced hyperammonemia, and 2 patients were identified in asymptomatic period by familial screening or tandem mass screening. Among 42 patients with hyperammonemia over 300 umol/L, 25 were managed by peritoneal dialysis or continuous renal replacement therapy, whereas 17 managed conservatively. The clinical outcomes were normal neurological development (19 patients, 32.2%), neurological deficit (26 patients, 44.1%), and death from hyperammonic encephalopathy (13 patients, 22.0%). Liver transplantations were performed in 3 patients. This study portrayed the full-spectrum of clinical and molecular genetic characteristics of UCDs in Korea. More efforts are needed for early identification and urgent management of each patient with UCD to improve the long-term neurological outcome.

2343T

Leigh syndrome and myoclonic epilepsy caused by novel mutation in AIMF1 gene. *T. Honzik[1], A. Vondrackova[1], V. Stranecky[2], M. Rodinova[1], H. Kratochvilova[1], H. Hansikova[1], P. Klement[1], M. Magner[1], S. Mazurova[1], J. Zeman[1], M. Tesarova[1].* 1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Czech Republic; 2) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Czech Republic.

Mutations in X-linked AIMF1 gene, encoding the Apoptosis-Inducing Factor Mitochondrion-associated 1, cause disruption of the respiratory chain complexes, decrease stability of mtDNA and increase apoptogenic stimuli. Only 11 patients from 3 families were reported, so far. Severe encephalomyopathy and Leigh syndrome was present in four children and progressive axonal neuropathy, deafness and cognitive impairment were present in 7 patients (Cowchock syndrome; CMTX4). Here we describe another patient with neonatal onset of the disease with hypotonia, myoclonic epilepsy, hearing loss, visual impairment, severe encephalopathy, Leigh syndrome and lactic acidosis. In addition, the boy had hyporeflexia compatible with axonal neuropathy and died at the age of 18 months. Muscle biopsy demonstrated mild myopathic pattern with accumulation of SDH product. In tissues obtained at autopsy, the activity and amount of complex I and IV was decreased in muscle, brain and heart and the activity of complex II was upregulated. Targeted sequencing of mitochondrial exome revealed novel hemizygous mutation c.1391T>G (p.Leu464Trp) in AIMF1 gene on chromosome X. Conclusions: X-linked AIMF1 encephalomyopathy due to apoptosis-inducing factor defect should be considered in the differential diagnostics in patients with Leigh syndrome, axonal neuropathy and myoclonic epilepsy. Supported by IGA NT 14156/3 and IGA NT 13114/4.

2344F

Reduced levels of ATP7A protein in a male child with subtle features of Menkes disease in the absence of *ATP7A* mutation suggest a novel defect in transcriptional regulation associated with a unique disorder of copper metabolism. *K. Patel[1], G. Golas[2], D. Adams[2], C. Tifft[2], C.S. Holmes[3], D.S. Goldstein[3], L. Yi[1], W.A Gahl[2], S.G. Kaler[1].* 1) Molecular Medicine Program; NICHD/NIH, Bethesda, MD; 2) Undiagnosed Disease Program; NHGRI, NIH, Bethesda, MD; 3) Clinical Neuroscience Program; NINDS, NIH, Bethesda, MD.

We evaluated a 6 year old white male from the NIH Undiagnosed Disease Program for a possible Menkes disease variant. His term birth was normal with Apgar scores recorded as 8 at both 1 and 5 minutes. Early neurodevelopment was markedly delayed. Since infancy, he has remained a diagnostic conundrum despite extensive testing. Based on normal plasma neurochemicals at age two years, Menkes disease was excluded. Physical examination findings at age 6 years included dysmorphic facies, coarse hair texture, alternating esotropia, generalized hypotonia, mild joint laxity, and moderate intellectual disability. Hair analysis showed *pili torti* in 4/25 (16%) of hair shafts obtained at age 9 months, and 7/35 (20%) hair shafts at age 6 years. A positive control hair specimen, obtained from a 3 year old classic Menkes patient receiving copper injection treatments, showed 6/25 (24%) *pili torti*. A brain MRI/MRA at age 6 years showed vascular dilation and tortuosity of the major arteries within the Circle of Willis. The patient had low-normal levels of serum copper (64 µg/dl) and ceruloplasmin (172 mg/L). Plasma catechol ratios were normal when tested at age 2 years and again at age 6 years, indicating normal peripheral activity of dopamine-beta-hydroxylase (DBH), a copper-dependent enzyme. However the catechol ratios in his cerebrospinal fluid at age 6 years were just within the range from a cohort of classic Menkes disease patients. DNA sequence analysis failed to reveal any abnormality in the *ATP7A* coding regions, splice junctions, and known regulatory elements. However, Western blot assay of the patient's fibroblasts revealed approximately 90% reduction in ATP7A protein quantity compared to wild-type cells. The residual ATP7A protein was normal-sized and exhibited proper intracellular trafficking itineraries in response to varying concentrations of copper. The patient's phenotype does not fit well with any of three recently identified autosomal recessive copper metabolism phenotypes caused by mutations in: 1) a copper chaperone (CCS), 2) an acetyl CoA transporter (AT-1), or 3) the sigma 1 alpha subunit of adaptor protein complex 1 (AP1S1). We therefore hypothesize a defect in an unrecognized transcriptional regulator of *ATP7A* and predict that delineation of the precise molecular basis of this syndrome (underway via whole exome sequencing) will reveal yet another novel cause of abnormal human copper metabolism, and one which may be amenable to therapeutic intervention.

2345T

Mutational analysis of ATP7B gene in 19 Vietnamese patients with Wilson Disease. *H. Cam Tu, Tr. Thinh Huy, T. Tue Nguyen, Ph. Le Anh Tuan, T. Van Thanh, Tr. Van Khanh.* Center For Gene Protein research, Hanoi Medical University, Hanoi, Hanoi, Viet Nam.

Wilson disease (WD) is an autosomal recessive inherited disease caused by abnormalities of the copper-transporting protein encoding gene ATP7B with over 600 mutations described. Identification of mutations has made genetic diagnosis of WD feasible in many countries. In this study, we examined ATP7B for mutations in 19 unrelated Vietnamese WD patients who were diagnosed as WD base on typical clinical symptoms. Their clinical characters are Kayer-Flesicher rings presence, Neurological symptom, Serum ceruloplasmin (around 10mg/dl), Hepatic dysfunction. All 21 exons and exon-intron boundaries of the ATP7B gene were amplified by polymerase chain reaction from the genomic DNA of the patients and then analyzed by direct sequencing. Totally, we identified 15 mutations, of which 4 mutations were novel: p.S105Stop (c.417T>A), g.81007T>A, c.47/48-CGGCG, p.D1027H (c.3079G>C) and 11 mutations were reported previously: p.K832R (c.2650A>G), p.V456L (c.1523G>C), p.V1140A (c.3419T>C), c.-118/117+GCCGC, p.T850I (c.2549C>T), p.L1371P (c.4112T>C), p.A604P(c.1967G>C), p.Y715D (c.2143T>G), p.Y765G (c.2295C>G), p.G943D (c.2827G>A), p.C985T (c.2945G>A). Among these, p.V456L (c.1523G>C) (25%), c.47/48+CGGCG (19 %) and 118/117+GCCGC (8%) were the most frequent mutations. Five patients were homozygous, fourteen patients were compound heterozygote. Our study will broaden our knowledge about ATP7B mutations in WD patients in Vietnam, and be helpful for clinical genetic testing. Keywords: ATP7B gene, mutational analysis, Wilson disease.

2346F

Intellectual disability and movement disorder caused by a homozygous mutation in N-glycanase (NGLY1) gene: A novel disorder of glycoprotein metabolism. *M.A. Jones, C. da Silva, A. Bibb, S. Warren, M. He, M.R. Hegde, M.J. Gambello.* Dept of Human Genetics, Emory University, Atlanta, GA.

NGLY1 encodes N-glycanase, a deglycosylating enzyme, which is proposed to be a component of the ER-associated degradation (ERAD) pathway. This pathway is responsible for the degradation of misfolded glycoproteins. Defects in NGLY1 are predicted to result in the accumulation of misfolded glycoproteins due to impaired degradation. We performed clinical whole exome sequencing on a 2 year-old girl and identified a homozygous nonsense mutation c.1201A>T (p.R401X) in exon eight of the NGLY1 gene. Both parents are carriers of this mutation. An older sister with a similar phenotype also has the homozygous nonsense mutation. Two unaffected siblings do not carry this mutation. The clinical features of the 2 yr old girl and her sister include severe intellectual disability, abnormal movements, absent speech, hypotonia and mild dysmorphic features. She had an extensive negative evaluation including: a chromosomal microarray, plasma amino acid, urine organic acid and acyl carnitine analysis, transferrin and N-glycan analysis, brain MRI, muscle biopsy, mitochondrial genome sequencing, lysosomal enzyme panel, and MECP2 analysis. The p.R401X mutation has been recently reported in another individual in trans with a c.1891delC mutation. Similar phenotypic features shared among all three individuals include developmental delay and involuntary movements. The movement disorder in our patient and her sister were notable for occasional hand-wringing reminiscent of classic Rett syndrome. We are currently screening MECP2 negative patients as well as children with similar phenotypes for mutations in NGLY1 to determine the prevalence of this novel neurogenetic disease. We propose that NGLY1 should be considered in the differential diagnosis for patients with severe intellectual disability and abnormal movements. Identification of additional patients with this rare disorder will allow further characterization of the phenotypic spectrum and natural progression of the disease.

2347T

Comprehensive molecular investigation of the pathologial effects of Menkes disease missense mutations. *L.B. Møller, T. Skjørringe, S.S. Thorborg.* Kennedy Center, Rigshospitalet, Glostrup, Denmark.

As an attempt to predict the functional consequence of substituted nucleotides in the ATP7A gene we have investigated the effect of 36 selected missense mutations in the ATP7A gene according to the resulting ATP7A transcript and protein including the intracellular localization. ATP7A encodes a copper-translocating ATPase belonging to the family of P-type ATPases. Mutations of ATP7A lead to Menkes disease, which is an X-linked, lethal disorder of copper metabolism. The ATP7A protein has a dual function: In basal copper levels ATP7A is located in the trans-Golgi network where it is responsible for delivery of copper across the membrane for cuproenzyme biogenesis. In excess intracellular copper levels, the steady state distribution of ATP7A shifts to the plasma membrane where the protein is responsible for pumping copper from the cell. Fibroblasts obtained from Menkes patients with missense mutations in the ATP7A gene was used as source for ATP7A mRNA and protein investigation. Investigation of ATP7A transcript was performed by RT-PCR and real-time PCR. ATP7A Protein quantitation and localization revealed was performed by Western blotting and immunofluoresecence. Notably, the pathogenic effect of almost one third of the missense mutations is not due to functional importance of the affected residue, as the mutations affect splice sites, splicing silencers, mRNA stability or protein stability, leading to absence or reduction of the resulting ATP7A protein. More than two third of the mutations abrogates copper dependent trafficking, probably by affecting phosphorylation, dephosphorylation, copper binding or copper transport. Only two out of 36 missense mutations show normal copper dependent trafficking. These results emphasized the importance of using a comprehensive spectrum of analyses to reveal the deleterious effect of missense mutations.

2348F

The effect of homozygous deletion of BBOX1 gene on carnitine level and fatty acid beta-oxidation. *A. Rashidi-Nezhad[1,2], S. Talebi[3], S.M. Akrami[3], A. Reymond[2].* 1) Maternal, Fetal & Neonatal Research Center, Tehran University of Medical sciences, Tehran, Iran; 2) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 3) Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.

Carnitine is a key molecule in energy metabolism and functions in the transport of activated fatty acids into the mitochondria. Carnitine hemostasis achieved and remained through oral intake, renal reabsorption and de novo biosynthesis. Unlike the dietary intake and renal reabsorption, the importance of de novo biosynthesis pathway in carnitine hemostasis remains unclear, due to lack of animal models or patients with the defect of this pathway. Here we evaluate metabolic and molecular study on carnitine level in a patient deficient in carnitine de novo biosynthesis. The proband was a 42 months old girl with microcephaly, speech delay, growth retardation and minor facial anomalies. ArrayCGH was used to detect small genomic imbalances. Quantitative PCR (qPCR) was used to confirm array results. Carnitine profile and metabolic investigation were performed according to routine laboratory protocols. The proband showed 221 Kb interstitial homozygous deletions at 11p14.2. Since BBOX1 gene is located in this region, the proband has no de novo biosynthesis of carnitine. Although free carnitine level in the proband was in normal range, acylcarnitine to free carnitine ratio (AC/FC ratio) was high but near the upper limit of reference range. Almost all the other evaluated metabolites were normal. In conclusion,we present the first report of a patient with the complete defect of carnitine de novo biosynthesis. Although this condition can result in a slight increase in AC/FC ratio but, it may not give rise to clinical manifestations of carnitine deficiency disorders.

2349T

Glycerol Kinase Transgenic Mice Have Increased Risk For Obesity. *C. Ho[1], A. Badjatiya[1], K.M. Dipple[1,2].* 1) Dept of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Dept of Pediatrics, Division of Genetics, University of California. Los Angeles, CA.

Glycerol kinase deficiency (GKD) is an X-linked inborn error of metabolism with a lack of genotype-phenotype correlation. Glycerol kinase (GK in humans, Gyk in mice) not only phosphorylates glycerol, but also performs other protein functions that may explain the pathogenesis of GKD. We recently developed a liver-specific *Gyk* transgenic mouse strains to understand the alternative functions of GK, including its role in adipogenesis and obesity. Male wildtype (WT/WT), heterozygous (WT/T) and homozygous (T/T) transgenic mice were placed on either regular chow or high fat (HF) diet for 12 weeks and monitored for weight gain and percentage body fat. Fasting glucose and cholesterol levels were measured from blood plasma to assess their risk for obesity and type II diabetes mellitus (T2DM). WT/T and T/T mice on HF diet gained more weight, 2.8% and 11.3% respectively, than WT/WT mice ($p<0.05$). Nuclear magnetic resonance analysis revealed both wildtype and transgenic mice had at least two-fold increase in percentage body fat as a result of consuming the HF diet. The elevated body weight in transgenic mice was also validated based on weighing of various tissues, including the mass of liver, visceral fat, perirenal fat and subcutaneous fat pads. T/T mice on chow diet had significantly greater mass in the liver tissue (25.8%), visceral fat pads (53.3%), and perirenal fat pads (40.2%) compared to the WT/WT mice ($p<0.05$), suggesting that *Gyk* overexpression is associated with abdominal obesity. Blood glucose and cholesterol tests showed that T/T mice on HF diet had higher triglycerides levels (16.4%), total cholesterol levels (41.3%), and fasting glucose levels (45.4%) relative to WT/WT, also implying high risk for obesity and T2DM ($p<0.05$). This study demonstrates that GK is involved in fat deposition and adipogenesis, and its overexpression increases risk of obesity and T2DM in mice. In addition, we propose that our *Gyk* transgenic mouse strains may serve as a useful model for obesity and T2DM research.

2350F

Novel heterozygous and homozygous mutations in the gap junction protein gamma-2 (*GJC2*) gene. *G. Hobson, D. Stubbolo, N. Manolakos, V. Funanage.* Nemours Biomedical Res, A I duPont Hosp Children, Wilmington, DE.

Recessive mutations in the gap junction protein gamma-2 (*GJC2*) gene cause Pelizaeus-Merzbacher like disease (PMLD), a rare hypomyelinating leukodystrophy. The clinical phenotype of PMLD is similar to that of the X-linked Pelizaeus-Merzbacher disease (PMD) and includes features such as spasticity, ataxia, dystonia, nystagmus, and developmental delay. A distinguishing feature of PMLD is the presence of brainstem auditory evoked potentials, which are typically absent in PMD. There have been reports that dominant point mutations in *GJC2* are associated with hereditary lymphedema type IC. Associated features include mild to severe swelling of the limbs and recurrent cellulitis. For the first time in our Molecular Diagnostics Lab, we report a cohort of patients and family members who have mutations in *GJC2* . For each patient, the entire coding region, surrounding intron, and promotor region of *GJC2* were sequenced. Family members, when available, were screened for the mutation(s) seen in the proband. Any changes from the reference sequence were analyzed via SIFT, PolyPhen, and Mutation Taster. These mutations include novel and previously reported single heterozygous, homozygous, and compound heterozygous mutations. The mutations are scattered throughout the entire coding region, as well as two different mutations in the promoter region. Three patients from different families had the c.-167A>G promotor mutation, which supports the previously reported hypothesis that this is a common pathogenic mutation in *GJC2*. The mutations p.Pro90Ser, p.Arg104Vfs*106, p.His252Arg, and p.Thr398Ile were each seen in two different families. Both the p.Pro90Ser and p.Thr398Ile have been reported previously, which suggests that these two mutations are also relatively common. For those patients with only one heterozygous mutation in the *GJC2* gene having PMLD symptoms, we suggest the possibility of an additional mutation being present in a region not sequenced, variable expressivity, or autosomal dominant inheritance. Further parental and familial genetic testing and phenotypic review should be completed to help make this distinction.

2351T

Arylsulfatase A Deficiency in Thai MLD Patients. *S. Nujarean[1], L. Thampratankul[2], L. Choubtum[3], D. Wattanasirichaigoon[1].* 1) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, 10400 Thailand; 2) Division of Neurology, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, 10400 Thailand; 3) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, 10400 Thailand.

Metachromatic Leukodystrophy (MLD) disorder is the lack of Arulsulfatase A (ARSA) enzyme which missing of ARSA enzyme lead to damage nervous system, kidney and other organs. MLD is an autosomal recessive disease caused by mutations in ARSA gene which ARSA gene is located on chromosome 22q13.33.It contains 8 coding exons and encodes 539 amino acid (NM_000487.5).This study; we identify mutations in ARSA gene in Thai MLD patients. The following in 8 exons of ARSA gene.

2352F

Unexpected findings of hexosaminidase B deficiency in a 39 yr old asymptomatic female identified during routine carrier screening for Tay-Sachs disease. *A. Schneider[1], R.B. Keep[2], M. Sobel[4], J. Liao[3], C. Yu[3], R. Kornreich[3].* 1) Dept Gen, Albert Einstein Med Ctr, Philadelphia, PA; 2) Quest Diagnostics, Horsham, PA; 3) Mount Sinai School of Medicine, New York, NY; 4) Abington Reproductive Medicine, Abington, PA.

Sandhoff disease is a rare autosomal recessive lysosomal storage disorder caused by mutations in the HEXB gene which results in the deficiencies of both Hexosaminidase A and B. Tay-Sachs disease (TSD), a related GM2 gangliosidosis, results from mutations in the HEXA gene causing the deficiency of Hexosaminidase A only. While similar disorders, Sandhoff disease is much less common than TSD and has no known ethnic predilection. Sandhoff disease presents with a wide range of phenotypes, from infantile to adult-onset, all of which are progressive to varying degrees. Detection of TSD and Sandhoff carriers is routinely performed via a standard heat inactivation Hex A% and total hexosaminidase activity assay using an artificial 4-MUG substrate. This assay can also diagnose TSD and Sandhoff diseases. Individuals affected with Sandhoff disease have a very low total hexosaminidase activity (<100 nmol/hr/mg) with a disproportionately high percentage Hex A (high 80s to 100%). This patient is a 39-year-old asymptomatic female of Sephardic Jewish ancestry. Her parents are from Syria and Yemen and are non-consanguineous. She was referred for a clinical genetics evaluation by an infertility practice due to abnormal enzyme results suggestive of Sandhoff disease. Blood was received for confirmational purposes. The leukocyte percent Hex A activity was 89.3%, which is higher than our Sandhoff carrier range (76.4 - 85.2%, 80.2 % ± 2.8, n=16). The total hexosaminidase activity was 456 nmol/hr/mg, which is reduced to 27% of the normal mean (1659 nmol/hr/mg, n=3939) or 57% of the Sandhoff carrier mean (798 nmol/hr/mg, n=16). Sequencing of the HEXB gene revealed four variants: c.214C>T, c.619A>G, c.1627G>A, and c.*81_*82delTG (homozygous). The variant c.1627G>A is a known heat labile allele in Oriental Jews. Both c.619A>G and c.*81_*82delTG are common polymorphisms which have been reported to reduce Hex B activity. In addition, a variant of unknown significance, c.214C>T causing a Leu->Phe substitution at position 72, is predicted to be a deleterious mutation by Polyphen and SIFT. Although, the effects of these variants in vivo are difficult to predict, her in vitro hexosaminidase B activity was indeed reduced; therefore, the late-onset clinical manifestation of GM2 accumulation cannot be excluded in this patient.

2353T

Comprehensive mutation screening of the AGXT gene in patients with primary hyperoxaluria type-1. *S. Khaliq[1,2], A. Abid[1], S. Shahid[1], S. Hashmi[3], S. Sultan[4], T. Aziz[5], S.A.H. Rizvi[5], S.Q. Mehdi[1].* 1) Centre for Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation , Karachi, Pakistan; 2) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan; 3) Department of Pediatric Nephrology, (SIUT), Karachi, Pakistan; 4) Department of Pediatric Urology, (SIUT), Karachi, Pakistan; 5) Department of Urology, (SIUT), Karachi, Pakistan.

AAbstract: Primary Hperoxaluria type 1 (PH1) is a rare autosomal recessive disorder caused by the deficiency of alanine glyoxylate aminotransferase (AGT) enzyme encoded by the AGXT gene. It is characterized by recurrent nephrolithiasis, nephrocalcinosis and end stage renal disease. The three most common mutations in the exons 1, 4 and 7 have been found in more than 50% of the cases. Hence, mutation screening of these three exons has suggested being the 1st-line test to avoid liver biopsy. Since the facilities for liver biopsy are not commonly available in this region, mutation screening is the only reliable procedure for the diagnosis of PH1. Mutation analysis was carried out by direct sequencing of the AGXT gene in 150 patients. The age of onset of PH1 was 1-16yrs. They rapidly progressed to renal failure at adolescence. The study was approved by the IRB and conformed to the Tenets of the Declaration of Helsinki. We have found 17 mutations in 44 patients in which three mutations were novel. They included one splice-site, four indel and twelve missense mutations. The most common p.G350D mutation was found in 12 patients (8%). The age at onset of PH1 in these patients was 5-16 yrs. Of these, three children developed renal failure. One patient had encountered transplant rejection due to the recurrence of oxalosis. Another common mutation p.G190R was found in 7 cases (4.6%). The mutation p.G82E was present in three patients on the background of the AGT-Ma allele. Interestingly, the most commonly reported mutation c.508G-A (p.G170R; 20-40% allele frequency) was not found in our patients. Another common c.33dupC (12% allele frequency in Caucasians) that leads to protein truncation was found in only five patients (3.3%). AGT-Mi allele was found in 12 patients (8%) that mostly occurred in combination with the p.G350D mutation which is not reported elsewhere. Linkage disequilibrium among the AGT-Mi was not found in this study. Mutations in the AGXT gene have been identified in more than 90% of the cases worldwide. However, they were found in only 30% of the cases in this cohort. It is noteworthy that the most commonly reported mutations are not prevailing in our patients, therefore, mutation screening of the whole AGXT gene needed to be carried out in these patients.

2354F

Mutation of the iron-sulfur cluster assembly IBA57 gene causes lethal myopathy and encephalopathy. *N. Ajit Bolar[1,8], A. V. Vanlander[2,8], C. Wilbrecht[3,8], N. Van der Aa[1], J. Smet[2], B. De Paepe[2], G. Vandeweyer[1], F. Kooy[1], F. Eyskens[4], E. De Latter[2], G. Delanghe[5], P. Govaert[5], J.G. Leroy[2], R. Lill[3,6,7], R. Van Coster[2], L. Van Laer[1], B. Loeys[1].* 1) Department of Medical Genetics, University of Antwerp, Antwerp, Edegem, Antwerp, Belgium, Belgium; 2) Department of Pediatrics, Division of Pediatric Neurology and Metabolism, University Hospital Ghent, 9000 Ghent, Belgium; 3) Institut für Zytobiologie, Philipps-Universität Marburg, Robert-Koch Str. 6, 35032 Marburg, Germany; 4) Provinciaal Centrum voor de Opsporing van Metabole Aandoeningen (PCMA), Department of Pediatrics/Metabolic Diseases, Faculty of Medicine and Health Sciences, Antwerp University Hospital and University of Antwerp, 2000 Antwerp, Belgium; 5) Department of Neonatology, Paola Children's Hospital ZNA Middelheim, 2000 Antwerp, Belgium; 6) Max-Planck-Institut für terrestrische Mikrobiologie, Karl-von-Frisch-Str. 10, 35043 Marburg, Germany; 7) LOEWE Zentrum für Synthetische Mikrobiologie SynMikro, Hans-Meerwein-Str., 35043 Marburg, Germany; 8) The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors.

The iron sulphur [Fe-S] proteins play an important role in redox reactions of the mitochondrial electron transport chain. The de novo synthesis and maturation of these proteins is highly complex and involves more than 25 biogenesis factors. In this study, we have identified two siblings from consanguineous parents who died perinatally from a condition characterised by generalised hypotonia, respiratory insufficiency, anthrogryposis, microcephaly, congenital brain malformations and hyperglycinemia. Analysis of the catalytic activities of the mitochondrial respiratory complexes I and II indicated deficiency in skeletal muscle, suggestive of an inborn error in the mitochondrial iron-sulfur cluster (ISC) biosynthesis pathway. Homozygosity mapping revealed the IBA57 gene, which is known to be involved in the biosynthesis of mitochondrial [4Fe-4S] proteins and present in the largest homozygous region on chromosome 1, as a candidate gene. Mutation analysis of IBA57 identified a c.941 A>C transversion causing the amino acid change p.Gln314Pro. Biochemical analysis of skeletal muscle and skin fibroblasts of affected individuals indicated severely decreased amounts of IBA57 and a decrease in various 4Fe-4S proteins and in proteins covalently linked to lipoic acid. IBA57 depleted HeLa cells reflected biochemical defects consistent with observations in patient derived cells. Defects could be rescued by the introduction of wildtype IBA57 and partially by mutant IBA57. Further functional analysis revealed an increased sensitivity of mutant IBA57 to degradation via proteolysis. Our findings suggest that the mutation leads to functional impairment and degradation below physiologically critical levels, resulting in the condition observed in the patients. In conclusion, we have identified a novel metabolic disorder presenting with a lethal complex biochemical phenotype caused by defective assembly of the ISC protein, IBA57.

2355T

Genetic variation in a gene involved in glycosphingolipid biosynthesis. *O.M. Amaral, A.J. Duarte, E. Pinto, I. Ribeiro, L. Lopes, D. Ribeiro.* Department of Human Genetics, INSA, IP , 4000-Porto, Portugal.

The main objective of this work was to investigate the possible existence of genetic variation in the UGCG gene. The UGCG gene encodes an enzyme essential in the first step of the glycosphingolipid biosynthesis process. Its genetic variation could lead to differences in biosynthesis and be related to phenotypic divergence in various genetic diseases of the glycosphingolipidoses group. In order to test this hypothesis we attempted to identify the extent of variation in the UGCG gene in order to relate it to phenotypic variation. Methods and samples: DNA was extracted from blood samples and/or fibroblast cell lines using an automated apparatus. Biological samples were obtained from healthy donors, with informed consent. In addition, all traceable identification was removed, so as to guarantee their anonymous nature. Skin fibroblast cell lines were obtained from the Corriel Institute (USA). The UGCG gene (exons and flanking intronic regions) of six control individuals was sequenced using standard methods. Results: In this work we present the identification and distribution of genetic variations among the control samples studied. The results obtained with the different samples showed the existence of several polymorphic changes. Discussion: Polymorphisms in the UGCG gene may interfere with the amount of substrate available for degradation in specific diseases along the same pathway. Thus, the degree of genetic variability might influence the phenotypic expression as well as the lysosomal burden. Conclusion: Assessment of variation in the UGCG gene should be considered, particularly in patients who do not comply with the expected genotype/phenotype correlations. Additional information: This work was carried out with financial support obtained from FCT-Portugal: project PIC/IC/82822/2007(2009); AJD and DR were beneficiaries of BI grants from Fundação da Ciência e Tecnologia (FCT/MCTES) - Portugal. Corresponding author: Olga Amaral, olga.amaral@insa.min-saude.pt.

2356F

A zebrafish model of cblC disease displays growth retardation that improves with vitamin B12 therapy. *N.P. Achilly[1], J.L. Sloan[1], K.S. Bishop[1], M.S. Jones[1], V.J. Hoffman[2], R.B. Sood[1], C.P. Venditti[1].* 1) Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Office of the Director, National Institutes of Health, Bethesda, MD.

Cobalamin C disease (*cblC*) is the most common inborn error of intracellular cobalamin metabolism. It is caused by mutations in *MMACHC*, a gene responsible for processing and trafficking intracellular cobalamin. Defects in this pathway impair the function of two cobalamin-dependent enzymes: methylmalonyl-CoA mutase and methionine synthase. Disease manifestations can include growth failure, anemia, congenital microcephaly, heart defects, and progressive blindness. At present, the pathological basis of these symptoms remains unknown, and no animal model exists. To replicate clinical manifestations experienced by patients with *cblC* disease, we created a series of loss of function alleles in the zebrafish orthologue of *MMACHC* using zinc-finger nucleases. Of these, we chose p.L44PfsX21 (hg12) and p.G32VfsX48 (hg13), transmitted by two independent founders, for phenotype analysis. F2 $mmachc^{hg12/hg12}$ and $mmachc^{hg13/hg13}$ fish survived the embryonic period but displayed growth impairment after 14 days post-fertilization (dpf). By 21 dpf, the standard length (SL) and height at the anterior of the anal fin (HAA) were significantly reduced; $mmachc^{hg12/hg12}$ fish (SL 6.94 ± 0.07, HAA 0.77 ± 0.03 mm) and $mmachc^{hg13/hg13}$ (SL 7.40 ± 0.07, HAA 0.86 ± 0.01 mm) fish were smaller than the wild-type and heterozygous fish (SL 10.39 ± 0.18, HAA 1.48 ± 0.03 mm) ($p<0.0001$). Histological examination of $mmachc^{hg12/hg12}$ fish revealed a complete absence of the secondary lamellae in the gills, which contain specialized cells for gas and ion exchange. Thinner retinal layers and a possible defect in the morphology of the photoreceptor outer segments were also observed. The concentration of methylmalonic acid (MMA), a classic biomarker of *cblC* disease, was elevated by 289-fold in $mmachc^{hg12/hg12}$ fish. OH-cobalamin (OH-cbl) injections are the main treatment administered to the patients and ameliorate some of the disease-related complications. When $mmachc^{hg12/hg12}$ fish were maintained in water supplemented with OH-cbl (100 µg/ml) for 21 days, SL increased by 25% ($p<0.05$) and HAA increased by 30% ($p<0.01$) compared to the untreated group. The zebrafish model of *cblC* disease we generated recapitulates several of the phenotypic and biochemical features of *MMACHC* deficiency, demonstrates a response to conventional therapy, and should be useful to delineate the pathophysiological mechanisms in this common disorder of cobalamin metabolism.

2357T

Identification of a novel mutation in the human ARSB gene on chromosome 5q14.1 for Muchopolysacharidosis type VI patients in southwest Colombia. *M.A. Acosta Aragón[1], J.R Lago[2], F. Barros[2], A.M. Carracedo Alvarez[2-3].* 1) Pediatrics Department, College of Medicine, University of Cauca, Popayán, Cauca, Colombia, Ph.D; 2) Clinical Hospital of Santiago de Compostela, Galicia, Spain M.D; 3) Institute of Legal Medicine, College of Medicine, University of Santiago de Compostela, Galicia, Ph.D.

Introduction: The MPS VI or Maroteaux-Lamy Syndrome is a recessive multisystemic progressive lysosomic storage disease caused by a deficiency of N-acetylgalactosamine 4-sulfatase enzyme or Arylsulfatase B. A total of 32 patients MPS VI patients are identified in Colombia, sixteen in Cauca Department (southwest Colombia) corresponding to 50% of the total cases with Maroteaux-Lamy syndrome registered in the country. All sixteen patients were identified clinically and by enzyme assay. Two of these individuals with severe form of disease belong to an Amerindian reservation(Guambiano ethnicity). Objetive: to analyze the genomic variations in the Arysulfatase B gene in two patients with severe phenotype of disease and identify ethnic and family backgrounds of the MPS VI Patients in Cauca Department. Subjects: We studied two native patients and their relatives.Data was obtained from charts and families of patients including ethnic and family backgrounds. Methods:It was PCR/sequencing of the 8 exons and their flanking regions in Arysulfatase B gene.We accomplished the validation of exonic changes by computational methods (Alamut 2,HGVS 2). We used 20 single nucleotide polymorphisms (SNPs) for haplotype characterization with a Sequenom Mass Array analyzer. Results: We characterized both alleles in the patients and their relatives identifying a novel mutation p.Ser403X no reported before, both in homozygous or heterozygous form.In addition we identified the same haplotype in the two homozygous patients and their heterozygous relatives when analized this gene with intragenic SNPs.These results together with the genealogy analysis, strongly suggest an inbreeding effect in this population. Conclusions: A novel mutation in the human ARSB gene was reported. It produces a premature stop codon. These results emphasize the broad molecular heterogeneity of Maroteaux-Lamy syndrome and contribute to the establishment of a genotype/phenotype correlation in this disease. The high frequency of MPS VI patients in the Cauca Department and the ethnic characteristics are suggestive that a population genetic factor can be responsible,such as in this study an isolated population with strong inbreeding and consanguinity.Studies with another families of this region with molecular characterization of the mutation by sequencing and the phylogeographic identification of the origin and dispersion of the gene will be performed to clarify this ethnic prevalence.

2358F

Two novel mutations in acid α-glucosidase gene in two patients with Pompe disease. *A. Aykut[1], H. Onay[1], M. Kose[2], E. Erbas Canda[2], E. Karaca[1], M. Coker[2], F. Ozkinay[1].* 1) Department of Medical Genetics, Ege University Faculty of Medicine, IZMIR, bornova izmir, Turkey; 2) Department of Pediatric Metabolic Disease, University Medical Faculty, Izmir, Turkey.

Glycogen storage disease type II (GSD-II), also known as Pompe disease (OMIM #232300), is a rare progressive autosomal recessive disorder characterized by deficiency of acid α-glucosidase (GAA) that results in abnormal glycogen deposition in the muscles. Mutations in the GAA gene cause Pompe disease. The infantile-onset form is the most severe form. Most of the infantile-onset form patients present with hypotonia and cardiomyopathy in early infancy. The first case presented is a second child of consanguineous parents, born at term after an uneventful pregnancy. At the age of 9 months infantile type Pompe disease was suspected based on hypertrophic cardiomyopathy. Unfortunately her previous sister died from hypertrohic cardiomyopathy suggesting Pompe disease at 7 months old. The diagnosis was confirmed by measuring acid alpha-glucosidase activity in serum (0.4 µkat/kg protein). Although enzyme replacement therapy was started at 9 months, patient died from sudden cardiac arrest at 1 year old. Genetic analysis of the parents revealed a novel missense mutation p.Q682R in GAA gene which was not previously reported. Second case, is a child of nonconsanguineous parents, born at term after an uneventful pregnancy. Infantile type Pompe disease was suspected based on hypotonia/generalized muscle weakness at 6 months. Sensorineural hearing loss and optic atrophy were detected. Echocardiography was normal. Levels of creatine kinase, or creatine kinase-myocardial band isoenzyme were typically elevated. She had GAA enzyme activity that was less than 1 percent of normal. She died from respiratory weakness at 2 years old. Genetic analysis of the mother revealed a novel nonsense mutation p.Q255X in GAA gene which was not previously reported and father had p.R224W which was previously described. We report two families wherein molecular diagnosis for Pompe disease was performed and we have identified two novel mutations in the GAA gene which were not previously described.

2359T

New dominant *ALDH18A1* mutation in two unrelated children with neurodevelopmental delay, cataracts, cutis laxa, and intracranial arterial tortuosity. *J. Ganesh[1], A.E. Lin[2], I. Sahai[2], E.H. Zackai[1], S. Chadwick[1], T.L. Toler[2], P.H. Byers[3,4], U. Schwarze[3].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) Mass General Hospital for Children, Boston, MA; 3) Department of Pathology, University of Washington, Seattle, WA; 4) Department of Medicine, University of Washington, Seattle, WA.

Cutis laxa syndromes are genetically heterogeneous. To date, mutations in *ALDH18A1*, encoding Δ^1-pyrroline-5-carboxylate synthase (P5CS), have been reported in five families. In four families biallelic mutations were present, and in the fifth there was a causative heterozygous mutation. We identified an apparently dominant *ALDH18A1* mutation (c.413G>A, p.Arg138Gln) in two unrelated infants with a complex phenotype that overlapped with the findings in reported patients and included failure to thrive, neurodevelopmental delay, bilateral cataract, intracranial arterial tortuosity, skin laxity, and a metabolic profile of low serum ornithine, citrulline, arginine, proline, and mildly elevated ammonia. P5CS is a bifunctional mitochondrial enzyme with both gamma-glutamyl kinase (γ-GK) and gamma-glutamyl phosphate reductase activity. It catalyzes a critical step in the biosynthesis of proline and ornithine, a substrate of the urea cycle and precursor in arginine biosynthesis. The heterozygous mutation that we identified in both infants results in substitution of an evolutionarily conserved arginine by glutamine at position 138 within the γ-GK domain. In the first child (P1) the mutation occurred *de novo* on the paternally derived allele. In addition there was a maternally derived polymorphic variant (c.896C>T, p.Thr299Ile) that has a heterozygote frequency of about 20% among European-Americans. In the second child (P2) the mutation arose *de novo* and there was no background polymorphic minor allele. We identified no other alteration in the gene sequence or relative mRNA abundance of the products of the two alleles in fibroblasts. These findings are consistent with the concept that the substitution of arginine by glutamine at this site represents a new dominant mutation and is sufficient to cause the phenotype. P1 was treated with L-arginine, leading to resolution of irritability, feeding difficulties, and tremors within a few weeks, and sustained improvement in growth and development after sixteen months. In P2 treatment with citrulline, ornithine, proline, and ascorbic acid was initiated with some improvement. Although mutations in *ALDH18A1* appear to be a rare, both dominant and biallelic mutations should be considered in young children with skin laxity, neurodevelopmental delay, cataracts, and arterial tortuosity and in whom sequence analysis of the growing number of other "cutis laxa genes" has not led to a diagnosis.

2360F

Molecular Characterisation of known mutations in Congenital Adrenal Hyperplasia patients (CYP21A2 gene) : Genetic & Diagnostic implications. *R. KHAJURIA[1], A. KUMAR[1], D. PAL[1], U. SHARMA[1], A. BHANSALI[2], R. WALIA[2], R. PRASAD[1].* 1) biochemistry, PGIMER, CHANDIGARH, India; 2) ENDOCRINOLOGY,PGIMER,CHANDIGARH,INDIA.

Abbreviations : SW-Salt Wasting, SV-Simple virilising, NC- Non classical, ACRS-Amplification created restriction site. Background : Congenital Adrenal Hyperplasia (CAH) is frequent autosomal recessive disease, with a wide range of clinical manifestations from severe classical form to late onset form. Most cases of CAH, the inherited inability to synthesize cortisol, aldosterone with subsequent overproduction of androgens, are caused by mutations in steroid 21 Hydroxylase gene. As a complement to hormonal measurements, mutation analysis of CYP21A2 gene is potentially important tool in diagnosis of steroid 21 Hydroxylase defiency and genetic counselling. Objective & methods : Our aim was to determine the frequency of common CYP21A2 gene mutations. Clinical and hormonal evaluations were used to categorize the patients in Salt Wasting (SW), Simple Virilizing (SV) and Non Classical (NC) forms. About 95 % of mutant alleles have apparently been transferred from linked pseudogene (CYP21P) to active gene (CYP21A2).-Molecular analysis of CYP21A2 was performed for detection of common mutations viz large gene deletions by PCR, ACRS, restriction method and product sequencing in 38 patients. Results : Disease causing mutations were identified in total 38 patients comprising SW(n=12), SV(n=18) and NC(n=8). Single gene deletion was found with frequency of 18.4 % in SW, SV patients whereas homozygosity was found with 7.9% in these cases. Heterozygous R356W mutation was found in SW (n=4), SV(n=3) and NC (n= 1). Heterozygous i2g mutation was found in SW (n=1) case. Heterozygous Q318X mutation was also present in SV(n=2) patients. Compound heterozygous patients were present in the cohort. Conclusions : This is a comprehensive study showing deleterious mutations in functional,CYP21A2 gene in our CAH population.

2361T

Spectrum of hyperhomocysteinemia in the pediatric and adolescent age group with MTHFR genotype in a north Indian cohort. *A. Lomash, S. Kumar, S. Kumar Pandey, A. Singh, S. Kumar Polipalli, S. Kapoor.* Division of Genetics, Department of Pediatrics, Maulana Azad Medical College, New-Delhi 110002, India.

Objective:Spectrum of hyperhomocysteinemia in the pediatric and adolescent age group is rarely reported from Indian ethnicity.We evaluated the clinical spectrum & association of MTHFR C677T & A1298Cgene polymorphisms with plasma homocysteine conc.in North Indian Cohort Materials & Methods:We analysed a cohort of 40patients referred to us for estimation of Plasma Homocysteine.Fasting plasma total homocysteine conc.were measured using RP-HPLC & genotype frequencies of the MTHFR gene at positions C677T & A1298C were ascertained using PCR-RFLP method.The age groups at presentation were divided into 0-12years(pediatric),12-18years(adolescent)&>18years(adult).Statistical calculation for genotype frequencies was carried out with Kruskal Wallis test using STATA software & p value less than <0.05 were considered significant Results:20subjects demonstrated hyperhomocysteinemia in the pediatric age.The mean age at presentation was 6.22years(6.22±SD2.83)The commonest presentation was stroke(60%),diminution of vision(20%),developmental delay(20%),neuroimaging finding was an infarct(30%)& lens dislocation(20%).Six cases presented in the adolescent age group with mean age of 13.6years(13.6±SD 1.57).The commonest reason for referral was stroke75% and 25%with lens dislocation.Seizures were seen in 33%cases.In the adult age group,14 cases with mean age at presentation was 25.43years(25.43 ±SD.5.97),presented with Deep Venous Thrombosis(29%),Diminution of vision(29%),Stroke(14%)Recurrent abortions(14%),Osteoporosis(14%),young MI with family history(14%).The frequencies of the AA,AC,CC genotype in cases&controls were 15%,72.5%& 10% and 87.5% & 12.5%.The frequencies of the CC,CT,TT genotype in cases&controls were 50%,50% and 75% & 25% respectively Conclusion:Stroke was the commonest reason for referral across all age groups.The commonest reason for suspicion of abnormality in Homocysteine metabolism remained suggestion of classic disease due to lens subluxation.Adult onset suspicion was clinically heterogeneous.The odds ratios of the A1298C genotypes were significantly different(17.0RR)between Patients and normal controls.The frequencies of C677T haplotype were significantly higher(3RR)in the patients than in the control group.A very Low frequency of T allele was demonstrated.Hyperhomocysteinemia is a significant contributor to pediatric and adolescent stroke.Genotypes A1298C &C677T appear to confer significant susceptibility to the development of Hyperhomocysteinemia.

2362F

A homozygous *UQCRC2* mutation cause a neonatal onset metabolic decompensation due to complex III deficiency. *N. Miyake[1], S. Shoji[2], Y. Goto[3], N. Matsumoto[1].* 1) Yokohama City University, Yokohama, Kanagawa, Japan; 2) Genetics Division, Department of Pediatrics, LAC+USC Medical Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 3) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, NCNP, Kodaira, Tokyo, Japan.

The mitochondrial respiratory chain generates energy as adenosine triphosphate (ATP) by means of the electron-transport chain and the oxidative-phosphorylation system. The mitochondrial respiratory chain is composed of five complexes: I, II, III, IV, and V. Mitochondrial complex III (CIII) deficiency is a relatively rare disease presenting high clinical and genetically heterogeneous. Until now, mutations in four genes have been known to cause autosomal recessive CIII deficiencies: *UQCRB, UQCRQ, BCS1L and TTC19*. *UQCRB* and *UQCRQ* encode components of CIII itself, while *BCS1L* and *TTC19* produce mitochondrial assembly factors. Here, we report three patients from a consanguineous Mexican family presenting with neonatal onset of hypoglycemia, lactic acidosis, ketosis, and hyperammonemia. According to the evidence of consanguinity, we hypothesized that this disease inherited in autosomal recessive fashion. By linkage analysis using SNP mapping array, we narrowed down the linked region to 36 Mb. At the same time, we utilized whole exome sequencing for one affected individual and picked up the homozygous variants. Then, we successfully found a homozygous missense mutation in *UQCRC2* (NM_003366), which encodes mitochondrial ubiquinol-cytochrome c reductase core protein II, in all three affected individuals. In its native state, the CIII monomer is quickly converted into a catalytically active homodimer that is incorporated into a supercomplex, and this supercomplex functions as a single enzyme. Based on structural modeling, the mutation (p.Arg183Trp) was predicted to destabilize the hydrophobic core at the subunit interface of the core protein II homodimer. Furthermore, *in vitro* studies using fibroblasts from the patient clearly indicated CIII deficiency, as well as impaired assembly of the supercomplex consisting of complexes I, III, and IV. This is the first described human disease caused by *UQCRC2* abnormality.

2363T

Assaying interallelic complementation in vivo at the Mut locus with adeno-associated (AAV) viral gene delivery. *J.S. Sénac, C.P. Venditti.* National Human Genome Research Institute, National Institute of Health, Bethesda, MD.

Methylmalonic acidemia (MMA) is caused by mutations of the mitochondrial gene methylmalonyl-CoA mutase (MUT). Among the wide spectrum of partial activity or *mut*⁻ mutations, the mechanism of enzymatic impairment has been defined only for selected mutations, such as p.G717V, which has been documented to be an adenosylcobalamin K_m mutant. We generated transgenic mice that ubiquitously express the murine p.G717V homologue (p.G715V) as a stable transgene (*Mut*$^{-/-}$;Tg$^{INS\text{-}CBA\text{-}G715V}$) and have established that this new model mimics the physiologic and phenotypic manifestations observed in *mut*⁻ MMA patients, such as dietary induction of severe methylmalonic acidemia and the inducible formation of megamitochondria in the liver. Previous publications have demonstrated that cell lines harboring some MUT mutations, such as p.G717V mutation, could be 'rescued' by alleles with distinct mutations, such as p.R93H. Using a yeast expression approach, we first established that the co-expression of the murine homologues of these mutations, Mut p.R91H and p.G715V, produced a recombinant Mut enzyme with improved kinetic parameters compared to either allele, supporting previous cell culture observations. We next hypothesized that we could recapitulate interallelic complementation in vivo using an rAAV vector to deliver the p.R91H *mut* transgene to *Mut*$^{-/-}$;Tg$^{INS\text{-}CBA\text{-}G715V}$ mice. A single retro-orbital injection of ($5e^{10}$ GC) of AAV2/9-CBA-Mut-pR91H was performed in adult mice. Following delivery, weight and metabolites were monitored. As early as one week post injection, *Mut*$^{-/-}$;Tg$^{INS\text{-}CBA\text{-}G715V}$ mice demonstrated a significant weight increase (p=0.04) compared to treated heterozygote littermates, an effect similar in magnitude to that observed in MMA mice treated with a gene therapy vector configured to express wild type Mut. Our results suggest that the phenotypic improvement is secondary to increased Mut activity mediated by interallelic complementation of mutant alleles in vivo. We are conducting detailed physiologic and phenotypic characterization of the treated *Mut*$^{-/-}$;Tg$^{INS\text{-}CBA\text{-}G715V}$ mice to further define the mechanism(s) and consequences of interallelic rescue. Using a *mut*⁻ mouse model and rAAV mediated gene delivery, we establish a new platform to study the trans-effects of mutant MUT alleles and the foundation to examine small molecule activators of the MUT activity.

2364F

A molecular genetic study of Japanese families of Creatine Transporter Deficiency. *T. Wada[1], M. Tachikawa[2], S. Ohtsuki[3], S. Itoh[3], H. Shimbo[4], H. Osaka[4].* 1) Medical Ethics and Medical Genetics, Kyoto University School of Public Health, Kyoto, Kyoto, Japan; 2) Division of Membrane Transport and Drug Targeting, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan; 3) Department of Pharmaceutical Microbiology, Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan; 4) Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Japan.

Cerebral creatine deficiency syndrome (CCDS) is considered to be a potentially treatable intellectual disability syndrome. The common clinical feature of CCDS is intellectual disability, expressive speech and language delay, autistic behavior and epilepsy, and is composed of three genetic syndromes, Creatine transporter deficiency (CTD; MIM 300036), L-arginine:glycine amidinotransferase deficiency (MIM 602360), and guanidinoacetate methyltransferase deficiency (MIM 601240). CTD is an example of X-linked Intellectual Disability (ID), caused by mutations in SLC6A8 on Xq28. Although this is the second most frequent genetic cause of ID in Europe or America after Fragile X syndrome, there are few reports of Japanese cases and many patients should remain to be diagnosed in Japan. We have established a new simple high performance liquid chromatography screening method to determine the concentrations of guanidinoacetic acid, creatine, and creatinine in urine, which are diagnostic for differentiating these three types of CCDs. We have examined the urine samples of 106 patients (74 males and 32 females) with developmental disabilities at our medical center using this HPLC method as diagnostic screening test, and found three cases of CTD, one de novo case, and two familial cases, showing the characteristic pattern of the urine analysis, elevated creatine/creatinine ratio, and normal level of guanidinoacetic acid. And the other two familial cases whose 1H-MRS study indicated CTD as their diagnosis, and the urine analysis showed the specific pattern for CTD. All five cases have mutations in the SLC6A8 gene, including exon 5-13 deletion (de novo); c.1681 G>C, p.Gly561Arg in exon 12; c.321_323del CTT, p.Phe107del in exon 2; c.1661 C>T, p.Pro554Leu in exon 12; c.514 T>C, p.Cys172Arg in exon 3. All mutations are novel, and all mothers except one are carriers of the mutations. Our result indicates that our HPLC method is useful for diagnosis of CDDS, and a systematic diagnostic system of this syndrome should be established in Japan to enable us to estimate its frequency and develop the treatment.

2365T

Epidemiological studies of *POLG* mutations in Polish patients with mitochondrial disease of unknown etiology - preliminary data. *D. Piekutowska-Abramczuk, J. Trubicka, S. Łuczak, P. Kowalski, D. Jurkiewicz, E. Ciara, M. Borucka-Mankiewicz, M. Pelc, M. Krajewska-Walasek, E. Pronicka.* Medical Genetics, Children's Memorial Health Institute, Warsaw, Poland.

Studies on molecular background of mitochondrial diseases (MD) indicate *POLG* mutations as one of the most common defects (up to 25%) in this group. The spectrum of *POLG*-related disease inherited in autosomal recessive or dominant manner is broad and still not fully defined. Currently, it includes the main phenotypes of AHS, MCHC, MEMSA, MIRAS, SCAE, SANDO, PEO, and less frequently LS, MNGIE, MERRF, MELAS, and SMA-like. Worldwide epidemiological studies point the widespread occurrence of the common c.1399G>A (p.A467T), c.2542G>A (p.G848S), and c.2243G>C (p.W748S) mutations up to 2%, in different European populations but also stress significant differences in the frequency of carriers depending on its' ethnicity. Retrospective group of 150 Polish patients with clinical suspicion of mitochondrial disease and negative results of standard molecular analysis (including identification of common point mtDNA mutations: m.3243A>G in *MTTL1*, m.8344A>G in *MTTK*, m.8993T>G/C in *MTATP6*; large-scale mtDNA rearrangements; and nuclear encoded *SURF1* and *SCO2* mutations (in selected cases) were enrolled to the study. DNA samples were isolated from the blood, saliva, urine, muscle, and liver biopsy specimens with phenol-chloroform extraction. Genotyping of the samples for the presence of recurrent *POLG* mutations was performed by Real-time PCR technique with specific TaqMan allele discrimination assays on Light Cycler 480 II instrument (Roche). Identified mutations were sequenced using ABI PRISM dye terminator cycle sequencing kits and automatic sequencing system Applied Biosystems 3130. Only W748S variant was found in the studied MD group; in one case in homozygous and in six cases in heterozygous form, all in patients with Alpers-Huttenlocher syndrome. Significant predominance of p.W748S mutation indicates similarity of the genotype identified in Polish and Nordic patients, that is not common in another part of the world. Evaluation of the carrier frequency of the common *POLG* mutations in a large cohort of MD is under way. There is a possibility of the presence of new data. The study was financed by National Science Centre, projects no. 2012/05/B/NZ2/01627 and 2857/B/P01/2010/39, CMHI S126/12 and EU Structural Funds, project POIG.02.01.00-14-059/09.

2366F

A synonymous polymorphic variation in ACADM exon 11 affects splicing efficiency and may affect fatty acid oxidation. *G.H. Bruun, T.K. Doktor, B.S. Andresen.* Biochemistry and Molecular Biology (BMB), University of Southern Denmark, Odense, Denmark.

Medium-Chain acyl-CoA dehydrogenase (MCAD) is an important gene in the mitochondrial fatty-acid beta-oxidation of medium-chain fatty acids and its deficiency results in the most frequent fatty acid oxidation defect in humans. In recent studies combining genome-wide association and metabolomics, a single-nucleotide polymorphism (SNP), rs211718C>T, located far upstream of the MCAD gene (ACADM) was found to be associated with improved beta-oxidation of medium-chain fatty acids. We examined the functional basis for this association and identified linkage between rs211718 and the intragenic synonymous SNP c.1161A>G in ACADM exon 11 (rs1061337). Employing ACADM minigenes we show that c.1161A is associated with aberrant splicing of exon 11, which is corrected by c.1161G. This may result in more full length MCAD from c.1161G alleles. RNA-oligonucleotide affinity purification analysis suggests that the improved splicing of the c.1161G allele is due to changes in the relative binding of the splicing regulatory proteins SRSF1 and hnRNP A1, indicating that this sequence variation either abolish an exon splicing enhancer (ESE) or create an exon splicing silencer (ESS) or both. Analysis of publicly available RNA-seq data from several hundred cell lines, show significantly more reads with c.1161G than with c.1161A in heterozygous individuals, supporting that c.1161A>G is a functional SNP, which leads to higher MCAD expression, perhaps due to improved splicing. We used the same minigene based approach for functional testing of other sequence variations in ACADM exon 11, among them the prevalent disease-causing c.985A>G mutation. These results further supported the notion that nucleotide changes in ACADM exon 11 are not neutral and may have effects by affecting splicing efficiency. This study is a proof of principle that synonymous SNPs are not neutral and that all single nucleotide changes may potentially affect splicing. By changing the binding sites for splicing regulatory proteins they can have significant effects on pre-mRNA splicing and thus protein function.

2367T

Patients affected with Fabry disease have an increased prevalence of Raynaud's phenomenon: an investigation of 222 patients. *D.P. Germain[1, 2], O.I. Atanasiu[2], A. Cordier[2], K. Benistan[2].* 1) Division of Medical Genetics, University of Versailles, Garches, France; 2) Referral Center for Fabry disease, Hôpital Raymond Poincaré, Garches, France.

Background: Fabry disease (FD, OMIM 301500) is an X-linked inborn error of glycosphingolipid metabolism due to the deficient activity of alpha-galactosidase A, a lysosomal enzyme. FD is the paradigm of a multi-systemic disease with a variety of signs and symptoms and a large phenotypic variability. While the progressive systemic deposition of globotriaosylceramide (Gb3) and secondary metabolites (lysoGb3) is known to have protean clinical manifestations with, in particular, vascular and autonomic nervous system involvement, the prevalence of Raynaud's phenomenon was not previously studied. **Patients and methods**: the presence of Raynaud's phenomenon was non-invasively investigated in 222 consecutive patients affected with classic (n = 207) or late onset (n = 15) FD, through medical history assessment, questionnaire, photographs and physical examination. All clinical cases of FD were confirmed by alpha-galactosidase A enzymatic assay and/or genotyping of the GLA gene as part of an exhaustive baseline evaluation prior to enzyme replacement therapy. **Results**: a total of 34 patients (15.3%) affected with classic FD were found to have Raynaud's phenomenon. **Discussion**: this is the first evidence of an increased prevalence of Raynaud's phenomenon in patients affected with Fabry disease. This further expands the clinical phenotype of the disease and may constitute an additional endpoint to monitor the outcome of enzyme replacement therapy or emerging therapies (active site specific chaperones). Studies are ongoing to characterize the pathophysiologic mechanism(s) of Raynaud's phenomenon in Fabry disease.

2368F

Natural History of Gangliosidosis, and Therapeutic Interventions. *J. Utz[1], R. Zeigler[2], J. Raymond[3], P. Karachinski[3], C.B. Whiltey[1].* 1) Adv Therapies, Pediatrics, Univ Minnesota, Minneapolis, MN; 2) Neuropsychology, Pediatrics, Univ Minnesota, Minneapolis, mn; 3) Neurology, University of Minnesota, Minneapolis, MN.

Background: The ganglioside accumulation has been implicated in the mechanisms of central nervous system (CNS) neurodegeneration in several lysosomal diseases including Tay-Sachs disease, Sandhoff disease, GM1-gangliosidosis and even the mucopolysaccharidoses. The ongoing natural history study has improved understanding of MRI findings and neurocognitive decline, as well as identifying new findings of clinical presentation in gangliosidoses that may be useful as clinical outcome measurements, and include increased intracranial hypertension,cardiac anomolies, gum tissue and dental disease, rapidly progressing megencephaly, and neurogenic bladder. Monotherapies such as miglust, an inhibitor of glucosylceramide synthase in the glycosphingolipid pathway, and chaperone theray with pyrimethamine have yielded disappointing results with no benefit observed in most cases. Microglial activation with release of inflammatory mediators is increasingly recognized as an important aspect of ganglioside-associated CNS damage and may provide an additional therapy target. Hypothesis: The DANSER regimen, an oral combination therapy for the gangliosidoses that targets glycosphingolipid precursor accumulation while simultaneously targeting microglial activation and downstream inflammatory processes, may provide a more comprehensive treatment strategy intended to slow and/or stabilize disease progression of the gangliosidoses. Methods: The DANSER regimen includes miglustat with ketogenic diet to enhance bioavailability of miglustat and minimize its side effects, pyrimethamine to reduce ganglioside accumulation for GM2 diseases, ibuprofen and anti-oxidant therapy to reduce damaging CNS inflammatory processes. Outcome monitoring at baseline, 4 to 6 months, 12 months, and then annually thereafter, includes volumetric MRI, neurocognitive testing, neurological exam, serum and CSF inflammatory markers, leukocyte enzyme activity, and ophthalmology exams. Natural history data and DANSER regimen data provide a retrospective and prospective assessment of natural disease pathology and the safety and efficacy of of a combined therapy approach. Results: To date, 10 patients have provided natural history data and started DANSER regimen. Overall, the DANSER regimen has been well tolerated. Neuropsychiatric evaluations have shown improvement in some domains for patients using the goal miglustat dose. These preliminary results warrant continued study.Supported by NIH U54NS065768.

2369T

Biochemical and hematologic laboratory studies in a cohort of patients with Barth Syndrome. *H. Vernon[1, 2], R. McClellan[2], A. Cordova[1], R. Kelley[2], Y. Sandlers[2].* 1) Inst Gen Med, Johns Hopkins Hosp, Baltimore, MD; 2) Kennedy Krieger Institute, Baltimore, MD.

Barth Syndrome is a rare X-linked disorder characterized by dilated cardiomyopathy, skeletal myopathy and neuropenia. It is caused by defects in Tafazzin, an enzyme responsible for modifying the acyl chain component of cardiolipin. In order to gain a better understanding of clinical laboratory values in this rare disease, we measured hematologic and biochemical values in a cohort of patients. We analyzed urine organic acids, red blood cell fatty acids, and plasma 3-methylglutaconic acid (3-MGC) in 28 individuals with Barth Syndrome from ages 10 months to 30 years. The average plasma 3-MGC level in patients was 1087.5 nmol/L +/- 435 (nl. 162+/- 68 nmol/L). Our previous principal component analysis of urine organic acids in Barth Syndrome patients revealed elevations in beta hydroxyisobutyrate, aconitate, 3-MGC, and 2-ethylhydracryate compared to controls. In this current cohort, the average urine values were: aconitate 34.9 µg/mg Creatinine (+/- 16 µg/mg Cr), beta hydroxyisobutryate 21.7 µg/mg Creatinine (+/- 27 µg/mg Cr), 3-MGC 44.6 µg/mg Creatinine (+/- 25 µg/mg Cr), and 2-ethylhydracryate 6.7 µg/mg Creatinine (+/- 25 µg/mg Cr). We measured 55 fatty acids in red blood cells, and found only that C16:1 trans fatty acid was significantly different than control values, with Barth Syndrome patients having an average level of 0.312 ug/mL +/-0.08 vs. normal controls with an average level of 0.820 ug/mL +/-0.2. Complete blood counts were measured in 17 affected individuals, from age 3y11m to 30y2m. Average white blood cell count was 5664 cells/mL +/- 1798, average hemoglobin was 13.5 mg/dL +/- 1.4, average absolute neutrophil count (ANC) was 1935 cells/mL +/- 1225, and average absolute monocyte count (AMC) was 894 cells/mL +/- 449. The range for ANC was 300-4900 cells/mL, with 4 of 17 patients having neutrophil counts <1000 cells/mL. The range for AMC was 500-2400 cells/mL, with 5 of 17 patients having monocyte counts at or above 1000 cells/mL. Prealbumin levels were measured in 18 affected individuals, from ages 3y11m-30y2m months. Values ranged from 8 to 22 mg/dL (nl. 20-40), with an average of 16.9 +/- 4.02. 13 of 18 measurements were below the lower limit of normal. R-squared analysis revealed no correlation between age and these laboratory parameters. Our studies augment the knowledge of the laboratory values to be expected in individuals with Barth Syndrome, and expand upon the expected biochemical laboratory profiles.

2370F

The emerging natural history of cross-reactive immunologic material (CRIM)-negative Infantile Pompe disease patients treated with recombinant human GAA. *K.B. Sheets, S.G. Banugaria, S.M. DeArmey, D.S. Bali, C.W. Rehder, P.S. Kishnani.* Duke University Medical Center, Durham, NC.

Deficient lysosomal acid alpha-glucosidase (GAA) leads to glycogen accumulation resulting in progressive, debilitating muscle weakness. Enzyme replacement therapy (ERT) with recombinant human GAA (rhGAA) prolongs survival in patients with Infantile Pompe disease (IPD); however cross reactive immunologic material (CRIM)-negative (CN) status is a well-appreciated prognostic factor influencing poor clinical outcome. CN patients lack residual GAA expression and typically mount an immune response with high sustained antibody titers (HSAT) resulting in significant clinical decline and ultimately death at median age of 27.1 months despite continued ERT. Immune tolerance induction (ITI) with ERT initiation is now standard care for CN IPD. Nonetheless, sporadic CN cases lacking significant immune responses have emerged in recent literature. We aimed to assess natural history and treatment outcomes in CN IPD to determine how often patients mount immune responses against rhGAA. As part of an Institutional Review Board approved study we completed retrospective analysis of outcome measures for 20 CN IPD patients receiving ERT monotherapy. Median age at diagnosis was 2.5 months and 3.5 months at ERT initiation. Of the 20, 45% patients required ventilation by median age 11.3 months [8.2-25.3 months]. Most developed anti-rhGAA antibodies (17/20), with 82.4% (14/17) mounting HSAT [median peak 204800; range:51200-1638400]. Three maintained intermediate antibody titers [peak values 25600 and 51200], while three had low antibody titers peaking at 200 and 800. Median age of death for 15/20 patients was 29.6 months [14.7-51.4 months]. Of the remaining five, the single patient with HSAT at 76 months of age is the only case to receive ITI in the entrenched setting, as previously published. One patient at 69 months has intermediate titers and antibodies values continue to fall without ITI. Three patient outliers at 73, 62 and 39.5 months old maintain low titers without ITI. Variable immune responsives in this small subset of CN IPD warrants further investigation into the influence of genotype, age at start of ERT, potential HLA involvement and other genomic factors. Our data provide further evidence that CN IPD patients receiving ERT alone develop HSAT with reduced survival and poorer clinical outcomes, which necessitates prophylactic ITI to prevent HSAT as opposed to prolonged ITI treatment in the entrenched setting with greater safety risks and less likelihood of success.

2371T

Hepatic adaptations during infection: implications for inborn errors of metabolism. *P. McGuire[1], L.N. Singh[1], S. Matsumoto[2], K. Saito[2], T.N. Tarasenko[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) National Cancer Institute, National Institutes of Health, Bethesda, MD.

Although individually rare, the inborn errors of metabolism (IEM) constitute a medically important class of disorders, the untreated clinical course of which can feature severe neurologic disease and even death. Viral infections play a significant role in precipitating life-threatening acute decompensations in IEM. Acute decompensations are defined as a functional deterioration in metabolic status and may result in a range of hepatic metabolic perturbations including hypoglycemia, acidosis and hyperammonemia, depending on the disorder. We hypothesized that hepatic metabolic adaptations or perturbations that normally occur during infection will not be tolerated in IEM. To help define the hepatic metabolic response to infection, we employed mouse models of respiratory infection (influenza A/PR/8) and simulated viremia in C57Bl/6 (B6). After 5 days of infection with influenza, B6 mice displayed hepatic sensitivity characterized by increased AST and ALT and liver mtDNA depletion. mRNA microarray studies revealed antiviral and immune response signatures suggestive of hepatic sensing of influenza. Global biochemical profiling using a multi-tiered (i.e. GC-MS and LC-MS/MS) metabolomic platform showed significant reductions of long chain fatty acids and components of coenzyme A synthetic pathway consistent with increased beta- oxidation. To simulate immune activation due to viremia, mice were injected with the viral dsRNA mimic, poly I:C. During this sterile immune activation, hepatic metabolomic profiling showed significant depletion of several cofactors and their precursors including nicotinamide, NAD+, FAD, riboflavin and FMN. Since these co-factors are involved in various aspects of energy metabolism, we profiled in vivo mitochondrial function by MR spectroscopy using hyperpolarized ^{13}C-pyruvate. Following intraperitoneal poly I:C treatment for 3 days, B6 mice displayed significant increases in ^{13}C-lactate and 13C-alanine, consistent with mitochondrial perturbations and decreased entry of ^{13}C-pyruvate into the TCA cycle. Overall, our findings suggest that infection may result in increased utilization of long chain fatty acids, which may be functionally impaired if the carnitine cycle or fatty acid oxidation is affected by the underlying IEM. In addition, immune activation may also have more direct effects on hepatic energy metabolism, which has implications for patients with mitochondrial disease and other IEM with significant mitochondrial dysfunction.

2372F

Distinctive neurocognitive and neuropsychiatric phenotype for mucopolysaccharidosis type I (Hurler-Scheie syndrome) associated with mutation L238Q of the alpha-L-iduronidase gene. *A. Ahmed, C. Whitley, R. Cooksley, E. Shapiro.* Department of Pediatrics, University of Minnesota, Minneapolis, Minnesota.

The lysosomal enzyme alpha-L-iduronidase hydrolyzes terminal iduronic acid from heparan sulfate and dermatan sulfate, and is an essential step in GAG degradation. Mutations of its gene, IDUA, yield a spectrum of mucopolysaccharidosis (MPS) type I clinical disorders. The IDUA mutation, c.712T>A (p.L238Q) was previously noted as a mild mutation. In a longitudinal study of MPS brain structure and function (Lysosomal Disease Network), we found this mutation in 6 of 23 Hurler-Scheie syndrome patients. We hypothesized that L238Q, when paired with a nonsense mutation, is significantly more severe than other missense-nonsense combinations. Methods: Of 6 patients with a L238Q mutation, the L238Q allele was paired with a nonsense mutation in 4 patients, paired with a deletion in 1, and with a splice site mutation in another. This group was compared to 6 Hurler Scheie patients closely matched in age with 3 with missense-nonsense, 2 missense-missense, and 1 splice site-nonsense genotypes. Medical history was compiled into a Physical Symptom Score (PSS). Assessment of IQ, attention, memory, spatial ability, adaptive function and psychological status were measured. Results: No group differences were found in age at evaluation (17.8 and 19.3 years), duration of ERT, or PSS. By history, all were reported to be average in IQ (4/6 with documentation) in early childhood. All (100%) of the L238Q group had a psychiatric history and sleep problems compared to none (0%) of the comparison group. Significant differences were found in depression and withdrawal on parent report measures. IQ was lower in the L238Q group (mean IQ 74) than the comparison group (mean IQ 95; $p < 0.008$). Attention, memory, and visual-spatial ability scores were also significantly lower. Three occurrences of shunted hydrocephalus, and 4 of cervical cord compression were found in the L238Q group; the comparison group had1 occurrence of unshunted hydrocephalus and 2 of cord compression . Discussion: The missense mutation L238Q, when paired with a nonsense mutation, is associated with significant, late-onset brain disease: psychiatric disorder, cognitive deficit, and general decline starting at a later age than in Hurler syndrome with a mutation-related rate of GAG accumulation and its pathologic sequelae. This particular genotype-phenotype may provide insight into the genesis of psychiatric illnesses more broadly. Consideration of methods for early, brain-targeted treatment in these children is necessary.

2373T

Oral manifestations in a patient with infantile onset Pompe disease: newly recognized findings.
L.S. Kehoe[1], T. Anupama[2], P. Kantaputra[3]. 1) 1.Division of Genetics and Metabolism, Children's National Medical Center, Washington, DC; 2) 2.Department of Dentistry, Children's National Medical Center, Washington DC, USA; 3) 3.Craniofacial Genetics Laboratory, Division of Pediatric Dentistry, Department of Orthodontics and Pediatric Dentistry, Faculty of Dentistry, Chiang Mai University, Chiang Mai 50200, THAILAND.

Pompe disease is a lysosomal storage disease caused by mutations in the acid alpha glucosidase (GAA) gene. Mutations in GAA impair glycogen degradation, leading to excessive glycogen storage in lysosomes. The clinical spectrum of the disease ranges from a rapidly progressive infantile form to a slowly progressive late-onset form. The hallmarks of the classic infantile form are progressive, generalized muscle weakness and hypertrophic cardiomyopathy. A 9-year-old female first presented with weakness and hypertrophic cardiomyopathy at 3 months old. The diagnosis of infantile onset Pompe was confirmed by enzyme and molecular analyses. Enzyme replacement therapy (ERT) was started at age 5 months. Over time she has experienced generalized muscle weakness and has been G-tube dependent. A recent echocardiogram was normal. Physical examination showed myopathic facies, bilateral ptosis, hypernasality and generalized muscle weakness with no ectodermal dysplasia features. Oral examination revealed malocclusion related to mid-face hypoplasia and prognathic mandible. Panoramic radiographs at age 8 years demonstrated very large pulp chambers with poorly formed dentin of the maxillary permanent central and lateral incisors. The tooth roots were extremely short. Taurodontism of the maxillary permanent first molars and hypodontia were noted. There is no family history of hypodontia or other dental anomalies. Dental and craniofacial abnormalities in infantile onset Pompe disease have been understudied. Malocclusion has been described in at least one patient with the infantile form. Our patient presented with hypodontia, dental anomalies and anterior crossbite secondary to mandibular prognathism and hypoplastic maxilla which have never been reported in infantile onset Pompe survivors. Tooth agenesis or hypodontia is one of the most common congenital anomalies in humans and can be an isolated finding or be part of a genetic syndrome. Mutations in several genes involved in early craniofacial development and transcription factors cause isolated hypodontia. The correlation between GAA gene function and tooth agenesis is unknown. Further studies are strongly suggested. ERT has modified the natural history of the disease. Patients live longer leading to the discovery of new findings. Dental anomalies in Pompe disease might be an overlooked part of this syndrome. Since early dental care may improve orodental health, a thorough oral examination is encouraged.

2374F

Assessment of Bone Mineral Density by Dual Energy X-ray Absorptiometry in Patients with Mucopolysaccharidoses.
H. Lin[1,2,3,4], S. Lin[1,2,3], C. Chuang[1], M. Chen[1,2,3], D. Niu[4,5]. 1) Departments of Pediatrics and Medical Research, Mackay Memorial Hospital, Taipei, Taiwan; 2) Department of Medicine, Mackay Medical College, New Taipei City, Taiwan; 3) Department of Early Infant Care and Education, Mackay Medicine, Nursing and Management College, Taipei, Taiwan; 4) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan; 5) Department of Pediatrics, Taipei Veterans General Hospital, Taipei, Taiwan.

Patients with mucopolysaccharidoses (MPS) have a skeletal dysplasia that is associated with poor bone growth and mineralization. However, only limited information exists regarding the assessment of bone mineral density (BMD) in relation to age and treatment in these disorders. Therefore, we performed dual energy x-ray absorptiometry (DXA) to characterize lumbar spine BMD in 30 MPS patients. Our study population was all Asian, mostly male (70%), and predominantly pediatric (87% <19 years of age). The median age was 10.8 years (range 5.0 years to 23.7 years). MPS types included I (n=2), II (n=12), IIIB (n=2), IVA (n=9), and VI (n=5). For the 26 patients <19 years of age, standard deviation scores (Z-scores) for height, weight, body mass index (BMI), and BMD were -4.53 ± 2.66, -1.15 ± 1.55, 0.74 ± 1.23, and -3.03 ± 1.62, respectively, and all were negatively correlated with age ($p < 0.05$). There were 21 (81%), 8 (31%), 0 and 18 patients (69%) below -2 Z-score for these four parameters, respectively. However, after correction for height-for-age Z-score (HAZ), HAZ adjusted BMD Z-score was -0.7 ± 1.24. Eight patients (31%) met our center's criteria for osteopenia (HAZ adjusted BMD Z-score <-1), and 4 patients (15%) had osteoporosis (HAZ adjusted BMD Z-score <-2). Of 8 patients with MPS I, II or VI who underwent follow-up DXA after receiving ERT for 1.0 to 7.4 years, all showed an increase in absolute BMD values, and 3 out of 5 patients who started ERT at the pre-pubertal age revealed an improvement in HAZ-adjusted BMD Z-score. These findings and the follow-up data can be used to develop quality of care strategies for patients with MPS.

2375T

A Retrospective Review of the Natural Course of Mucopolysaccharidosis VII.
A. Montano[1,2], N. Lock-Hock[3], R. Steiner[4], B.H. Graham[5], M. Szlago[6], R. Greenstein[7], M. Pineda[8], A. Gonzalez-Meneses[9], M. Çoker[10], D. Bartholomew[11], M. Sands[12], R. Wang[13], R. Giugliani[14], A. Macaya[15], G. Pastores[16], A.K. Ketko[17], F. Ezgü[18], A. Tanaka[19], L. Arash[20], M. Beck[20], R.E. Falk[21], K. Battacharya[22], J. Franco[23], K. White[24], G. Mitchell[25], L. Cimbalistiene[26], W.S. Sly[2]. 1) Dept Pediatrics, Saint Louis University, St Louis, MO,USA; 2) Dept Biochemistry and Molecular Biology, Saint Louis University, St Louis, MO,USA; 3) Metabolic and Clinical Genetics, Kuala Lumpur Hospital, Kuala Lumpur, Malaysia; 4) Oregon Health Sciences University, Portland, OR, USA; 5) Baylor College of Medicine, Houston, TX,USA; 6) Laboratorio de Neuroquimica, FESEN, Buenos Aires, Argentina; 7) University of Connecticut Health Center, CT, USA; 8) Sant Joan De Deu Hospital, Barcelona, Spain; 9) Virgen del Rocío Hospital, Sevilla, Spain; 10) Ege University, Faculty of Medicine, Izmir, Turkey; 11) Children's hospital, Columbus, OH, USA; 12) Washington University in St Louis, St Louis, MO, USA; 13) Children's Hospital Orange County,CA, USA; 14) Medical Genetics Service/HCPA & Department of Genetics/UFRGS, Porto Alegre, Brazil; 15) Hospital Universitari Vall d'Hebron, Barcelona, Spain; 16) New York University Medical Center, NY, USA; 17) University of Michigan Health Systems, Ann Arbor, MI,USA; 18) Gazi University Faculty of Medicine, Ankara, Turkey; 19) Osaka City University Graduate School of Medicine, Osaka, Japan; 20) Childrens Hospital, Johannes Gutenberg University Medical Center, Mainz, Germany; 21) Genetics Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA; 22) Genetic Metabolic Disorders Service, The Children's Hospital at Westmead, Sydney, Australia; 23) São Paulo University, Sao Paulo, Brazil; 24) Seattle children's Hospital, Seattle, WA, USA; 25) Department of Pediatrics, Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montreal, Canada; 26) Vilnius University Hospital, Santariskiu Klinikos, Vilnius, Lithuania.

Mucopolysaccharidosis VII (MPS VII; Sly syndrome) is a lysosomal storage disorder caused by the deficiency of β-glucuronidase (GUS). To understand the natural history of MPS VII, extensive knowledge of the spectrum of disease severity, the rate of progression and distribution of specific symptoms in untreated patients with MPS VII is needed. In this study, we investigated the onset, severity, management and natural progression of the disease via a retrospective review of MPS VII cases. Anonymous information of signs and symptoms of MPS VII patients was collected by conducting interviews to physicians and by reviewing published literature. To date, 112 patients from 23 countries were analyzed. Of these, 64 patients had an infantile or adolescent form of the disease and 48 patients were classified with non-immune infantile hydrops fetalis (NIHF). About 70 percent of patients with the infantile or adolescent forms of the disease had symptoms one year before diagnosis. In contrast, most patients with NIHF were diagnosed during pregnancy, as early as 20 weeks of gestation. Some patients survived only a few months and died from respiratory, renal or cardiac insufficiencies. A few improved thereafter resembling later onset forms. The presentation of the disease is heterogeneous with phenotypes ranging from severe to mild exhibiting normal intelligence. Our study shows that over 50% of individuals with late onset forms are characterized mostly with macrocephaly, coarse facies, pulmonary and cardiac disease, hepatosplenomegaly, cognitive impairment, short stature, and dysostosis multiplex. The data collected in this study will help to establish the natural history of MPS VII disease and to determine potential clinical endpoints.

2376F

Cobalamin C disease with fulminant hyperammonemic presentation in the neonatal period. *N.S. Abul-Husn, A.C. Yang, C. Yu, H. Chen, G.A. Diaz, J.D. Weisfeld-Adams.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Combined methylmalonic acidemia (MMA) and homocystinuria, cobalamin C type (cblC disease) is the most common inherited disorder of cobalamin (vitamin B12) metabolism. CblC disease is caused by mutations in the *MMACHC* gene, which lead to failed synthesis of adenosylcobalamin and methylcobalamin, essential cofactors for the mitochondrial enzyme methylmalonyl-CoA mutase (encoded by the *MUT* gene) and the cytosolic enzyme methionine synthase, respectively. CblC disease typically presents with progressive neurological deterioration in the newborn period, with later progression to multi-systemic disease. Although the clinical manifestations can vary, cblC disease is phenotypically quite distinct from either isolated MMA or isolated homocystinuria. Here, we describe the unusual case of a neonate from consanguineous parents who presented with severe hyperammonemic encephalopathy and metabolic acidosis, a presentation not typically seen in cblC disease. Newborn screen showed elevation of C3 acylcarnitine (12.56 µmol/L; normal range < 7.00 µmol/L) with low plasma methionine (0.18 µmol/L; normal range > 0.25 µmol/L). Initial plasma ammonia level was 1150 µg/dL (normal range 15 - 50 µg/dL). Plasma homocysteine was measured at 146 µmol/L (normal range 5 - 15 µmol/L), suggesting a defect in both adenosylcobalamin and methylcobalamin (i.e. cblC, cblD, or cblF disease). Urine methylmalonic acid was massively elevated at 20,482 mmol/mol creatinine (normal range < 5 mmol/mol creatinine), similar to levels observed in cases of isolated MMA caused by *MUT* deficiency prior to initiation of treatment at our center. Post-mortem studies showed homozygosity for the common c.271dupA allele in *MMACHC* in DNA recovered from patient fibroblast cultures, confirming the diagnosis of cblC disease. Further studies using next-generation sequencing revealed a heterozygous missense mutation (c.1991C>T) in the *MUT* gene that is predicted to be pathogenic using standard *in silico* analyses. This suggests that variants in other genes associated with cobalamin metabolism may play a role in modifying the clinical presentation of cblC disease, resulting in catastrophic presentations early in the neonatal period.

2377T

Molecular analysis of patients diagnosed as *cblC*. *A. Brebner, D. Watkins, D.S. Rosenblatt.* Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

Patients with combined homocystinuria and methylmalonic aciduria and mutations in the *MMACHC* gene constitute the *cblC* complementation group, the most common inborn error of vitamin B_{12} (cobalamin) metabolism. Approximately 550 patients are known to exist worldwide. Sanger sequencing of *MMACHC* in 39 *cblC* patients diagnosed by somatic cell complementation identified at least one mutation in 36. Four of these mutations were novel: two missense mutations [c.158T>C (p.L53P) and c.566G>A (p.R189H)]; one nonsense mutation [c.292C>T (p.Q98X)]; and a mutation that results in a deletion of exon 2. The missense mutation c.158T>C was seen in four patients that were asymptomatic at the time of ascertainment. Three were found by newborn screening for methylmalonic aciduria, and one from a metabolic screen in an adult with a presumably unrelated cancer. One patient developed symptoms only when she was 14 weeks pregnant, at age 24. In three cases the c.158T>C mutation was found in combination with a previously known *MMACHC* mutation. In the fourth, no second mutation was identified within the coding sequence, in the 5' and 3' UTRs, or in the *MMACHC* promoter. Sequencing of cDNA identified c.158T>C in apparent hemizygosity, suggesting an unidentified mutation in a non-coding region that eliminated expression of the second allele. The c.566G>A mutation was identified in one asymptomatic patient that was picked up on newborn screen. Both missense mutations affect conserved residues and were predicted to be probably damaging by Polyphen-2 with scores of 1.000. The nonsense mutation c.292C>T was seen in one severely affected early onset patient who presented with seizures, hypotonia and poor feeding. In another patient that died in infancy, sequencing of gDNA seemed to show the patient was homozygous for the common c.271dupA mutation. However amplification of the cDNA showed a heterozygous deletion of exon 2; there were two products, one normal size and the other 200bp shorter. Sequencing demonstrated the shorter product was lacking exon 2. This brings the total number of mutations seen in *cblC* patients to 80. In one of the three patients with no identified *MMACHC* mutation, re-examination of results of cellular studies showed an ambiguous complementation assay and a biochemical picture consistent with the *cblF* or *cblJ* disorder. Sequencing of the *LMBRD1* gene revealed that this patient was homozygous for the common c.1056delG mutation, demonstrating the diagnosis of *cblF*.

2378F

Glycemic deregulation in Congenital Central Hypoventilation Syndrome. *F. Moreau[1], K. Braun[2], J. Amiel[3], T. Dery[4], G. Jedraszak[4], S. Goudjil[1], G. Kongolo[1], H. Trang[5], P. De Lonlay[6], M. Mathieu[4], A. Leke[1], G. Morin[4].* 1) Pediatric Reanimation, Amiens University Hospital, Amiens, France; 2) Pediatric Endocrinology, Amiens University Hospital, Amiens, France; 3) Genetic department, Necker-Enfants Malades Hospital, Paris, France; 4) Genetic department, Amiens University Hospital, Amiens, France; 5) Ondine curse rare disease reference center, Robert Debré Hospital, Paris, France; 6) Inherited metabolic diseases rare disease reference center, Necker-Enfants Malades Hospital, Paris, France.

Background Congenital Central Hypoventilation Syndrome (CCHS; OMIM 209880) is a life-threatening disorder with an impaired ventilatory response to hypercarbia and hypoxemia. This core phenotype is associated with lower-penetrance anomalies of the autonomic nervous system including Hirschsprung disease and tumors of neural-crest derivatives. CCHS is due to heterozygous mutations of PHOX2B gene, mainly polyalanine repeat expansions. In the past 12 years, recurrent hypoglycemia has been reported in 6 CCHS patients and was always related to high plasmatic insulin level. Interestingly, a recent study in CCHS patients observed a chronic tendency to hyperglycemia. Case report The delivery of this female baby occurred by caesarian section at 40 WG for fetal bradycardia. Apgar score was 9/9/9. Apneas led to tracheal intubation. Abdominal distension required a discharge colostomy and Hirschsprung disease was diagnosed. After recurrent sleeping apneas, the diagnosis of CCHS was confirmed (PHOX2B: c.590dup, p.Gly199ArgfsX161 - J. Amiel). First hypoglycemia was observed at 6 weeks of life, consecutively to seizure. The hypoglycemias were profound, recurrent, occurred after 1.5-2.5 hours fast, without ketone bodies, sensitive to Glucagon. Insulinemia was 21.1 µUI/ml (glycemia 1.9mmol/L). Enrichment of alimentation with carbohydrates and diazoxide (10-15 mg/kg/d) allowed a good control. During continuous glucose monitoring, recurrent hyperglycemia was observed. Discussion Hyperinsulinism in CCHS patients is probably not coincidental, but the characteristics of the hypoglycemias are different of those observed in hyperinsulinism by channelopathies, suggesting a different mechanism. Insulin hypersecretion in CCHS could be secondary to pancreatic hypertrophy, PHOX2B-/-mice presenting an increased number of β cells. An abnormal function of carotid body glomus cells could also be suspected. Functions of glomus cells combine O2 and glucose sensor to prevent neuronal damage by acute hypoxia and/or hypoglycemia. Carotid glomus cells derive from neural crest and CCHS is a disease of neural crest derivatives. In two CCHS patients, the corpus bodies were found smaller than controls with a marked decrease in the number of glomus cells. But carotid bodies are histologically normal in mice harboring a heterozygote +7 alanines expansion. However, the role of fast in hypoglycemias, and the coexistence of hyperglycemias suggest carotid glomus insensitivity to glucose.

2379T

Fabry Disease in Northern Ireland.
F.J. Stewart[1], A. Muir[2], J. McOsker[2], T. Jardine[2], A. Wilson[1], P. McKeown[2]. 1) Dept Genetic Medicine, Belfast City Hosp, Belfast, United Kingdom; 2) Belfast Heart Centre, Royal Victoria Hospital, Belfast, United Kingdom.

Fabry disease is a multi-system lysosomal storage disease due to a deficiency of alpha galactosidase. It is caused by mutations in the GLA gene which is on the X chromosome. It differs from many X-linked disorders in that females frequently show clinical features. Affected individuals may show the classical form or a more attenuated form. Treatment for this condition is available in the form of enzyme replacement therapy with agalsidase beta (Fabrazyme) and agalsidase alpha (replagal). The incidence has been estimated at approximately 1:50 000 males (Desnick 2001). However a newborn screening study suggested an incidence of 1:3000 with an 11:1 ratio of individuals with the more attenuated phenotype (Spada 2006). Northern Ireland has a population of 1.7 million. To date we have found 11 families who have Fabry disease within which are 46 individuals with a GLA mutation. All individuals with a mutation are followed up on an annual basis at our multidisciplinary Fabry clinic. Three families have the p.A13P mutation. All other mutations are seen in only one family and they are Exon 1 deletion, p.R392SfsX2, p.R220X, c.144delG, p.W209X, p.A13T, p.D313G and c.802_804delCA. 10 patients are aged under 18. The eldest female is aged 79. The eldest male is 75. The typical skin angiokeratoma are seen in 6 individuals - all male. 16 show signs of cardiac involvement. Two patients have had small strokes but most over the age of 30 show some bright changes on MRI. No patients have renal failure though some show evidence of mild proteinuria. 11 patients are on treatment with agalsidase alpha and 7 with agalsidase beta. The remaining patients are monitored on an annual basis with cardiac and renal investigations. They have either declined ERT treatment or have no clinical features to warrant treatment at this stage. The majority of our patients have come from family follow up of the index cases and the offer of genetic testing to at risk relatives. This may explain why we have a significant number of individuals not currently on treatment. These figures suggest an incidence of at least 1:37000 in our population. We feel that we do not have 100% ascertainment of cases and therefore we believe this may be an underestimate of the true incidence. We feel that family follow up of newly diagnosed cases and genetic testing of at risk relatives is essential as clinically asymptomatic individuals may miss out on the opportunity for monitoring and possible treatment.

2380F

Hepatocerebral mitochondrial DNA depletion syndrome with atypical brain MRI findings due to heterozygote mutations in the MPV17 gene including a novel mutation: clinical, biochemical and molecular analysis.
A. Iglesias. Dept Pediatrics, Div Med Gen, Columbia Univ Med Ctr, New York, NY.

Hepatocerebral mitochondrial DNA depletion syndrome is a known clinical entity. Mutations in the MPV17 gene are causative, but rare and the natural history of the phenotype is incomplete. Moreover, liver transplant is ethically controversial due to brain involvement. A case of a 13 month-old boy with the diagnosis is discussed. He was born full term after uneventful pregnancy. Parents are from Pakistan and are first cousins. Jaundice was diagnosed at 3 months improving by 5 months; but recurred at 9 months. At 1 year he was found to have elevated direct bilirubin, alkaline phosphatase, AST/ALT and AFP and negative metabolic evaluation except for high lactate and negative PFIC testing. Normal chromosomes and SNP-array with areas of homozygosity were done. Needle liver biopsy was done: giant cell hepatitis, cirrhosis with micro/macro vesicular steatosis; abnormal mitochondria were seen on EM. mtDNA testing on liver was inconclusive. Brain MRI: mild delay in myelinization in the internal capsule and white matter changes sparing cerebellum and basal ganglia. Liver transplantation was discussed, but not performed due to clinical severity and uncertain diagnosis. A muscle biopsy was performed: positive mtDNA depletion (>95%). Molecular testing for mtDNA depletion syndrome was done. Heterozygosity for 2 mutations in the MPV17 gene: c.284dupG [p.Phe96LeufsX17] and c.293 C>T [p.Pro98Leu (P98L) was found confirming the diagnosis. Both mutations are predicted to be pathogenic. The latter, P98L has been previously reported in Pakistan. The former, although predicted to be pathogenic, has not been previously reported. Due to clinical complications the patient passed away few days after the muscle biopsy without being transplanted. Therefore, although the diagnosis did not affect the patient's outcome, it helped to console the parents and plan for future pregnancies. In conclusion, this case well exemplifies the value of how a strong clinical suspicion, despite inconclusive liver biopsy results, can help to achieve a definitive diagnosis. Moreover, the brain MRI changes, although non-typical, raised concerns about the ethical decision about liver transplant. In hindsight, the diagnostic decisions taken seemed appropriate; but future characterization of additional cases with mutations in the MPV17 gene (especially this particular genotype) will be needed to address this clinical dilemma while teaching us about this fascinating condition.

2381T

Genetic Associations with Neonatal Acylcarnitine Levels.
C.J. Smith[1], A.M. Momany[2], S.L. Berberich[3], J.L. Murray[2], K.K. Ryckman[1]. 1) Department of Epidemiology, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) State Hygienic Laboratory, University of Iowa, Iowa City, IA.

Background: Short-chain acylcarnitines are critical for normal cellular metabolism. Variation in acylcarnitine levels may be indicative of a variety of adult metabolic deficiencies such as Type 2 Diabetes Mellitus and may be associated with specific genotypes. Recent genome-wide association studies of the adult human metabolome have identified genetic variants associated with multiple acylcarnitine levels; however, to our knowledge, no studies have examined this at birth. We have previously shown that acylcarnitine levels are highly heritable at birth. **Objective:** We examined single nucleotide polymorphisms (SNPs) in candidate genes implicated in adult metabolism with the normal variation in short-chain acyclarnine levels in newborns. **Methods:** Our analysis included genetic data and analyte measurements for 425 term born (40 weeks gestation) singleton neonates. We examined 10 short-chain acylcarnitines including C0, C2, C3, C3-DC, C4, C4-DC, C5, C5:1, C5-OH and C5-DC as measured by tandem mass spectrometry as part of routine newborn screening performed by the State of Iowa Hygienic Laboratory. We excluded from analysis any neonate that had an abnormal newborn screen, was on total parenteral nutrition or those for whom the sample was not obtained between 24 and 72 hours after birth. DNA was extracted from the residual dried blood spot card and 22 SNPs in 7 candidate genes were genotyped using the Fluidigm® SNP Genotyping platform. We analyzed each SNP-analyte level for association using Kruskal-Wallis tests. **Results:** Of the 220 SNP-analyte associations 8 were significant by Bonferroni correction (p-value <2×10-4). Three SNPs (rs3794215, rs3916, rs555404) in the short chain acyl-CoA dehydrogenase gene (ACADS) were associated with neonatal C4 levels. The GG genotype of rs3794215 (P=9.9×10-20), the CC genotype of rs3916 (P=2.0×10-27) and the CC genotype of rs555404 (P=1.6×10-17) were all associated with higher neonatal C4 levels. There were also significant (P<0.01) associations with 2 SNPs (rs11746555 and rs17622208) in the solute carrier family 22 (SLC22A5) gene with C0, C4-DC, C5 and C5-OH. **Conclusions:** We have replicated SNP associations observed between adult metabolic traits and the ACADS and SLC22A5 genes in newborns. This research has important implications for not only detection of rare inborn errors of metabolism but also life course epidemiology.

2382F

Presenting optional newborn screening in the context of mandatory newborn screening: The case of Duchenne muscular dystrophy.
S.E. Lillie[1], B.A. Tarini[2], B.J. Zikmund-Fisher[1]. 1) Department of Health Behavior and Health Education, University of Michigan, Ann Arbor, MI; 2) Department of Pediatrics, University of Michigan, Ann Arbor, MI.

Purpose: Many mandatory newborn screening (NBS) programs have expanded to include optional tests that require parents to make explicit decisions. Parents' decisions may depend not only on NBS information, but also the context in which it is presented. We examined whether presenting optional NBS for Duchenne Muscular Dystrophy (DMD) with or without the context of mandatory NBS would influence intended utilization of and attitudes towards DMD NBS. Methods: U.S. adults recruited from Amazon Mechanical Turk (N=2,991) read a hypothetical vignette and learned about DMD and DMD NBS for their (hypothetical) newborn boy. Participants were randomized to one of three vignettes that presented information about DMD NBS by itself (no context), with a description of mandatory tests that just listed the total number of tests ('bundled' mandatory context), or with a listing of each mandatory test ('unbundled' mandatory context). We assessed intended utilization and attitudes towards the optional test. Results: Most participants reported they would utilize DMD NBS (68.6%). Participants were more likely to choose DMD NBS if they also read about DMD NBS in the context of mandatory NBS (either bundled or unbundled) versus when the test was presented by itself with no larger context (Bundled: OR=1.54, p<0.01; Unbundled: OR=1.35, p<0.05). Participants who read about DMD NBS with no context of mandatory NBS reported such testing to be less important and reported that they would worry more about DMD NBS results than those provided with the context of mandatory NBS (F=3.40, p<0.05; F=3.48, p<0.05). Conclusions: Future optional NBS programs should describe how optional tests fit into overall mandatory NBS programs. Increasing attention to how the framing of communications about testing influences parental behavior will become more and more important as optional NBS programs continue to expand.

2383T

Lysosomal Disease Newborn Screening in Mexican Population. *J. Navarrete[1], A. Rivera[2], A. Limon[1], E. Covian[1].* 1) Dept Gen, Hosp Sur PEMEX, Mexico City, Distrito Federal, Mexico; 2) School of Medicine, Universidad Anahuac, Mexico Norte, Mexico City, Distrito Federal, Mexico.

The goal of newborn screening is almost always to make a specific diagnosis for the purpose of providing medical intervention to avoid or ameliorate symptoms. In Mexico we began with newborn screening since 1977 but with very few inborn errors of metabolism, phenylketonuria, galactosemia, congenital hypothyroidism, sickle cell anemia and cystic fibrosis, after 10 years other disorders including some aminoacidopathies have been included. Petroleos Mexicanos is a big institution with aproximately 10 thousand workers and their families. Since 2005 a larger newborn screening has been done to all newborns in this insitution through all the country. We test for most aminoacidopathies including organic acidurias, hemoglobinopathies, G6PD deficiency, adrenal hyperplasia, cystic fibrosis, biotindase deficiency and since August 2012 we included immunodeficiency syndrome, Gaucher disease, Niemann Pick(A/B) disease, Pompe disease, Krabbe disease, Fabry disease, MPS I. We present are results of 2500 newborns screened since last year, their follow up and early treatment.

2384F

Enhancing case detection of selected inherited disorders through expanded newborn screening in the Philippines. *C. Padilla[1,2].* 1) Newborn Screening Reference Center, National Institutes of Health, University of the Philippines Manila; 2) Department of Pediatrics, College of Medicine, University of the Philippines Manila.

Background. Newborn screening in the Philippines currently includes screening for 6 disorders - congenital hypothyroidism (CH), congenital adrenal hyperplasia (CAH), phenylketonuria (PKU), glucose-6-phosphate dehydrogenase (G6PD) deficiency, Galactosemia (GAL) and maple syrup urine disease (MSUD). With improved newborn screening technologies, the potential of screening more than 70 disorders is possible. Worldwide, there has been a move towards increasing the number of disorders included in newborn screening panels. The California Newborn Screening Program (CNSP) screens for over 70 disorders and its database includes a large number of Filipino newborns. Objectives. To ascertain the profile of Filipino newborns screened through the CNSP and to extrapolate these data to the Philippine newborn population in order to assess the potential value of expanding the Philippine newborn screening program. Methods. The newborn screening database of the CNSP was reviewed. Projections based on the California data were made relative to expanded newborn screening and related outcomes in the Philippines. Results: From 2005 to 2011, a total of 3,460,839 newborns were screened in the CNSP which included 111,127 Filipinos. Among the Filipinos, 199 were confirmed as having one of the screened disorders categorized as follows: endocrinology disorders (51); hemoglobinopathies (109); amino acid disorders (6); organic acid disorders (7); fatty acid disorders (10); and other disorders (16). Extrapolating these findings to the Philippine newborn population predicts the detection of significant additional cases of screened disorders including: 2180 hemoglobinopathies, 140 organic acid disorders, 200 fatty acid disorders, and 240 other disorders. Conclusion: Data from the CNSP show significant cases of serious disorders detected by newborn screening in Filipino babies that are not currently detectable in the limited newborn screening program in the Philippines. Expanding the panel of screened disorders to approximate that in the CNSP will result in significant numbers of additional case detections in the Philippines that will save lives and reduce unnecessary negative health outcomes as a result of early detection and treatment.

2385T

Hyposialylation in glomerulopathies is mitigated by N-acetylmannosamine therapy. *M. Huizing[1], M.C. Malicdan[1], P. Leoyklang[1], O. Okafor[1], T. Yardeni[1,2], M.F. Starost[3], P.M. Zerfas[3], Y. Anikster[2], A. Volkov[4], B. Dekel[5], J.B. Kopp[6], W.A. Gahl[1].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD, USA; 2) Metabolic Disease Unit, Sheba Medical Center, Tel Hashomer, Tel Aviv, Israel; 3) Division of Veterinary Resources, ORS, NIH Bethesda, MD, USA; 4) Department of Pathology, Sheba Medical Center, Tel Hashomer, Israel; 5) Pediatric Nephrology Unit, Sheba Medical Center, Tel Hashomer, Israel; 6) Kidney Disease Section, NIDDK, NIH, Bethesda, MD, USA.

Biallelic mutations in murine *Gne*, coding for UDP-GlcNAc 2-epimerase/ManNAc kinase, the key enzyme in sialic acid biosynthesis, result in glomerular disease with podocyte effacement due to hyposialylation. We showed that oral supplementation with the sialic acid precursor N-acetylmannosamine (ManNAc) ameliorated the proteinuria and improved the podocyte foot process architecture and glomerular sialylation status of mutant mice. A panel of fluorescent-labeled lectins (including WGA, SNA, HPA, and PNA) applied to kidney sections, indicated aberrant sialylation of predominantly O-linked glomerular glycans in mutant mice kidneys; this normalized after ManNAc treatment. Since hyposialylation has sporadically been suggested in human glomerulopathies, we applied the lectin panel to renal tissue sections from 40 patients with unexplained glomerulopathies. An unexpectedly high number of biopsies (8) had glomerular hyposialylation similar to that seen in our mouse model, indicating that this condition may occur relatively frequently, and also that ManNAc may be a therapy. To gather more preclinical data, we induced podocyte hyposialylation in mice by intraperitoneal injection of (*Vibrio cholera*) sialidase, removing sialic acids from glycans. Sialidase-injected mice developed proteinuria and renal failure in a dose-dependent manner. Their glomerular glycoproteins were hyposialylated and their podocytes were effaced, similar to our *Gne* knock-in mouse model. Importantly, oral prophylaxis and treatment with ManNAc significantly reduced their proteinuria and podocyte injury. Although the exact mechanisms and consequences of glomerular hyposialylation requires further study, oral ManNAc therapy could benefit patients with glomerular hyposialylation; ManNAc has minimal toxicity, is easily (orally) administered and could replace or supplement existing therapies. Moreover, ManNAc is currently being tested in a Phase 1 clinical trial for the treatment of the rare hyposialylation disorder GNE myopathy; it could be repurposed for trials in patients with glomerular hyposialylation.

2386F

THERAPEUTIC HYPOTHERMIA FOR HYPERAMMONEMIC METABOLIC CRISES: A PILOT STUDY. *U. Lichter-Konecki[1], J. Poeschl[2], D. Dimmock[3], S. Baumgart[4].* 1) Div Gen & Metabolism, Children's National Med Ctr, Washington, DC; 2) Neonatal Intensive Care, Dept. Pediatrics, Children's Hospital, University of Heidelberg, Heidelberg, Germany; 3) Division of Genetics, Dept. of Pediatrics, Medical College of Wisconsin, Milwaukee, WI; 4) Div Neonatology, Children's National Med Ctr, Washington, DC.

Background: Neonates with urea cycle disorders (UCDs) or organic acidemias (OAs) and acute hyperammonemia (HA) and encephalopathy are at great risk for brain injury, intellectual disability, and death. Nutritional support and renal replacement therapy are used to treat the severe hyperammonemia of neonatal onset disease. Neuroprotection during rescue treatment may improve neuropsycholigical outcome. Animal experiments and small clinical trials in acute liver failure indicate that therapeutic hypothermia (TH) is neuroprotective in HA. We report results of a feasibility and safety pilot study of whole body cooling for neonates with acute HA and encephalopathy. Methods: Encephalopathic, hyperammonemic neonates with symptoms of UCDs or OAs requiring dialysis were enrolled. The whole body of the patients was cooled to 33.5 °C +/- 1°C. After 72h they were slowly rewarmed. Data of age-matched historic controls were also collected for comparison. Results: Seven patients were cooled using the pilot study protocol and their data were compared to data of seven historic controls. All patients survived dialysis and TH, 6 patients were discharged home, five feeding orally. The main complication was hypotension. Conclusion: TH treatment for hyperammonemic neonates was feasible and safe when administered in intensive care units experienced in cooling and dialyzing neonates and managing metabolic crises. TH however adds to the complexity of the treatment of these critically ill children and should not be done unless proven efficacious in a randomized clinical trial.

2387T

Polymyositis: A Common Misdiagnosis For Late Onset Pompe Disease. *V. Kasturi[1], S. Jain[1], V. Juel[2], P. Kishnani[1].* 1) Medical Genetics, Duke University Medical Center, Durham, NC; 2) Department of Neurology, Duke University Medical Center, Durham, NC.

Background: Pompe disease (PD) is a progressive neuromuscular disorder caused by acid alfa glucosidase (GAA) deficiency leading to accumulation of glycogen in lysosomes. PD is broadly categorized into 2 subtypes. Infantile PD is rapidly progressive and fatal within the first year of life, secondary to cardiorespiratory failure without treatment. Late onset Pompe disease (LOPD) is primarily recognized as a proximal limb girdle muscle dystrophy with pulmonary involvement that leads to respiratory failure. There are many other presentations for LOPD including ptosis, WPW syndrome, basilar artery aneurysm, lingual weakness, rigid spine syndrome, polymyalgia rheumatic, and polymyositis amongst others. LOPD is often misdiagnosed with significant diagnostic delays. With the advent of ERT, and the knowledge that early treatment impacts a better clinical outcome, early recognition of the disease is critical. Case Presentation: We present a case of a 44-year-old male admitted to the ICU in acute respiratory failure leading to cardiac arrest. His past medical history included a diagnosis of 'idiopathic scoliosis syndrome' and low back pain from the age of 29. He reported difficulty keeping up with his lifestyle due to his muscle weakness, tiredness and hyper somnolence. There was a history of weight loss and muscle wasting in the past 18 months. His CPK, thyroid function testing, ANA, anti Ro, La, RNA polymerase, Troponin, Rheumatoid Factor, and myositis antibodies were negative. EMG showed findings consistent with an inflammatory myopathy. A muscle biopsy showed nonspecific myopathic findings and no glycogen accumulation. A diagnosis of polymyositis was made and the patient was started on steroids. During the recent hospitalization, given findings of slowly progressive adult onset myopathy with prominent involvement of axial, limb girdle, and respiratory muscles, a diagnosis of Pompe disease was suspected. This diagnosis was confirmed by low enzyme activity and two mutations in GAA gene, c. -32-13T>G and c.1827delC: (p. Tyr609X). Diagnostic delay was 16 years from the initial onset of symptoms and 2 years from onset of severe symptoms, similar to what has been reported in the literature. A review of the literature showed polymyositis as a common misdiagnosis for PD. We present this case to demonstrate the diagnostic odyssey often encountered in clinical practice, and to raise awareness about polymyositis as a common misdiagnosis of Pompe disease.

2388F

Emerging Next-generation Therapies for Pompe Disease. *R.J. Moreland, K.M. Taylor, A. McVie-Wylie, N.P. Clayton, B.M. Wentworth, R.K. Scheule, S.H. Cheng.* Rare Disease Science, Genzyme, Framingham, MA 01701.

Recombinant human acid α-glucosidase (rhGAA) is approved globally for the treatment of Pompe disease based on its ability to prolong invasive ventilator-free survival in infants, and improve or stabilize muscle and respiratory function in adults with the disease. The natural history, presentation, and progression of Pompe disease are extremely heterogeneous. Accordingly, the response to treatment is also variable and a minority of patients treated with rhGAA has shown marginal or no improvement. To address this, we are working on alternative therapeutic strategies that might be brought to bear in the management of this disease. These include the development of a second-generation enzyme (neoGAA) with improved potency and ability to target skeletal muscle. Preclinical studies with neoGAA in Pompe mice showed an approximately 5-fold enhancement in potency at clearing glycogen stores from a variety of tissues relative to unmodified GAA. Associated with this greater clearance in lysosomal glycogen are measurable improvements in motor function and coordination. Yet another approach that we are exploring is substrate reduction therapy that seeks to lower the rate of synthesis of tissue glycogen. As a first step, we have begun to validate potential drug targets that might support substrate reduction for Pompe disease.

2389T

Efficacy of long-term velaglucerase alfa on hematological and visceral parameters in treatment-naïve patients with type 1 Gaucher disease. *A. Zimran[1], I. Kisinovsky[2], E.A. Lukina[3], D. Elstein[1], D. Zahrieh[4], E. Crombez[4], P. Giraldo[5].* 1) Shaare Zedek Medical Center and Hebrew University-Hadassah Medical School, Jerusalem, Israel; 2) Your Health SA, Buenos Aires, Argentina; 3) Hematology Research Center, Moscow, Russia; 4) Shire, Lexington, MA, USA; 5) CIBERER and Hospital Universitario Miguel Servet, Zaragoza, Spain.

Purpose: We assessed the long-term efficacy of velaglucerase alfa in patients with type 1 Gaucher disease (GD1). Methods: Treatment-naïve GD1 patients aged ≥2 years received velaglucerase alfa (45 or 60 U/kg every other week [EOW]) in two Phase III trials and an extension study. Results: 39 GD1 patients were randomized to 45 or 60 U/kg EOW and enrolled in the extension study. Mean treatment duration was 54.5 (SD 9.9) months. At baseline: median age, 29 years (range 6, 62); n=8 <18 years; n=21 male; n=9 splenectomized; median hemoglobin concentration, 10.9 g/dL (range 7.1, 14.4); median platelet count, 77×10^9/L (range 13, 310); and median volumes of the spleen and liver, normalized to body weight (BW), 2.80% BW (range 0.96, 13.03) and 3.77% BW (range 1.90, 7.96) respectively. Longitudinal analysis showed mean (95% CI) changes from baseline to months 24, 36, and 60, respectively, for: hemoglobin (g/dL), (n=38) 2.68 (2.25, 3.10), (n=37) 2.68 (2.26, 3.11), and (n=10) 3.32 (2.72, 3.92); and platelets (10^9/L), (n=37) +100.3 (76.1, 124.6), (n=34) +102.5 (77.9, 127.1), and (n=8) +88.3 (54.3, 122.4). Longitudinal analysis showed mean (95% CI) changes from baseline to months 24, 39, and 63, respectively, for: normalized splenic volume, (n=30) -66% (-71%, -59%), (n=27) -71% (-76%, -66%), and (n=9) -78% (-83%, -73%); and normalized hepatic volume, (n=39) -27% (-31%, -22%), (n=36) -33% (-37%, -28%), and (n=9) -39% (-44%, -33%). No new safety concerns were identified. Conclusions: Improvements from baseline in hematological and visceral parameters were maintained during the extension study in GD1 patients receiving long-term velaglucerase alfa treatment.

2390F

Delivery of α-*N*-acetyl-glucosaminidase via choroid plexus-directed viral gene therapy as an enzyme replacement in cerebrospinal fluid for Sanfilippo B syndrome. *S-H. Kan[1], S.Q. Le[1], M.R. Haddad[2], E-Y. Choi[2], A. Donsante[2], S.G. Kaler[2], P.I. Dickson[1].* 1) Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 2) Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA.

Sanfilippo syndrome type B (Mucopolysaccharidosis IIIB, MPS IIIB) is a rare autosomal recessive lysosomal disorder caused by the deficiency of α-*N*-acetylglucosaminidase (NAGLU) which leads to accumulation of heparan sulfate glycosaminoglycans (GAG). The development of effective systemic ERT for MPS IIIB is hampered by inadequate mannose 6-phosphorylation and resultant diminished cellular uptake of recombinantly-produced human NAGLU. In addition, NAGLU must traverse the blood-brain barriers to reach the brain and affect the neurological phenotype.

We created a modified NAGLU fused with the receptor-binding motif of insulin-like growth factor-II (NAGLU-IGF-II). NAGLU-IGF-II shows enhanced ability to enter cells compared to untagged NAGLU, due to its ability to utilize the cation-independent M6P receptor, which is also a receptor for IGF-II. Recombinant adeno-associated virus serotype 5 (rAAV5) is selectively tropic for choroid plexus epithelia when administered into the cerebral ventricles. The choroid plexus epithelia are specialized cells that project into the brain ventricles, produce cerebrospinal fluid (CSF), and exhibit a slow rate of turnover. Transduction of these cells with rAAV5 would be expected to enable a protein encoded by the vector transgene to be secreted into CSF. To deliver NAGLU-IGF-II to the CSF, a rAAV5 construct containing the NAGLU-IGF-II cDNA was produced. Under stereotaxic guidance, a pilot *in vivo* study in MPS IIIB mice was performed using intracerebroventricular injection of 5×10^{10} rAAV5 viral particles. At 16 days post-AAV5 injection, robust NAGLU activity was detected (1.47-4.35 units/mg protein) in brain sections of affected mice, versus 0.00 in untreated affected mice, and 0.22 in heterozygotes ($p<0.05$ for each comparison). Immunohistochemistry showed expression of NAGLU-IGF-II in choroid plexus epithelia, as well as detectable enzyme in hippocampal and amygdala neurons from brain hemispheres both ipsilateral and contralateral to AAV5 injection. Evaluation of the effects of treatment on GAG storage is underway.

This therapeutic approach combines the advantages of choroid plexus-directed gene therapy with NAGLU-IGF-II might provide a permanent and effective treatment that delivers enzyme to CSF for the treatment of Sanfilippo B syndrome.

2391T

Humoral immune response against enzyme replacement therapy (ERT) alters enzyme distribution in Mucopolysaccharidosis I mice. *S.Q. Le, S-H. Kan, M. Vera, Y. Herskovitz, M. Srour, P.I. Dickson.* Dept Pediatrics, LA Biomed Harbor-UCLA, Torrance, CA.

Mucopolysaccharidosis type I (MPS I) is an inherited and progressive lysosomal storage disease which is caused by lack or low level of α-L-iduronidase (IDU), an important enzyme participating in glucosaminoglycan (GAG) metabolism. Enzyme replacement therapy with recombinant human IDU (rhIDU) is available, but a humoral immune response to rhIDU may limit its effectiveness.

We studied the tissue and cellular distribution of rhIDU in naïve and sensitized *Idua*$^{-/-}$ mice to determine the effects of the humoral immune response on rhIDU distribution. Weekly exposure to rhIDU was given to sensitize *Idua*$^{-/-}$ mice (n=13) via tail-vein injection from 4 to 16 weeks of age. Of the ten surviving sensitized mice (three died prior the completion of the injections), five showed anti-rhIDU IgG antibody titers ranging from 3 to 30 OD units/ml and were designated 'high-titer,' while five had titers <1 OD units/ml and were designated 'low-titer'. Incubation of MPS I fibroblasts with serum from high-titer mice led to ~40% reduction in rhIDU uptake compared with antibody-free serum. Sensitized, high-titer mice had an average ~33-45% lower enzyme activities in high reticuloendothelial (RE) content organs (liver, spleen, thymus) and ~33-70% lower enzyme distribution in low RE organs (lung, kidney, heart, brain) compared to naïve mice by the fluorescence labeled enzyme. Cellular rhIDU distribution in these organs was examined using immunostaining with antibodies against IDU. In high-titer mice, rhIDU was mainly localized in Kupffer cells in liver, while low-titer mice showed rhIDU in both Kupffer cells and hepatocytes. The humoral immune response against rhIDU may alter the distribution of the enzyme on a tissue and cellular level. This finding may have implications for human MPS I patients undergoing enzyme replacement therapy with rhIDU.

2392F

Lysosomal Disease Network. *C.B. Whitley[1], B. Diethelm-Okita[1], J.R. Utz[2], J.C. Cloyd[2], D.C.C. Erickson[1], E. Shapiro[3], Investigators of the Lysosomal Disease Network.* 1) Gene Therapy Ctr, Univ Minnesota, MMC 446, Minneapolis, MN; 2) Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN; 3) Neuropsychology, University of Minnesota, Minneapolis, MN.

The lysosomal diseases have become a test bed for some of the most innovative and advanced experimental treatments. Collaborative clinical research on these rare disorders and their treatment is absolutely crucial to make substantial progress. The Lysosomal Disease Network (LDN) brings together more than 500 researchers and clinicians across the country, Patient Advocacy Groups (PAG), and other interested partners, and has generated a synergistic research and educational consortium to advance treatment of these diseases. In NIH-funded multi-center program, longitudinal studies of the natural history of 11 lysosomal disease categories and 11 pilot studies of measurement of outcome and phase I/II clinical trials are focused on several themes. Central nervous system (CNS) disease has been the most difficult to treat as well as to measure. A significant focus will be on quantitative methods of CNS structure and function providing a standard toolbox across the network in the Mucopolysaccharidoses (MPS), Batten disease, Niemann-Pick type C, Mucolipidosis type IV, Late Infantile Neuronal Ceroid Lipofuscinosis, Glycoproteinoses, GM2-gangliosidoses, Sandhoff disease, and Wolman disease. A study on Pompe disease focuses primarily on the immune modulatory factors affecting treatment response. Additionally, we include a study on bone disease in the MPS and a set of innovative studies on Fabry disease in which collaborators will carry out the natural history of kidney structure and function, pulmonary function as a marker of disease progression in children, and identification of Fabry disease among high-risk populations. The LDN provides support for all of these projects, leveraging additional resources from PAG and industry, in the hope of fostering research on other lysosomal diseases and providing the impetus for more in-depth studies of pathophysiology and treatment. The LDN provides substantial support for at least two post-doctoral trainees each year for career development in lysosomal diseases as well as an international research meeting (The 10th annual **WORLD Symposium**, Feb 10-15, 2014, San Diego, CA, USA) for sharing of research findings, education, and network synergy. The website **www.LysosomalDiseaseNetwork.org** provides an educational, research, and clinical resource for the Network, patients, physicians, and the public. (Support by NINDS, NIDDK, ORDR-NCATS, U54NS065768).

2393T

The effect of citrulline and arginine supplementation on lactic acidemia in MELAS syndrome. *A. El-Hattab[1], L. Emrick[2], W. Williamson[2], W. Craigen[2], F. Scaglia[2].* 1) King Fahad Medical City, Riyadh, Saudi Arabia; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.

Introduction Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome is a common mitochondrial disorders in which energy depletion due to mitochondrial dysfunction can explain many of its the multi-organ manifestations. There has been growing evidence that nitric oxide (NO) deficiency occurs in MELAS and can play a major role in the pathogenesis of several complications including stroke-like episodes, myopathy, and lactic acidosis. Arginine and citrulline act as NO precursors and their administration can restore NO production in MELAS. Therefore, arginine and citrulline can be of therapeutic utility in treating NO deficiency-related manifestations in MELAS. Lactic acidemia is a cardinal manifestation of MELAS and results from an inability of dysfunctional mitochondria to generate sufficient ATP, leading to shunting of pyruvate to lactate. Moreover, hypoperfusion can result in lactic acidosis due to decreased oxygen delivery to peripheral tissues and shifting to anaerobic glycolysis. NO deficiency in MELAS can result in decreased blood perfusion and therefore aggravates lactic acidosis. In this study we hypothesized that arginine and citrulline supplementation will lower plasma lactate in patients with MELAS via increasing NO availability and improving perfusion. **Methods** We measured plasma lactate in 10 adults with MELAS before and after 48 hours of oral L-arginine supplementation. The study was subsequently repeated before and after L-citrulline. **Results** Average plasma lactate was lower after arginine (3.16→2.99 mmol/L) and citrulline supplementations (3.17→2.94 mmol/L). This reduction was statistically significant ($p<0.05$) and more consistent after citrulline supplementation. **Conclusions** The reduction in lactate after arginine and citrulline supplementation add more evidence to their potential therapeutic utility to restore the NO deficiency in MELAS. A previous study showed that both arginine and citrulline supplementations increase NO production in MELAS with citrulline resulting in a higher increment. In this study the lactate reduction was more significant and consistent after citrulline which can be due to the superiority of citrulline in increasing NO production leading to a better perfusion. These results also suggest that citrulline can have a better therapeutic effect. Additional assessments of the clinical effects of arginine or citrulline supplementation on different aspects of MELAS are needed.

2394F

Phenylbutyrate for therapy of pyruvate dehydrogenase complex deficiency. *R. Ferriero[1], A. Bourton[2], L. Bonafé[3], M. Baumgartner[4], D. Kerr[5], E. Morava[6], R. Rodenburg[6], M. Brivet[2], N. Brunetti-Pierri[1,7].* 1) Telethon Institute of Genetics and Medicine, Naples, Italy; 2) Laboratoire de Biochimie, AP-HP Hôpital de Bicêtre, Le Kremlin Bicêtre, 94270 France; 3) University of Lausanne, Lausanne, 1011 Switzerland; 4) Division for Metabolic Diseases, University Children's Hospital, Zürich 8032 Switzerland; 5) Center for Inherited Disorders of Energy Metabolism, Case Western Reserve University, Cleveland, OH 44106, USA; 6) Department of Pediatrics and the Institute of Genetic and Metabolic Disease (IGMD), Nijmegen Centre for Mitochondrial Disorders (NCMD), Radboud University Medical Centre, Nijmegen, The Netherlands; 7) Department of Translational Medicine, Federico II University of Naples, 80131 Italy.

Deficiency of nuclear-encoded pyruvate dehydrogenase complex (PDHC) is one of the most common inborn errors of mitochondrial energy metabolism. Most patients show progressive neurological degeneration and lactic acidosis. Affected patients present with a graded spectrum of severity ranging from the most severe Leigh syndrome with overwhelming lactic acidosis and death in the neonatal period to a milder form with carbohydrate induced ataxia. Most cases result from mutations in the X-linked gene for the E1α subunit (PDHA1) of the complex while fewer patients carry mutations in genes for E1β (PDHB), E2 (DLAT), E3 (DLD) and E3BP (PDHX) subunits or PDHC phosphatase (PDP1). PDHC activity is tightly regulated by phosphorylation and four isoforms of PDK (PDK1 to 4) phosphorylate the E1α subunit thereby reducing the activity of the complex. Conversely, dephosphorylation by phosphatases (PDP1 and PDP2) restores enzyme activity. We have previously shown that phenylbutyrate increases residual activity of PDHC by increasing the unphosphorylated enzyme and reduces lactic acidosis. Phenylbutyrate, and not its bio-product phenylacetate, increases PDHC activity through inhibition of E1α phosphorylation mediated by PDK1, PDK2, and PDK3 whereas PDK4 is not affected. We investigated phenylbutyrate on multiple cell lines harboring mutations in PDHA1 and genes encoding other components of the complex. Large deletions affecting PDHA1 and PDHX that result in lack of detectable protein were unresponsive to phenylbutyrate. PDHA1 mutations localized in the proximity of the catalytic site were responsive whereas mutations affecting the Arg349-α residue, involved in interaction with E1β subunit, were consistently unresponsive to the drug. PDHC activity was enhanced by phenylbutyrate in cells harboring PDHB, PDHX, and DLD missense mutations. Because phenylbutyrate is already approved for human use in other diseases, the findings of our studies have the potential to be rapidly translated for treatment of patients with PDHC deficiency and other forms of primary and secondary lactic acidosis. The results of the present study may help predicting whether a patient with PDHC deficiency is responsive to phenylbutyrate treatment based on the affected gene, type and location of the mutation.

2395T

A novel mutation in GLUD1 causing Hyperinsulinism Hyperammonemia in a patient with high density of homozygosity on microarray. *J.D. Odom[1], M. Gieron-Korthals[2], D. Shulman[2], P. Newkirk[2], E.J. Prijoles[3], A. Sanchez-Valle[2].* 1) Morsani College of Medicine, University of South Florida, Tampa, FL USA; 2) Department of Pediatrics, University of South Florida Morsani College of Medicine, Tampa, FL USA; 3) Clinical Division, Greenwood Genetic Center, Greenwood, South Carolina USA.

We describe a 28-month girl with congenital hyperinsulinism/hyperammonemia (HH) syndrome due to a de novo mutation (His507Tyr) in the glutamate dehydrogenase gene (GLUD1) not previously reported in the literature. She initially presented with intrauterine growth restriction, patent foramen ovale, peripheral pulmonic stenosis, and dysmorphic features. Her initial exam revealed wide anterior fontanel, bilateral posterior hair whorls, blue sclera, flat nasal bridge and posteriorly rotated low set ears. Chromosomal microarray analysis utilizing Affymetrix 6.0 platform revealed arr. (1-22, X) x2, with a high density of noncontiguous regions of homozygosity indicating possible consanguinity. Family history was unremarkable except that consanguinity was suspected as both parents were from a small village in Mexico. She had multiple short runs (1-10Mb) of allele homozygosity throughout the genome, consistent with limited outbreeding. She was suspected to have a recessive syndrome that had not yet been identified. At 8 months old, she presented to the emergency room with new onset seizures and hypoglycemia (39 mg/dL). Seizures were described as episodes of unresponsiveness and staring; body stiffness occurred with some episodes and cyanosis with most. She had frequent seizures until 18 months of age, most of which were triggered by hypoglycemia. Electroencephalogram showed encephalopathy with epileptic activity in the left centro-parietal region. Her hypoglycemic workup revealed excessive insulin level (6 µIU/mL at blood glucose of 44 mg/dL), negative urine ketones, low beta-hydroxybutyrate, and hyperammonemia (150-200 umol/L). DNA analysis revealed a novel mutation in GLUD1 that has not been previously described, predicted to be deleterious by SIFT and PolyPhen-2.1.0 programs. Parental DNA testing confirmed the mutation to be de novo, supporting the diagnosis of this autosomal dominant metabolic disorder. Her hypoglycemia initially responded well to 10 mg/kg/d diazoxide, however, she was subsequently admitted with recurrent seizures and gastroenteritis. Glycemic control was achieved with 15 mg/kg/d diazoxide and a leucine-restricted diet. At present, seizures are controlled with levetiracetam. She is thriving, but has global mild developmental delay that is attributed to her chronic hyperammonemia or an underlying unidentified genetic disorder whose locus is in one of the areas of homozygosity.

2396F

Pathogenic study of mitochondrial complex I deficiency and Leigh syndrome in Drosophila model. *Z. Li[1], K. Zhang[1], J. Guan[1, 4], C. Haueter[1, 2], H. Sandoval[1], H. Bellen[1, 3], B. Graham[1].* 1) Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX 77030; 2) Methodist Hospital, Houston, TX 77030; 3) Howard Hughes Medical Institute, Houston TX 77030; 4) Medical Student, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030.

Mitochondrial disorders, including oxidative phosphorylation (OXPHOS) disorders, affect at least 1 in 5,000 individuals. Isolated complex I deficiency is the most frequently observed OXPHOS defect (64%) and is a common cause of Leigh Syndrome with no effective treatments available. NDUFS3 is a Fe-S cluster-containing subunit of Complex I in electron transport chain. Oxidative stress is an important pathogenic factor in OXPHOS deficiencies as well as in diseases with secondary defects of mitochondrial metabolism. We have generated mutant NDUSF3 alleles in Drosophila through P-element excision of CG12079, the Drosophila ortholog of NDUFS3. Phenotypic characterizations include molecular, biochemical,immunohistochemical, electrophysiological and behavioral analyses as well as treatment with antioxidants. Recessive mutations of NDUFS3 cause severe deficiency of Complex I activity with abnormal assembly. Most animals without NDUFS3 fail to pupate and die during larvogenesis, however, rare adult escapers survive and experience a shortened lifespan. NDUFS3-deficient flies exhibit increased levels of oxidative stress demonstrated by fluorescence markers and decreased mitochondrial aconitase activity. Evidence for mitochondrial proliferation includes dramatically increased citrate synthase activity and increased Spargel gene expression (Drosophila ortholog of PGC-1 α). Mutant flies also exhibit neurological dysfunction manifested as increased sensitivity to mechanical stress, abnormal morphology of larval neuromuscular junction, defective climbing ability and perturbed retinal function. The adult mutant NDUFS3 phenotypes are fully rescued by ectopic genomic rescue of NDUFS3 and partially rescued by vitamin E analogs supplemented in food. These results demonstrate that the loss of NDUFS3 in Drosophila causes Complex I deficiency and increased oxidative stress. These studies also demonstrate that Drosophila is a useful model system for investigating the pathogenesis of mitochondrial disease as well as for exploring therapeutic strategies.

2397T

Long term follow up of 15 patients with methylmalonic acidemia following solid organ transplantation. *J.L. Sloan, I. Manoli, C.P. Venditti.* NIH/NHGRI/GMBB, Bethesda, MD.

Methylmalonic acidemia (MMA) is a devastating disorder, recalcitrant to medical management in many cases. Solid organ transplantation, including liver (LT), combined liver-kidney (LKT) or kidney transplants (KT), has been reported in >40 patients with isolated MMA. However, there is little systematic information on long-term outcomes and there are currently no guidelines regarding indications for transplant in MMA. We describe the natural history up to 15.4y post-transplant in 15 individuals with MMA (13 mut, 1 cblA, 1 cblB). There were 3 LT, 8 LKT and 4 KT recipients in our cohort . The mean age at transplant was 14.3 yrs (range 1.3-31.4 yrs) and mean follow up 5.1 yrs (range 0.8-15.4 yrs). All 11 patients status post LT or LKT remained metabolically stable (65.1 yrs combined follow-up), with the exception of one LKT patient, who had numerous hospitalizations due to gastrointestinal dysmotility. A previously reported mut0 patient who developed renal failure requiring KT 1-year post-LT and a worsening movement disorder after liberalizing her diet. Another mut0 patient with KT suffered repeated hospitalizations for metabolic decompensations and developed optic nerve atrophy 3 yrs after transplant. Transplantation improved metabolic parameters (72-93% decrease), primarily due to correcting renal function in those who received KT, but all had persistently elevated methylmalonic acid in plasma (71-1383 μM, nl <0.4), urine (166-4446 mmol/mol cr, nl<3) and total whole body MMA output (23-382 μmol/kg/day). Despite the massive MMA elevations, two LKT patients >12 yrs post transplant had eGFR >80 ml/min/1.73m2. Studies in this patient group, the largest assembled to date, afford the following conclusions: 1. LT and LKT, but not isolated KT, completely prevent ketoacidotic crises. 2. After LT, patients can develop renal failure and therefore require careful monitoring and this is also a theoretical risk for LKT and KT patients. 3. Although solid organ transplantation improves biochemical parameters, adequate metabolic control is required following the procedure due to the risk for late complications of the disease. Further studies are needed on the association of genotype and metabolic correlations pre- and post-procedure and the outcomes achieved, to develop guidelines about the optimal timing and procedure indicated for each patient.

2398F

Neuroimaging and neuropathology reveal dysmyelination in canine mucopolysaccharidosis I. *P. Dickson[1], J. Provenzale[2,3], S. Chen[2], I. Nestrasil[4], S-h. Kan[1], S.Q. Le[1], E. Lotshaw[1], J. Jens[5], J. Yee[1], N.M. Ellinwood[5], M.A. Guzman[6], C. Vite[7], V. Kovac[4], E.G. Shapiro[4].* 1) Dept Pediatrics, LA BioMed Harbor-UCLA Med Ctr, Torrance, CA; 2) Duke University Department of Radiology, Durham, NC; 3) Emory University Department of Radiology, Oncology and Biomedical Engineering, Atlanta, GA; 4) University of Minnesota Department of Pediatrics, Minneapolis, MN; 5) Iowa State University Department of Animal Science, Ames, IA; 6) St. Louis University Department of Pathology, St. Louis, MO; 7) University of Pennsylvania, School of Veterinary Medicine, Philadelphia, PA.

The cause of neurological deterioration due to mucopolysaccharidosis (MPS) is poorly understood. While many investigations focus on neurodegeneration, evidence suggests that white matter involvement may also be important. Hyperintense white matter lesions on brain imaging studies correlate with cognitive impairment in patients (1), and MPS patients show reduced fractional anisotropy (FA, a measure of white matter integrity) on diffusion tensor imaging of the corpus callosum that correlates with specific measures of attention (2). We performed neuroimaging and neuropathological evaluation of the corpus callosum (a major white matter structure) in canine MPS I. We studied nine MPS I dogs, four unaffected carriers, and four MPS I dogs treated with intrathecal recombinant alpha-L-iduronidase (IT rhIDU) 0.05 mg/kg every 3m from 4m to 21m. Dogs were imaged at end-study (mean age 24m for carrier, 21m for IT-treated, and 20m for untreated MPS I). Mean whole corpus callosum volumes were lower in MPS I dogs vs. carriers (p=0.02). Corpus callosum volume was preserved with IT rhIDU (p=0.006 vs. untreated MPS I). Diffusion tensor imaging demonstrated that fractional anisotropy was reduced in MPS I dogs vs. carriers (p=0.004 for genu and 0.008 for splenium of the corpus callosum), and the component of fractional anisotropy that was most abnormal was an increase in radial diffusivity (consistent with demyelination). The IT rhIDU-treated dogs showed fractional anisotropy values intermediate between carriers and MPS I dogs. Ultrastructural evaluation of the corpus callosum revealed thinning and disruption of the myelin sheath in MPS I dogs. Myelin-related genes showed small but consistent reduction in the expression by RT-PCR showed a small but consistent reduction in the expression of myelin-related genes in MPS I dogs vs. controls. Western blots showed reduced myelin basic protein and phosphorylated myelin basic protein in the corpus callosum of MPS I dogs. We quantified LC/MS showed abnormalities of myelin lipids by LC/MS, including a 37% reduction in sphingomyelin in MPS I dogs (p<0.001). Abnormal diffusion tensor imaging of the corpus callosum, which correlates with attention loss in MPS I patients, appears to be due to dysmyelination in MPS I dogs. Further studies, including determining the mechanism of these abnormalities, are underway. (1)Vedolin L, et al. Am J Neurorad 28 (2007) 1029-33. (2)Shapiro E, et al. Mol Genet Metab 107 (2012) 116-121.

2399T

Neuropsychiatric Outcomes in PKU Patients Treated With Sapropterin: Results from the Randomized, Controlled PKU ASCEND (PKU 016) Trial. *S. Prasad[1], B. Burton[2], A. Feigenbaum[3], M. Grant[4], R. Hendren[5], R. Singh[6], S. Stahl[3], C. Zhang[1].* 1) BioMarin, Novato, CA; 2) Children's Memorial Hospital of Chicago, Chicago, USA; 3) University California San Diego, San Diego, USA; 4) St. Christopher's Hospital for Children, Philadelphia, USA; 5) University California San Francisco, San Francisco, USA; 6) Emory University, Decatur, USA.

Background: Phenylketonuria (PKU) patients often exhibit attention deficits similar to attention deficit hyperactivity disorder (ADHD). Methods: PKU 016 is the largest health outcomes study in PKU. This randomized, placebo-controlled trial evaluated baseline neuropsychiatric impairment, specifically ADHD behaviors, symptoms of anxiety (HAM-A rating scale), depression (HAM-D rating scale) and executive dysfunction (BRIEF rating scale), and the effects of sapropterin dihydrochloride (Kuvan) on these impairments after 13 weeks of treatment in 206 children and adults with PKU. The primary endpoint was the total score on the ADHD Rating Scale (ADHD-RS) commonly used to evaluate symptoms of inattentiveness and hyperactivity. Results: Mean patient age was 22.5±11.6 years, with 42% aged <18 years. Baseline blood Phe was 841.4±473.1 μmol/L. 35% of the sample had symptoms of ADHD at baseline. 118/206 were sapropterin responders (defined as blood Phe level reduction ≥20%). Among sapropterin responders, ADHD-RS total score was improved in the sapropterin group compared with placebo (mean change from baseline -9.1±2.2 vs -4.9±2.0, P=0.085), driven by a statistically significant and clinically relevant change in the inattention subscale (-5.9±1.4 vs -2.5±1.3, P=0.036). Conclusions: Sapropterin was associated with improvement in inattention among children and adults with PKU who had a Phe response to sapropterin therapy.

2400F

Effect of Globotriaosylceramide (Gb3) on cytokine profile in dendritic cells and monocytes. Consequences for Fabry disease. *P. Rozenfeld, P. De Francesco, J. Mucci, R. Ceci.* Dept Immunolgia, Univ Natl de La Plata, Buenos Aires, Argentina.

Fabry disease (FD) is an X-linked genetic disorder characterized by the deficiency in the activity of the lysosomal enzyme α-galactosidase A (αGal). This defect leads to the accumulation of neutral glycolipids, mainly globotriaosylceramide (Gb3). Clinical manifestations in males include angiokeratoma, hypohidrosis, cornea verticillata, proteinuria, acroparesthesia. Death usually occurs at 4th decade of life because of kidney, cardiac or cerebrovascular problems. Previous results of our group showed increased expression and production of proinflammatory cytokines in peripheral blood mononuclear cells (PBMC) from patients with EF, in particular dendritic cells (DC) and monocytes (M).The aim of this study is to investigate the effects of Gb3 on the production of cytokines in DC and macrophages (MΦ) derived from normal PBMC monocytes treated with DGJ, an inhibitor of αGal.PBMC were isolated from normal buffy coats, and M were purified and cultured in the presence of GM-CSF and IL-4, or M-CSF, to induce their differentiation into DC and MΦ, respectively. Obtained cells were then cultured in the presence or absence of 20μM Gb3 and/or 200μM DGJ and the levels of IL-β, IL-6 and TNFα were analyzed in the supernatants. Both cultures showed a significant increase in production of IL-1β and TNFα, relative to the control, only in the case of combined Gb3 and DGJ treatment (DC p=0.0018 and p=0.0041, MΦ p=0.0002 and p=0.0054, respectively). A similar trend was observed for IL-6. Moreover, addition of a TLR4-blocking antibody completely abolished the observed effect in MΦ treated with Gb3 and DGJ.These results show that Gb3, possibly by interacting with TLR4, can induce a proinflammatory state, similar to that observed in PBMC from patients with FD, and could be directly involved in the pathogenesis of this disease.

2401T

Is common mutation in SCO2 associated with reproductive failure in the Polish population? Preliminary study. *S. Łuczak[1], D. Piekutowska-Abramczuk[1], A. Jezela-Stanek[1], K. Chrzanowska[1], M. Kugaudo[1, 2], A. Cieślikowska[1], A. Kochański[1], P. Kowalski[1], E. Ciara[1], J. Trubicka[1], D. Jurkiewicz[1], M. Pelc[1], M. Borucka-Mankiewicz[1], M. Krajewska-Walasek[1], E. Pronicka[1].* 1) Medical Genetics, The Children's Memorial Health Institute, Warsaw, Poland; 2) Department of Child and Adolescent Psychiatry, Medical University of Warsaw, Warsaw, Poland.

Mutations in the SCO2 gene (22q13) lead to severe COX deficiency observed mainly in muscles, heart and brain. A common, g.1541G>A (p.E140K), substitution is identified at least on one allele (in heterozygous or homozygous form) in all reported patients. Analysis of genotype-phenotype correlation revealed the presence of two distinct clinical patterns of disease: phenotype of early beginning (at birth or in fetus) with hypertrophic cardiomyopathy as a leading symptom, and neurological phenotype of delayed begining with encephalomyopathy. There are a few data of the possible association between mutation in SCO2 and fetal wastage, especially in fetuses with a heterozygous form of common mutation. The aim of this preliminary study was to attempt to verify the existence of probable association SCO2 mutations with early spontaneous abortions. We examined 64 DNA samples from couples with at least 3 miscarriages in family history (44 samples of whole blood on EDTA and 20 dry blood spots on Guthrie cards). Genomic DNA was eluted from dried blood spots using 5% Chelex-100 resin and extracted from peripheral blood leukocytes by phenol/chloroform procedure or by automatic isolation (MagNa Pure equipment). Genotyping for the presence of the specific mutation was performed by Real Time PCR analysis with Taqman probes (DNA concentration about 30 ng/µl). The probes were designed to 150 nt fragment of the SCO2 gene (position in the reference sequence in the NCBI database: gi|281182726:495-644). The study did not reveal the presence of common mutation in the examined group. Every patient was homozygous for the wild form. Since this is a preliminary study and the study group has less than 100 persons, the existence of correlation between reproductive failure and SCO2 mutations can not be excluded. Carriership of common mutations in the Polish population is relatively high (1:147), so the planned extension of the study group with early abortions seems to be fully justified. Based on reported data in all cases with positive family history (recurrent abortions, fetal wastage, neonatal deaths characterized by cardiomyopathy, myopathy or encephalopathy), screening for the g.1541G>A mutation should be performed. The study was financed by National Science Centre, project no. 1154/B/P01/2011/40, EU Structural Funds, project POIG.02.01.00-14-059/09 and CMHI project no. S126/12.

2402F

Frequency of inherited metabolic diseases (IMDs) in high-risk children in a North Indian tertiary care hospital. *S. Attri[1], A. Patial[1], P. Kumar[1], P. Singhi[1], S. Singhi[1], S. Sharda[1], C. Kumari[2], I. Dwivedi[1], S. Kapoor[2], L. Kratz[3].* 1) Pediatrics, Postgraduate Inst Medl Education & Research, Chandigarh, India; 2) Division of Genetics, Lok Nayak Hospital & Maulana Azad Medical College, New Delhi, India; 3) Biochemical Genetics Laboratory, KKI, Baltimore, USA.

Introduction: Newborn screening, not commonplace in our country necessitates the shift to high risk screening. We present our data from a cohort in North India. Aims and objectives: To study the frequency of various IMDs in high-risk children. Methods: A total of 220 suspected cases with red flag signs suggestive of a metabolic disorder were enrolled from March 2012 till January, 2013 from various wards of the Advanced Pediatric Centre like pediatric intensive care unit, neonatal intensive care unit, emergency department and out-patient neurodevelopment clinic. Children from age group 0-15 years were included. These patients were subjected to initial blood and urine screening tests followed by thin-layer chromatography, tandem mass spectrometry for dried blood spots, gas chromatography mass spectrometry of urine samples and high performance liquid chromatography on plasma and urine samples. Enzyme estimations were done for biotinidase and prolidase deficiencies. Results: Majority of the children (112/220) were in the age group of 1-15 years. Male to female ratio was 1.5:1. We were able to diagnose 13% children with IMDs. Eighteen children had hyperhomocysteinemia and/or homocystinuria, out of which seven had combined homocystinuria and methylmalonic aciduria. Other disorders detected were glutaric aciduria (2), fumaric aciduria (2), biotinidase deficiency (2), congenital lactic acidosis (1), primary hyperoxaluria (1), maple syrup urine disease (1), prolidase deficiency (1) and urea cycle defect (1). Conclusions: Amongst the investigated children, 13% were diagnosed with IMDs. The most common IMDs were homocystinuria and cobalamin defect. Implementation of a national newborn screening is of paramount importance.

2403T

Different Combinations of *ACADS* Mutations Cause SCAD Deficiency in an Infant and His Mother. *A.N. Dang Do[1], J. Cox[1], M. Johnson[1], L. Kratz[2], T. Wang[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Biochemical Genetics Laboratory, Kennedy Krieger Institute, Baltimore, MD.

Short-chain acyl-coenzyme A dehydrogenase (SCAD) deficiency is an autosomal recessive disorder of fatty acid oxidation caused by short-chain acyl-CoA dehydrogenase (*ACADS*) deficiency. More than 35 disease-causing mutations and two common variants (R171W and G209S) have been identified in *ACADS*, which encodes an enzyme responsible for the dehydrogenation of butyryl-CoA during the first steps of short-chain fatty acid beta-oxidation. Clinical presentations of patients with confirmed SCAD deficiency vary considerably from asymptomatic individuals to patients with poor feeding, hypoglycemia, hypotonia, developmental delays, and epilepsy. Since SCADD is an autosomal recessive disorder, parents of affected infants identified through newborn screening are rarely investigated. We report a case of metabolically and molecularly confirmed SCAD deficiency in an asymptomatic newborn male infant and his mother who had subtle symptoms of this disorder during periods of metabolic stress. The 6-1/2 week old infant was found to have elevated C4 and C4/C2 ratio by state newborn screening, and the suspected diagnosis of SCADD was confirmed by finding increased plasma butyrylcarnitine (C4: 0.82 nmol/ml; nl <0.62) and urinary ethylmalonate (EMA: 60 mg/g creatinine; normal < 30). Sequence analysis of *ACADS* in this infant identified three coding variants: one reported mutation, R107C; one apparent truncated variant, c.660-699del40; and one common variant, G209S. Careful review of the family history revealed that his mother had a history of fasting-related fainting episodes, which were relieved by taking snacks or sugar-containing drinks. Metabolic studies of the mother revealed an increased plasma butyrylcarnitine level (C4: 1.92 nmol/ml; normal <0.62) and increased urinary excretion of ethylmalonate (EMA: 122 mg/g creatinine; normal <15 for adult). Sequence analysis of *ACADS* in the parents showed that the mother carried two apparent deleterious mutations (R107C and c.660-699del40) in *cis* and a presumed functional variant, G281S, in *trans*; while the father was a carrier for a common variant, G209S. This case exemplifies the need for careful evaluation of parents of confirmed SCADD infants to identify subtle clinical symptoms suggestive of this recessive metabolic disorder.

2404F

A formal approach to evaluating the neuropsychiatric manifestations of PKU: Assessing the content validity of ADHD rating scales in phenylketonuria. *K. Mooney[1], S. Prasad[1], S. Kummer Shaffer[2], K.S. Gries[2], P. Auguste[2], D. Bilder[3], K.W. Wyrwich[2].* 1) BioMarin, Novato, CA; 2) United BioSource Corporation, Bethesda, MD; 3) University of Utah, School of Medicine, Salt Lake City, UT.

Background: Adults and children with phenylketonuria (PKU) demonstrate symptoms of Attention-Deficit/Hyperactivity Disorder (ADHD). Both the self-reported Adult ADHD Self-Rating Scale (ASRS) and parent-completed ADHD Rating Scale (ADHD RS-IV), based on DSM IV criteria, have been validated and successfully measure changes over time in features of inattentiveness and hyperactivity/impulsivity in ADHD. However, these instruments have not been validated for their appropriateness to measure these concepts in the PKU population. The objectives of this study were to: 1) elicit descriptions of participant-perceived PKU symptoms, and 2) assess the content validity of the ASRS and ADHD RS-IV in measuring inattentive and hyperactive/impulsive symptoms in adults and children with PKU. Methods: Following IRB approval, adults with PKU (n=13) and parents of children with PKU ages 8-17 (n=15) were recruited for one-on-one interviews exploring their or their children's PKU symptom experiences. At the end of each interview, adults completed the ASRS, parents completed the ADHD RS-IV (reporting their children's behaviors), and all participants provided their insights on the clarity, comprehensiveness, and relevance of the respective instrument in measuring PKU-related symptoms. Results: Participants reported inattention-related symptoms much more frequently than hyperactive/impulsive symptoms. The emergent participant-reported inattentiveness symptoms matched well with respective questions in the ASRS and ADHD RS-IV. Most participants reported that the ASRS and ADHD RS-IV questions were clear, meaningful, and relevant to their own, or their children's experience with inattentiveness, but they were less supportive of the hyperactive/impulsive subscale. Conclusions: The qualitative interviews demonstrated concept elicitation support for concepts related to inattentiveness. The ASRS and ADHD-RS IV demonstrated the clinical relevance of these instruments for assessing relevant inattentive symptoms associated with PKU in persons ≥8 years old. Additional research is ongoing to understand the psychometric properties of these ADHD instruments in the PKU population. Furthermore, exploration of psychiatric aspects of PKU and other metabolic diseases using valid measurement instruments yields important insights into the clinical manifestations of such disease.

2405T

The role of innate immune system activation and signaling in vascular disease in the MPS I canine model. *M. Vera[1], S.Q. Le[1], S. Kan[1], P. Dickson[1], R. Wang[2].* 1) Dept of Pediatrics, Division of Medical Genetics, Harbor-UCLA Medical Center, Torrance, CA; 2) CHOC Children's, Orange, CA.

Vascular disease in mucopolysaccharidosis I consists of regions of arterial intimal-medial thickening characterized by glycosaminoglycan storage, vascular smooth muscle proliferation, and elastin fiber disruption. The latter features are suggestive of innate immune system signaling pathways including those mediated by the Toll-like receptor family. These intraluminal lesions do not completely respond to enzyme replacement therapy or bone marrow transplantation, therefore understanding the initiating events leading to this phenotype is likely to be important in designing effective treatment. Our goal in this project is to identify and quantify the relationship between glycosaminoglycan storage and innate immune system mediated inflammatory signaling in MPS I vascular disease. Our hypothesis is that innate immune system stimulation leads to macrophage activation and production of cytokines, including TGFβ, that stimulate extracellular matrix remodeling and vascular smooth muscle cell proliferation. Our preliminary data indicate an increased expression of TGFβ and increased activation of downstream signaling mediators within these arterial lesions in MPS I canine arterial disease. We are continuing to study the causal relation between glycosaminoglycan storage and Toll-like receptor signaling in the development of this vascular phenotype.

2406F

Defective T-cell function in a mouse model of Citrullinemia Type I. *T.N. Tarasenko, J. Gomez-Rodriguez, P.J. McGuire.* National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Although inborn errors of metabolism (IEM) are often exacerbated by infection, whether having an IEM impacts immune function is less well characterized. Lymphocyte activation is accompanied by dramatic changes in metabolism suggesting that certain IEM may also affect immune cell function. Argininosuccinate synthetase (ASS) is present in multiple tissues and is known to play a role in the conversion of citrulline to arginine, and nitric oxide generation. ASS deficiency in humans results in the urea cycle disorder known as known as citrullinemia type I, and is characterized by grossly elevated plasma citrulline levels and hyperammonemia. In human tissues, we demonstrated ASS protein in components of the immune system including naïve T-cells, thymus and spleen, however, the role of ASS in these normal tissues remains to be defined. Based on the expression of ASS, we hypothesized that ASS deficiency will result in T-cell defects. To evaluate this question, we are characterizing immune function in hypomorphic $ASS1^{fold/fold}$ (*fold/fold*) mice. As previously published, the lifespan of *fold/fold* mice is reduced at 3 weeks. Interestingly, studies of immune organs revealed significantly smaller thymi and spleens, absent mesenteric lymph nodes, and a marked reduction in T-cell numbers despite normal expression of lineage and activation surface markers. Adeno-associated virus 8-mediated correction of liver ASS to enhance *fold/fold* survival resulted in a greater recovery of splenic T-cell numbers despite the persistence of the ASS enzyme defect in splenic T-cells. However, in vitro studies in *fold/fold* T-cells demonstrated abnormal differentiation into various T-cell subsets critical for immune responses including Th1, Th2, Th17 and Treg cells. In vivo characterization of *fold/fold* immune cells is currently underway using bone marrow chimeras with immunization and influenza infection protocols. In addition, a trans vivo approach using humanized mouse models is being utilized to explore the role of human ASS in immune organs and other immunologic niches. Finally, to further understand potential immune consequences of this IEM, clinical characterization of immune status in patients with ASS deficiency and related disorders is currently underway in the NIH Clinical Center. Overall, our work suggests that ASS deficiency may translate into specific immune defects and may have implications for patients with citrullinemia Type I.

2407W

Local and International Medical Geneticists and Genetic Counselors Collaboration: The First Vietnamese-North American Genetics Conference in Hanoi, Viet Nam. *K. Leppig[1], M. Laurino[2], D. Sternen[3], J. Thompson[4].* 1) Genetic Services, Group Health Cooperative, Seattle, WA; 2) Institute for Public Health Genetics, University of Washington, Seattle, WA; 3) Laboratory Medicine, Seattle Children's Hospital, Seattle, WA; 4) Molecular Genetics Laboratory, BC Children's and Women's Hospital, Vancouver, BC.

The commitment of local Vietnamese medical geneticists Dr. Dung Chi Vu at the National Hospital of Pediatrics and Dr. Nguyen Viet Nhan at Hue College of Medicine and Pharmacy served as the strong foundation for collaboration with international genetic professionals. The shared goals of the local and international partnership were to bring together genetic providers from across Viet Nam to discuss the scope of current genetics care provided in the country; facilitate collaboration and mentorship with international colleagues; discuss examples of new technologies that may allow more cost effective genetic testing; and highlight the field of genetic counseling as part of health care delivery. A longer term goal was to develop a sustainable genetic counseling training program within Viet Nam to anticipate the need for specialized professionals who can work with the increasing number of geneticists as newborn screening and medical genetics services are rapidly growing. To achieve these goals, the first Vietnamese-North American Genetics Conference was held in Hanoi, Viet Nam on March 25-26, 2013. Targeted attendees of the meeting were health care providers in Viet Nam who are involved in the care of children with genetic conditions and birth defects. Over 300 physicians, laboratory specialists, nurses, researchers, medical students, international non-governmental organization workers, and government officials attended. There were over 20 presentations by Vietnamese experts and internationally recognized North American geneticists. A genetic counseling workshop discussed the history of genetic counseling, roles of genetic counselors, establishment of the Philippines genetic counseling program as a model for strategic curriculum development, and the ethical, legal and sociocultural issues of genomic medicine. Conference attendees confirmed joint interest in the creation of a genetic counseling program within Viet Nam. As such, the 2nd Vietnamese-North American Genetics Conference is being planned in conjunction with the 2014 Asia Pacific Conference on Human Genetics in order to foster the ongoing local and international partnership of genetic professionals. Certainly, this year's conference offered transparency and recognition of the current needs and invaluable opportunities on how to best integrate the field of medical genetics and genetic counseling in Viet Nam's existing health care and educational system.

2408W

Making sense of a primary care role in genetics: Views of genetics professionals. *J.C. Carroll[1], S. Morrison[2], F.A. Miller[3], B.J. Wilson[4], J.A. Permaul[5], J. Allanson[2,6].* 1) Dept of Family & Community Medicine, Mount Sinai Hospital, University of Toronto, Canada; 2) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 3) Institute of Health Policy, Management and Evaluation, University of Toronto, Canada; 4) Department of Epidemiology and Community Medicine, University of Ottawa, Canada; 5) Department of Family Medicine, Mount Sinai Hospital, Toronto, ON, Canada; 6) Department of Pediatrics, University of Ottawa, Canada.

Background: Developments in genetics are widely expected to implicate primary care providers (PCPs), who, as first contact professionals with a foundational role in most health care systems, may need to be better equipped. Yet public policy efforts to enhance the relationship between primary care (PC) and genetics vary widely. Initiatives in some jurisdictions such as the UK seek to `mainstream' genetics knowledge and services, while in others such as Canada, top-down initiatives have not been pursued. Objectives: To explore the role of genetics health professionals (GHPs) in building relationships with the PC community. Method: Qualitative methodology. GHPs working at genetics centres in Ontario, Canada were invited to participate in semi-structured focus groups or telephone interviews. Purposeful sampling was used to ensure diversity by type of GHP, and type of centre (academic, community, northern). Transcripts were analyzed using a qualitative interpretive approach. Findings: 5 focus groups and 2 interviews were conducted with GHPs (6 clinical geneticists, 24 genetic counselors, 4 nurses, 4 lab staff, 3 administrators) from 5 of 10 regional genetics centres and 2 of 3 northern outreach clinics across Ontario. Through interactions with their patients, very limited direct interactions with PCPs, and occasional educational events that involved PCPs, GHPs sought to make sense of the nature of PCP practice, and to imagine a PCP role in genetic medicine. GHPs acknowledged their limited understanding of the PCP role (and the role of different PCPs) in health care, and had little awareness of how genetics might fit into PCPs' practice. GHPs saw a role for PCPs in genetics, but expressed concern at a perceived lack of knowledge of genetics, as reflected in both under and over referring, and PCPs' ability to incorporate new genomic discoveries into practice. GHPs also reflected on whether existing communication methods with PCPs were effective, and recognized a need to understand what PCPs might want from the genetics community. Conclusions: Widespread expectations of a PCP role in genetic medicine are shared by GHPs in Ontario, Canada. But in the absence of policy direction or organizational support, GHPs struggle to make sense of how this might work. In this context, in addition to identifying knowledge and capacity deficits, parameters of a productive collaboration between PCPs and GHPs need to be developed.

2409W

A Brief Curriculum for Physician Orientation to Clinical Whole Genome Sequencing. *M.A. Giovanni[1,2], J. Krier[2], J.L. Vassy[2,3], D. Lautenbach[3], R.C. Green[2,3], M.F. Murray[1] for The MedSeq Project.* 1) Geisinger Health System, Danville, PA; 2) Harvard Medical School, Boston, MA; 3) Division of Genetics, Brigham and Women's Hospital, Boston, MA.

The MedSeq Project is a randomized clinical trial examining the impact of whole genome sequencing (WGS) in clinical care. In this study, 200 patients will receive WGS or usual care in the context of either primary or specialty care. Physicians are engaged by the study as research subjects and asked to recruit patients from their practice. This report describes a brief educational curriculum developed and administered to physicians participating in the research in order to prepare them to receive clinical reports from WGS. All of the physicians engaged in the educational program have appointments at Brigham and Women's Hospital in Boston. So far, we have enrolled 10 primary care providers trained in internal medicine (4 men and 6 women) and 7 specialists trained in both internal medicine and cardiology (5 men and 2 women). A 6.5 hour continuing medical education (CME) course was designed and offered to participating physicians prior to patient engagement. The curriculum included two 75-minute in-person didactic sessions, and 12 self-study case-based educational modules, allowing multiple modalities to facilitate engagement of busy physicians. The case studies were designed to highlight key concepts in genomic medicine. In addition, a library of resources was created and made available online to supplement the course lectures and self study cases. Pre- and post -education surveys were conducted to measure the impact of the educational program. Additional data on the confidence and performance of physicians following the patient disclosures are being tracked. As genomic sequencing is integrated into the practice of medicine, motivated clinicians will seek opportunities to improve their knowledge and understanding of applied genomics prior to delivering results to their patients. A short curriculum of didactic sessions and self-study, coupled with relevant learning resources, has been created to enhance physician preparedness for delivering genomic results to their patients and data will be presented evaluating its effectiveness.

2410W

Teaching Physicians about Genomic Medicine. *I. Maya[1], I. Basel-Vanagaite[1,2,3,4], E. Basel-Vanagaite[1], A. Koifman[5,6], DM. Behar[7], R. Tomashov-Matar[1], R. Sukenik-Halevi[8], A. Reches[9], M. Shohat[1,2].* 1) medical genetics, Recanati Genetic Institute, Petah Tikva, Israel; 2) The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Pediatric Genetics, Schneider Children s Medical Center of Israel, Petah Tikva, Israel; 4) Felsenstein Medical Research Center, Rabin Medical Center, Petah Tikva, Israel; 5) Institute of Human Genetics, Soroka University Medical Center, Beersheba, Israel; 6) Ben-Gurion University of the Negev, Beersheba, Israel; 7) Rambam Medical Center, Haifa, Israel; 8) Meir Medical Center, Kfar Saba, Israel; 9) Souraski Medical Center, Tel Aviv, Israel.

Background: Due to new discoveries in genomic medicine, clinical guidelines are increasingly recommending the incorporation of genomic tests or therapeutics into routine care, especially as the transition of genetic knowledge from research laboratories into clinical practice is becoming more and more a part of health care systems. Primary care practitioners suffer from inadequate knowledge and skills in medical genetics and many are unaware of the technical, ethical, legal and psychosocial implications of genetic testing. Methods: We initiated a “genomic education” program for the purpose of teaching primary care practitioners new advanced knowledge on genomic medicine. We emphasized the main take-home messages for physicians, which were: risk calculation for various genetic diseases, recognition of the mode of inheritance from the pedigree, guidelines for decision-making on which molecular tests to use, and the interpretation of test results and their clinical implications. Results: To date, 48 physicians have participated in 3 “genomic education” program, which included lectures, workshops and guided tours in genetic laboratories. In the “pre-course” examination the average score was 57% (range 20%-80%), whereas in the “post-course” examination it was 77% (range 74%-100%). The average improvement in score as a result of the course was 21% (range 0%-80%). The physicians who participated in our program reported a very high level of satisfaction from the content as well as the concept of a “one-week update”. Conclusion: A one-week “genomic education” program is an effective strategy to update primary care physicians in order to improve their care of patients.

2411W

Hello Genetics (Alô Genética): A Brazilian strategy to provide guidance and education in genetics to primary health care providers. *T. Vieira[1,2], D. Nadler[1], L. Silva[1], N. Murcia[1], C. Giugliani[3], C. Rafaelli[1], R. Giugliani[1,4,5].* 1) Medical Genetics Service, Hosp Clinicas Porto Alegre, Porto Alegre RS, RS, Brazil; 2) Bioethics Service, Research and Post-graduation Group, Hosp Clinicas Porto Alegre, Porto Alegre, RS, Brazil; 3) Department of Social Medicine, Universidade Federal do Rio Grande do Sul School of Medicine, Porto Alegre, Brazil; 4) INaGeMP - Instituto Nacional de Genética Médica Populacional, Porto Alegre, Brazil; 5) Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Introduction: the integration of Medical Genetics (MG) into Primary Health Care (PHC) seems to be an alternative to develop actions of improve diagnosis, prevention and management of genetic conditions, and to facilitate access of patients to specialized health care. Therefore, it is desirable that PHC providers obtain basic knowledge about MG, including the immediate measures to be taken when dealing with the most important conditions. Aims: to provide remote support to PHC providers to support their actions when facing a patient/family with a genetic condition, and to provide a long range continuing education program in medical genetics to PHC providers. Methods: a remote support facility was set up in 2013 at Medical Genetics Service, Hospital de Clinicas de Porto Alegre, in the Southern part of Brazil, accessible by the toll-free telephone number 0800 642 6761(in addition to email and internet).This new service was advertised at public health services of the Porto Alegre area by displays, and to a broader area also by emails. In addition, a continuing free education program in Portuguese was built, which could be accessed over the internet by PHC providers. Results: in the first months of activity, `Hello Genetics' was accessed by phone, email and internet by a growing number of PHC providers, mainly from the geographic area where advertisement was performed with displays; registration for the internet courses is also being increasingly requested(courses to be started in Q3 2013), mostly from PHC providers from the same region. Conclusions: the proposed strategy seems to fill a gap, helping PHC providers to access specialized medical genetics centers, obtaining the information needed to better manage and refer patients, and improving knowledge in the subject. However, advertisement should be increased to enable the program to expand nationally.

2412W

Experimental and Computational Biologists: Lessons learned from a transdisciplinary training program. *M.A. Fernandez-Altuna[1], I. Imaz-Rosshandler[2], A.L. Escobedo-Ramos[1], M. Mata-Sotres[1], C. Rangel-Escareño[2].* 1) National Institute of Genomic Medicine, Education and Outreach Office, Mexico City, México; 2) National Institute of Genomic Medicine, Computational Genomics, Mexico City, México.

The National Institute of Genomic Medicine is the reference center of research on genomic medicine in México. It hosts scientists, technicians and students from many disciplines within the biological sciences. There is also the counterpart group of scientists in the fields of mathematics, statistics, physics and computational sciences. A collaborative effort between the director of the Education and Outreach office and scientists of the Computational Genomics consortium implemented a 4-week transdisciplinary training program. Idea adapted from the original cross-training program at the Broad Institute. The program was a real success the first time in 2012 and even more for the 2013 run. We would like to share the experiences from both years and present a full panorama of lessons learned from the initiative. We will provide educational perspectives on creating graduate programs that efficiently combine a quantitative and qualitative training in Biology.

2413W

Human Genetic Variation: A Flipped Classroom Exercise in Cultural Competency. *K. Tuttle[1], S. Dasgupta[2].* 1) Pediatrics, Massachusetts General Hospital, Boston, MA; 2) Medicine, Biomedical Genetics, Boston University School of Medicine, Boston, MA.

PURPOSE

Our understanding of human genetic variation has deepened through the Human Genome and International Hap Map Projects, which gave us a high-resolution view of human genetic variation and ancestry. Applying this knowledge to evaluation of ancestry-based genetic testing strategies is a key component of the practice of culturally competent medicine.

METHODS

To emphasize the clinical relevance of population genetics, we designed a required flipped classroom module for the first year Medical Genetics course at Boston University. The concepts of population genetics and Hardy-Weinberg equilibrium are introduced through an interactive on-line exercise. In class, the students then explore application of these concepts to clinical cases, specifically considering how geographic ancestry and genetic variation influence the appearance of false positive and negative results, the ramifications of resulting variants of unknown significance or incidental findings, and the psychosocial aspects of direct-to-consumer genetic testing.

RESULTS

The introductory exercise's pre-test and post-test as well as the course final exam allowed us to assess knowledge gains over time in understanding the interplay between historical human migration patterns and genetic variation in modern populations. 179 medical students were involved in the study. The percentage of students getting the correct answer on the conceptual population genetics questions on the pre-test, post-test, and final exam were 60.2%, 76.8%, and 95.4% respectively. Student performance on the quantitative Hardy-Weinberg questions on the pre-test, post-test, and final exam, demonstrated that 50.2%, 92.1%, and 62.4% answered correctly. We were also able to directly measure the positive impact of the flipped classroom module through the use of audience response technology and course survey results, which demonstrated strong student engagement with the content.

CONCLUSIONS

The curricular design of this session allows us to introduce complex, critical concepts relating to cultural competency in a scientifically rigorous way. In order to help our future physicians gain the scientific expertise to practice culturally competent medicine, we will need to promote integration of similar exercises into the pre-clerkship phase of medical school curricula.

2414W

How is high school students' genomic literacy correlated with their attitudes toward promotion of genomic studies? A comparison with a result of general public survey in Japan. *I. Ishiyama[1], T. Tanzawa[2].* 1) Department of Child Care and Education, Teikyo-Gakuen Junior College, Yamanashi, Japan; 2) Faculty of Education, Shizuoka University, Shizuoka, Japan.

This study aimed to 1) assess high school students' genomic literacy level, 2) reveal their attitudes toward genomic studies promotion related to medicine, and 3) identify associated factors to the approval of promotion. High school students were sampled from 2 high schools to be sure to contain a variety of academic level students. The number of respondents was 354. Attitude survey was conducted in the `Basic Biology' class at each school on January 2013. Students have already studied the unit of genetics. The date was analyzed using logistic regression models and compared with the results of nationwide opinion survey which was targeted general public and conducted in the same period. These surveys included the following same queries; 1) subjective and objective understanding of genome science, 2) pros and cons of medicine-related genomic studies promotion, and 3) attitudes toward science in general. Responses of subjective and objective understanding were calculated into scores of genomic literacy. Concerning high school students, the average score of subjective understanding was 8.6 (range 0-10), objective understanding was 5.3 (0-10), genomic literacy was total 13.5 (0-20). The average score of genomic literacy in general public was 10.5 (0-20). The term of `Genome' was recognized by 99.7% students, while 31.9% of general public responded that they never heard of the term. Result also showed that approximately 3/4 students favored genomic studies related to medicine. Pros and cons were divided into agree (74.8%) and cannot decide (24.4%), while only 0.8% expressed disagree. The similar result was shown in general public survey (agree: 71.6%, cannot decide: 27.6%, disagree: 0.9%). No correlation was found between literacy level and approval of promotion in students, whereas strong correlation was found in general public. Interest in science and technology in general, and willingness to use new products developed by technological innovation were correlated with the approval. In conclusion, even if high school students are knowledgeable about genomics, they don't necessarily support the genomic studies in medicine. If anything, it's important for them to develop their interest and trust in science through biology class.

2415W

Geneticists Teaching Genetics to New Orleans K-12 Schools and Universities. *F. Tsien, J. Loupe, A. Iyengar.* Department of Genetics, LSU Health Sciences Center, New Orleans, LA.

While the latest breakthroughs within the field of genetics continue to increase in complexity, they are more frequently mentioned in the media and are becoming increasingly relevant to our health, highlighting the growing need for the public to become more “genetics literate”. One way to better inform the public is to incorporate genetics concepts into K-12 school and college curricula in effective, novel ways. The LSU Health Sciences Center (LSUHSC) has created two programs: the LSUHSC/New Orleans Schools Science Partnership Program for 4th graders and the LSUHSC Hands-on Genetics and Career Workshop Program for middle, high school, and undergraduate students (established in 2006 and 2009 respectively). Geneticists have partnered with teachers to make genetic concepts and other science topics easier to understand and less intimidating, to improve academic achievements in science, to introduce K-12 and undergraduate students to diverse role models in the sciences (including females and underrepresented minorities), to increase students' awareness of potential career paths in science and genetics, and to train geneticists to present to the lay public and interact directly with the community. Instructors include faculty, graduate students, and postdoctoral fellows from the LSUHSC Genetics Department. Activities involve hands-on experiments and videos pertaining to genetic concepts found in the media (cancer, forensics), and topics from the students' curriculum (cells, DNA). We have established a science and genetics education “pipeline” that includes the elementary school, middle school, high school, and college levels, leading to genetic internships in our institution. Anonymous pre- and post- workshop evaluations and demographic questionnaires reveal that students understand and retain lesson information regardless of socioeconomic or academic background. Students' quantitative standardized test and classroom exam scores demonstrate gains following implementation of the programs. Qualitative evaluations of student gains in knowledge include a Science Jeopardy competition, science fair projects, internships, and anecdotal information from teachers and students. These programs have introduced genetic concepts to approximately 750 elementary school students, 220 middle school students, 1055 high school students, and 250 undergraduates. The hope is that these programs will ultimately inspire some students to pursue lifelong careers in the genetics field.

2416W

Accelerating public awareness of personal genetics. *M.E. Gelbart, L. Tomaselli, D. Waring, T. Wu.* Personal Genetics Education Project, Department of Genetics, Harvard Medical School, Boston, MA.

The Personal Genetics Education Project (pgEd; www.pged.org) is using a unique platform of strategies to address the widening gap between what researchers are learning at the frontiers of genetics and what the public understands. First, pgEd is working to advance awareness through television with Hollywood, Health & Society at the USC Annenberg Norman Lear Center. Second, pgEd is continuing to develop the on-line game, Map-Ed (www.map-ed.org), which we launched in March to spread awareness of core concepts in genetics. Third, pgEd organizes the annual GETed conference, which brings together experts in education, research, health, entertainment, and policy to develop strategies for accelerating public awareness. Fourth, pgEd is building a freely available on-line curriculum that addresses the benefits as well as the ethical, legal, and social implications of personal genetics. Finally, pgEd is training high school teachers in biology, health, social studies, and beyond to engage an entire generation of young people on the topic of personal genetics.

2417W

Parents and newborn screening decisions: empirical assessment of specific educational messages. *B.J. Wilson[1], B.K. Potter[1], J.C. Carroll[2], J. Little[1,3], D. Castle[4], D. Avard[5], P. Chakraborty[6,7], S. Craigie[1], H. Etchegary[8], L. Lemyre[9], F.A. Miller[10], G.A. Wells[1], J. Millburn[6], R. Rennicks White[11, 12], G. Tawagi[11], M. Walker[11, 12, 13], CIHR Emerging Team in Genomics in Screening.* 1) Department of Epidemiology & Community Medicine, University of Ottawa, ON, Canada; 2) Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 3) Canada Research Chair in Human Genome Epidemiology); 4) ESRC Innogen Centre, University of Edinburgh, Edinburgh, UK; 5) Centre of Genomics and Policy, McGill University, Montreal, QC, Canada; 6) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 7) Department of Pediatrics, University of Ottawa, ON, Canada; 8) Clinical Epidemiology, Memorial University Newfoundland, St John's, NL, Canada; 9) School of Psychology & Institute of Population Health, University of Ottawa, ON, Canada; 10) Institute of Health Policy, Management and Evaluation, University of Toronto, ON, Canada; 11) Department of Obstetrics, Gynecology and Newborn Care, The Ottawa Hospital, ON, Canada; 12) Clinical Epidemiology Program, Ottawa Hospital Research Institute, ON, Canada; 13) Department of Obstetrics and Gynecology, University of Ottawa, ON, Canada.

Background: In many jurisdictions, systematic parental education about newborn screening (NBS) seems to be limited, perhaps because screening may be seen as routine care and not a separate intervention requiring conscious decision-making by a parent. However, even where screening is mandated, parental education provides benefits such as promoting trust in health services and preparing for the process and potential outcomes of NBS. However, NBS programs may be concerned that more comprehensive education may lead to higher parental anxiety and lower acceptance rates. Debates on secondary bloodspot use are also relevant to this issue. `Education' is a broad concept, and there appears to be no empirical evidence on the most important content of education for parents. This study is designed to formally test the impact of specific components of potential NBS educational content on parents' decision making. **Objectives**: To measure and compare expecting mothers' responses to specific NBS educational messages in a jurisdiction where screening is encouraged but not mandated. **Methods**: The study sample is women with low-risk pregnancies attending routine second trimester ultrasound clinics in Ottawa, Ontario. Using a factorial survey, participants are randomized to receive different combinations of messages about NBS: the possibility of false positive/negative results; pain from the heel-prick; potential for over-diagnosis; storage/secondary use of bloodspots; and nature of parental consent (25=32 discrete message combinations being assessed). The primary outcome is the mean score (max 100) on the Decisional Conflict Scale (DCS). **Results**: From a target sample size of 500, we report data on the first 129 respondents. The mean DCS score is 27 (range 0-89). Over 90 percent intend to accept NBS; actual uptake will be tracked prospectively, with consent. In univariate analyses, no statistically significant differences are observed in DCS scores in those receiving/not receiving specific messages, with the exception of the storage/secondary use of bloodspots message (p=0.02), a lower mean DCS score associated with receiving specific information. Further analyses will examine interactions between specific messages, the effect of the absolute number of messages (cognitive burden), and associations with other respondent characteristics including their apparent understanding of key messages. Implications for design of NBS parental education will be discussed.

2418W

On 30th anniversary of Orphan Drug Act: a review of resources to inform and connect rare disease patients. *M.H. Dunkle[1], J. O'Leary[2], H. Hyatt-Knorr[3], M. Snyder[3], J. Lewis[3].* 1) Communications, National Organization for Rare Disorders (NORD), Danbury, CT; 2) Genetic Alliance, Bethesda, MD; 3) National Institutes of Health Genetic and Rare Diseases Information Center, Bethesda, MD.

The Orphan Drug Act (ODA) of 1983 brought major new incentives to encourage the development of therapies for patients with rare diseases. It also launched a coordinated effort, involving patient organizations, government agencies and others to address the unique and previously unaddressed needs of patients and families affected by rare diseases. During this 30th anniversary year, three of the primary sources of information and connection for rare disease patients provide a brief review of the history of rare disease information since 1983 and highlight a few of the major resources available to patients and families today. The three collaborating entities are the National Organization for Rare Disorders (NORD), which was established by patient organizations in 1983 after the ODA was enacted; Genetic Alliance, established in 1986 by NORD, the March of Dimes, and the Maternal and Child Health Bureau of the Health Resources and Services Administration; and the NIH Genetic and Rare Diseases Information Center (GARD), established in 2002 by the NIH National Center for Advancing Translational Sciences/ Office of Rare Diseases Research and the NIH National Human Genome Research Institute. Current initiatives include NORD's Rare Disease Database, regional patient organization meetings, and RareConnect, global online patient communities established by NORD and its European counterpart, EURORDIS; Disease InfoSearch, an online database of advocacy organizations and resources for genetic conditions and Genes in Life, to provide easy access to Genetic Alliance's many resources and tools related to health and genetics information; and the GARD Information Center for patients and their families, as well as researchers and medical professionals, in which trained Information Specialists staff a toll-free phone line, provide customized responses in English and Spanish, and create web content based on questions received from the public so that it is responsive to public `here and now' issues. Together, these three entities, NORD, Genetic Alliance, and GARD, are providing valuable and complementary information services to the rare disease community.

2419W

`Will he be bald?' Introducing the general public to the genomic era. *M. Kriek[1], K. Kraaijeveld[1,3], H. Sminia[2], B. Gravendeel[4], M. Schilthuizen[4], J.T. den Dunnen[1].* 1) Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, Leiden, Netherlands; 2) Netherlands Bioinformatics center (NIBC), Nijmegen, The Netherlands; 3) Hogeschool Leiden, The Netherlands; 4) Naturalis Biodiversity Center, Leiden, The Netherlands.

We stand today on the threshold of a new era. Within the foreseeable future analysis of complete genomes will play a prominent role in society in the broadest sense. It will have immediate implications for healthcare, agriculture, biodiversity research and many other aspects of the world around us. LeveDNA!* is a Dutch initiative that aims to create public awareness of DNA-sequencing and its applications. The focus is on high school students, as they already possess some background knowledge and their generation is thought to be exposed to DNA sequencing most frequently in the near future. Our approach is to invite the public to contrive interesting questions that can be addressed by DNA sequencing. LeveDNA! attempts to realize such ideas by finding support for sampling, sequencing and analysis. Secondly, we actively initiate projects that bring our target audience in contact with DNA sequencing, often in collaboration with scientific partners. Projects conducted so far include: (I) Sequencing the genome of a DJ from a popular Dutch radio station, followed by a dedicated evening of his daily radio show. During the three-hour broadcast, listeners were given the opportunity to ask questions about the DJ's genome. 'Is he a carrier of a genetic disorder? Will he be bald? Is he a night owl?' are some of the many questions asked. (II) LeveDNA!, in collaboration with an initiative to teach the public about evolution (Evolution MegaLab), made an attempt to identify genes that contribute to the colour patterning of garden snail shells. (III) Organisation of a DNA-day, where participants are taught about DNA and where they can isolate and analyse small parts of their genome. (IV) Shaking Science, a biodiversity project where students sampled fresh water around the Natural History museum; the sequence determined can be analysed by everybody using a simplified Blast tool. Finally, (V) high school students sequencing organisms in fresh water of a major river in the Netherlands, the Rhine. LeveDNA! creates a platform where the general public and experts directly interact, exchange knowledge and discuss social consequence of the new technologies. *LeveDNA!: “Leve” (Dutch) stands for both “life” and “hurray”.

2420W

Personalized medicine - what do people think? *L. Leitsalu[1,2], A. Allik[1], A. Metspalu[1,2,3].* 1) Estonian Genome Center of the University of Tartu, Tartu, Estonia; 2) Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Estonian Biocentre, Tartu, Estonia.

Estonia has several factors favorable for large-scale, innovative projects in the healthcare sector, such as implementing personalized healthcare due to several reasons. The factors that the genomic information originates from a single center, the health care sector is regulated on a national level, having one major healthcare funding office, and a single university educates all healthcare professionals, all contribute to a unified system. Additionally, the relatively small population size of 1.34 million makes it logistically more feasible to implement projects on a nationwide scale. The success of such a project depends largely on the support of the public. Public's perspectives were investigated through a phone survey by TNS Emor polling agency in spring of 2011. The sample included 1000 individuals between the ages of 15-74 and was composed to be proportional to the Estonian population structure with respect to age, gender, and nationality. The topics investigated included - role of genetics in common complex diseases, attitude towards the use of genetic information, motivators and barriers for genetic testing, and reported hypothetical effect on health behavior. The results of the first wave of the survey suggested that the majority is interested in genetic testing and that genetic information would reportedly increase the likelihood that doctor's recommendations for changes in health behavior will be followed. The most common motivator for testing was the opportunity for doctors to provide better care, while the most common barrier was fear for receiving unwanted information. The open-ended comments demonstrated several misconceptions about genetic testing and more information on the topic of genetic testing was frequently asked for. This spring the Government of Estonia has officially expressed its support for introducing personalized healthcare in Estonia. There has been an increasing amount of media coverage on topics of genetic testing and personalized medicine over the past few years. Now we are conducting a second wave of the survey to see whether any changes can be detected over the past two years. The results are used for the development of educational resources in order to improve genetic literacy among the general public to avoid unsound effects, unrealistic expectations and misconceptions, and facilitate effective use of genomic information in personalized healthcare.

2421W

Vignettes as an aid to deciding about genetic testing. *D. Zallen.* Dept Sci & Tech in Society, Virginia Tech, Blacksburg, VA.

Genetic tests are currently offered in doctors' offices and by commercial laboratories widely advertised on the Internet. Though genetic information can be very useful for some people, for others it can produce severe personal and family problems without any significant corresponding medical benefit. Within the confines of a standard medical office visit, it is difficult to provide consumers with the information that will enable them to make informed decisions about whether a particular form of genetic testing is right for them. Also, there are too few genetic counsellors. New educational approaches are needed. Consumers deciding whether or not to test need guidance not only in understanding the technical aspects of genetic testing but also in assessing its suitability with regard to their personal values, psychological makeup, and family dynamics.

As part of an investigation of new educational tools, we have developed a set of short dramatized vignettes based on more than 150 interviews (D.T. Zallen, “To Test or Not to Test”, Rutgers 2008) with consumers who were considering genetic testing for susceptibility to common disorders. Testing for predisposition to the late-onset form of Alzheimer's disease was selected as the central focus. The vignettes, presented by male and female actors of various ages and heritage backgrounds, draw on the actual experiences of consumers and the reasons they have given both for and against having testing.

These vignettes were reviewed in two stages by three different focus groups whose participants were recruited through ads placed in the major regional newspaper. All the focus groups found the dramatized vignettes compelling and balanced. The first group judged that the visual background needed to be softened and the actors placed in a more natural setting. The vignettes were then re-shot with such a setting. The prevailing view of the subsequent two groups was that the vignettes were more informative than relying solely on written text and were an effective way to help them identify their own value preferences and reach appropriate decisions.

The use of dramatized vignettes is a novel and effective approach to providing consumers with a means of clarifying the various value dimensions associated with genetic testing prior to deciding whether or not to undergo a genetic test.

2422W

Incorporating Bioinformatics into the Undergraduate Genetics Curriculum through an Authentic Research Project. *B.V. Bowling, E.D. Strome.* Biological Sciences, Northern Kentucky University, Highland Heights, KY.

Genomics and bioinformatics are advancing fields that have become an integral part of a broad spectrum of biological research. As a result, it is increasingly important for undergraduate genetics courses to include basic bioinformatic tools and genomic concepts to prepare students for the modern biology workforce. At Northern Kentucky University, we've introduced these through an authentic research project in which students investigate the role of *Saccharomyces cerevisiae* open reading frames designated as coding for `putative proteins' within the Saccharomyces Genome Database. Students perform wet-lab experiments using classic genetic techniques in order to create knockout strains for their gene of interest and then assess phenotypes; as well as bioinformatics analyses conducted on individually assigned ORFs and then data are synthesized in order to develop a hypothesis of the function of the gene product within the cell. The bioinformatics and comparative genomics analyses include multiple sequence alignment, conserved domain identification, cellular localization prediction, reviewing interaction and expression data, and gene deletion phenotypes. The *in silico* approaches allow students to apply genetic (and additional biological) concepts to an authentic research question, provide experience with bioinformatics tools, and challenge students to analyze and synthesize data from various sources. The project culminates in a brief presentation by each student in which they summarize the most important evidence collected for the role of their gene product in the cell. These activities have been shown to be effective in our sophomore-level, multi-section laboratory course involving several instructors. Instructors report increased engagement relative to previous labs and a substantial portion of the students (74%) felt that they learned even more from this project than other lab projects. The vast majority of the students indicate the research was challenging (93%) and that they are better prepared to take on investigating a problem without a known solution (82%). The lack of established answers contributed to student anxiety about the assignment. Many students have the urge to check to see if they are `right,' and instructors have to balance encouraging independence while also reassuring students of their analyses. Incorporating formal mechanisms for peer feedback throughout the process has helped with students' confidence in their work.

2423W

Integrating authentic research in human genetics into an undergraduate liberal arts curriculum to enhance the pace of rare disease gene discovery for underserved populations. *R. Jinks[1,2], B. Davis[1], D. Roberts[1], E. Rice[1], K. Brigatti[1], E. Puffenberger[3], K. Strauss[3,4].* 1) Biology, Franklin & Marshall College, Lancaster, PA; 2) Biological Foundations of Behavior Program, Franklin & Marshall College, Lancaster, PA; 3) Clinic for Special Children, Strasburg, PA; 4) Lancaster General Hospital, Lancaster, PA.

Next-generation sequencing accelerates disease gene discovery, especially for orphan diseases, though at present it outpaces functional studies needed to provide 'proof of causation.' This problem was underscored by a recent funding opportunity from the NHGRI requesting assistance with functional studies for 75 rare, potentially pathogenic alleles discovered through the NIH Undiagnosed Diseases Program. Working from a small, community-based, clinical laboratory, the Clinic for Special Children has identified more than 130 allelic variants associated with disability, disease, or untimely death among the Amish and Mennonite (Plain) populations of North America. In collaboration with the Clinic, we developed an HHMI-funded program that integrates rare disease gene discovery research into our undergraduate curriculum. Roughly 150 students per year in our introductory cell biology and neuroscience courses clone disease genes and study the functional impacts of gene variants through expression in mammalian cell culture. We use these authentic research experiences to teach key concepts in cell biology, genetics, and neuroscience. Students later build upon this experience in upper-level courses in neuroscience, genetics, cancer biology, and immunology in which they engage in semester-long research projects in small teams, each conducting functional studies of a single novel gene variant. These experiences engender talented undergraduates to assume greater research responsibility through independent study and summer research projects that transition them into the role of co-PI for the gene and disease they have chosen to study. Data published through this project (PLoS ONE 7:e28936) were recently used by another institution to diagnose a critically-ill, non-Plain newborn with lethal neonatal seizure-rigidity syndrome (Sci Tran Med 4:154ra135). This provides important proof-of-concept for integrating novel disease gene functional studies into a carefully structured undergraduate research curriculum. Students in our Public Health program are collaborating with the Clinic to develop handbooks to help parents care for children with special medical needs with the goal of producing one to two high-quality disease handbooks annually. Our program represents a model for engaging undergraduates in meaningful research at the front lines of biomedical science and public health in ways that directly impact the diagnosis and care of children with rare inherited disorders.

2424W

Inverted Curriculum: An Online Activity Replaces In Class Lecture to Teach Regulation of the Lactose Operon in a General Genetics Undergraduate Classroom. *T. Kelson[1], M. DeWall[1], S. Minnich[2], G. Arrizabalaga[2], J. Anderson[2], D. Knox[2].* 1) Biol Dept, Brigham Young Univ, Idaho, Rexburg, ID; 2) Imaging and Science Engineering Education Group, Univ of Idaho, Moscow, ID.

A revolution currently underway to change the traditional college classroom is being referred to as the "flipped" classroom. Instructors provide resources and require students to learn content on their own outside of lecture and then, class time is used to review, answer questions, practice problems, and apply concepts. In short, lecture is moved out of the class and `homework' is moved in. We are investigating a novel approach to undergraduate teaching in Genetics - a 3 credit undergraduate General Genetics class where students spend 2 hours in lecture each week and 1 hour outside of class in meaningful, supervised activities. In order to accomplish this we are investigating the use of online activities that will teach the material in much the same way that a faculty member would in a face-to-face lecture environment. The University of Idaho has developed an online tool to teach gene regulation in bacteria, specifically the lactose operon. We tested the efficacy of this online tool over a 2 semester time period, as follows - we taught the same material to students from 2 undergraduate General Genetics classes: one in a formal lecture setting (n=36) and the other through the use of this online tool (n=52). After instruction, both groups were administered the same homework assignment. Our results showed that both groups performed equally well on the same homework assignment (average score = 94%). As a follow-up to assess retention, the students were tested on this same material one month later, with results supporting that they did remember much of what they learned the first time (average score = 71%). Each participating student also completed a questionnaire. When asked about the online tool, most students responded that they found the online activity useful and that it helped their comprehension of the lactose operon. Based on our positive results over several semesters, we no longer lecture on the lactose operon in our classroom; instead, the students complete an online assignment prior to coming to lecture, and we now spend classroom time developing critical thinking skills in gene regulation. We use problem-based and teamwork activities in the classroom to enhance deeper learning of gene regulation. We have demonstrated that the use of technology can enhance, not replace, instruction of genetics topics that, in the past, have been difficult for some students to learn in lecture settings.

2425W

The Research Connection: Development of an integrated institutional pediatric research infrastructure as a model framework for multi-institution implementation and collaboration. *S.K. Savage[1], C.M. Clinton[1], W.A. Wolf[1,2], D.M. Margulies[1,2,3,4].* 1) Program in Genomics, Boston Children's Hospital, Boston, MA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA; 3) Division of Developmental Medicine, Boston Children's Hospital, Boston, MA; 4) Center for Biomedical Informatics, Harvard Medical School, Boston, MA.

Our objective is the development and implementation of a fully-integrated research infrastructure to power and enhance discovery within Boston Children's Hospital (BCH) and across collaborating institutions. To achieve this goal, we reviewed and assessed existing research resources and identified gaps and areas of need. We then designed a set of core research and clinical services, called the Research Connection, to supplement existing operations and handle the varied needs of BCH investigators, clinicians, and collaborators. The BCH Biobank is patient-centric repository of leftover clinical samples linked with electronic medical record data to facilitate biospecimen-based research. By offering Biobank enrollment across the entire institution, we attain broad reach over a diverse population of children and adults, creating a large and rich research cohort. Claritas Genomics, BCH's new genetics and genomics diagnostic laboratory, provides cutting-edge measurements, analyses, and revisable reporting for both research and clinical samples. BCH's Interpretive Genomic services allow for refined interpretation, clinical correlation, and specialty care management, drawing on clinical and scientific expertise within the institution. Finally, our novel approach to integrated clinical and research genetic result return leverages existing and expanding BCH clinical services to provide appropriate result disclosure and genetic counseling. This design allows our research infrastructure to be optimized to accommodate multi-institutional collaboration via a new collaborative agreement structure called the Pediatric Research Network. To support the concept of a Pediatric Research Network, we have initiated collaborative relationships with other pediatric hospitals, where components or the entire Research Connection model may be adopted and tailored to meet their unique institutional needs. By implementing a common research infrastructure like the Research Connection across multiple sites, coupled with a collaborative agreement across institutions, we create an opportunity for an extensive 'virtual' research repository to support large-scale research, foster collaborations, and enhance discovery that would otherwise not be possible. This structure may be the future of genomic research design for large and small institutions.

2426T

Revisiting the role of gamete and embryo donor registries on the transmission of genetic information between donation relatives, in light of the principle of beneficence. *V. Couture[1], M.-A. Dubois[1,2], R. Drouin[1], J.-M. Moutquin[3], C. Bouffard[1].* 1) Division of Genetics, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Quebec, Canada; 2) Me Hélène Guay Office, Montreal, Quebec, Canada; 3) Department of Obstetrics and Gynecology, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Quebec, Canada.

Introduction: In the context of assisted reproductive technology, the transmission of genetic information between relatives may have implications for prevention, diagnosis, treatment, and reproductive decision-making. In the case of gamete and embryo donor-conceived people, where there is no previous contact between donors and their sibling, this issue acquires greater complexity. The creation of gamete and embryo donor registries is one of the strategies that has been developed to make information available between donation parties. However, the nature, flow, preservation, and accessibility of genetic and medical information depend on the type of registry. Whether national, institutional or independent, each type may have different repercussions on the health protection of donor-conceived people, donor, siblings, and offspring. Under these conditions, our objective was to identify the ethical implications of the transmission of genetic and medical information through donor registries, in light of the principle of beneficence. Methods: Bioethical analytical approach: (1) typology of donor registries (2), analysis of their strengths and weaknesses, and (3) introduction of the principle of beneficence in the analysis of their impact. Results: A donor registry has as its main benefits the opportunity to track hereditary disease and prevents future gamete donation by an affected donor. It can also help in the diagnosis of genetic disease as well as in the prevention of late-onset conditions. According to the principle of beneficence, the needs and values of all donation parties have to be considered. A registry must address the issues surrounding the uncertain liability of donors associated with the transmission of genetic and medical information, as well as the protection of their anonymity. An effective counseling service has to be provided to avoid unnecessary distress for all the parties. To maximize the benefits of gamete and embryo donor registry, the national type with proper counseling and clear guidelines appears to be the most respectful solution to promote the health of all donation parties. Conclusion: Despite the complexity of genetic and genomic issues for the development of a national donor registry, the principle of beneficence offers us the opportunity to pursue the present and future needs of all donation parties.

2427W

Participant views of re-consent and broad consent in cancer genetics research. *K. Edwards[1, 2], D. Korngiebel[3], L. Pfeifer[2], J. Scott[4], N. Shridhar[1], D. Kaufman[5], D. Bowen[6], C. Condit[7].* 1) Institute for Public Health Genetics, School of Public Health, University of Washington, Seattle, WA; 2) Epidemiology, School of Public Health, University of Washington, Seattle, WA; 3) Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA; 4) National Coalition for Health Professional Education in Genetics, Lutherville, MD; 5) Genetics & Public Policy Center, Johns Hopkins University, Washington, DC; 6) Community Health Sciences, School of Public Health, Boston University, Boston, MA; 7) Communication Studies, University of Georgia, Athens, GA.

Background: Institutional Review Boards and researchers protect participants in genetic research, but the preferences of participants do not sufficiently inform policy currently. With the proliferation of biobanks intended to support future studies that go beyond the scope of their original consent, fully understanding participant preferences for re-consent is a pressing concern. The Participant Issues Project (PIP) addresses this gap. Methods: PIP study participants were drawn from the Northwest Cancer Genetics Registry and included cancer patients, controls, and relatives. Thirty telephone interviews were conducted and analyzed using content and thematic analysis. Questions asked participants to consider diverse scenarios: the research focus had changed from the original study; re-consent of a minor achieving majority; for-profit studies; and the use of broad consent forms. Results: The majority of participants agreed that re-consent was necessary when the study direction changed significantly (e.g., a different disease) or a child participant became an adult. Participants favored re-consent in order to provide study information and control of the use of their data, However, participants felt re-consent was not needed if the study shifted to examine a different gene's effect on the same cancer. Several participants indicated that genetics was a broad umbrella that covered the gamut of gene-related research, and some expressed a strong preference that research make best possible use of study data for the greater good. Most participants' willingness to participate in research would not be affected if the researcher or institution profited or if a broad consent form were used. Conclusions: There were specific scenarios where the majority of participants felt re-consent was needed and other scenarios where it was not needed. For many participants, the perceived value of re-consent appears to lie largely in keeping participants informed of future studies and updated on the research progress, while for others, it is to control the use of their data. It is not known whether satisfying the need for information will change participant perspectives on the desire for re-consent.

2428T

Balancing Patient Privacy While Supporting Rich and Convenient Access to Clinical Genomic Data. *E.R. Riggs[1], J. Berg[2], C. Bustamante[3], D.M. Church[4], W.A. Faucett[1], M. Feolo[4], D.H. Ledbetter[1], D. Maglott[4], C.L. Martin[1], D. Metterville[5], J. Mitchell[6], R. Nussbaum[7], J. Ostell[4], S. Plon[8], H. Rehm[5,9], L. Rodriguez[10], W. Rubinstein[4], M. Watson[11], International Collaboration for Clinical Genomics.* 1) Geisinger Health System, Danville, PA; 2) University of North Carolina, Chapel Hill, Chapel Hill, NC; 3) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 4) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; 5) Laboratory for Molecular Medicine, Partners Healthcare Center for Personalized Genetic Medicine, Cambridge, MA; 6) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; 7) Division of Genomic Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA; 8) Departments of Pediatrics and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 9) Department of Pathology, Harvard Medical School, Boston, MA; 10) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 11) American College of Medical Genetics and Genomics, Bethesda, MD.

It is becoming clear that access to large datasets of genomic variants and observed phenotypic consequences is critical to our understanding of human genomic variation and its applications to basic science, public health, and improved patient care. However, these datasets must be maintained in ways that both respect patient privacy and support active use by the scientific community. The International Collaboration for Clinical Genomics (ICCG) is a group of laboratories, clinicians, and researchers working closely with the National Center for Biotechnology Information (NCBI) to develop ClinVar, a publicly available repository of genomic variants, associated phenotypic information, and clinical classifications with supporting evidence. Laboratories are beginning to submit their data, which include a mixture of both de-identified and potentially identifiable information regarding variants and individual cases observed during the course of clinical testing and research studies. Given that a large amount of ClinVar's data will be derived from patient testing results and may warrant different levels of protection, it is critical that ICCG develops clear policies regarding patient privacy, and appropriate data access. Current data access policies and requirements of similar databases were initially developed to accommodate traditional research applications of the data, and remain appropriate for these types of applications; however, many of the requirements, such as IRB-approved protocols, annual usage reports, and the requirement to fully download entire datasets in order to search for variants may represent barriers to potential clinical users. Possible solutions may include a clear statement of appropriate use, alternative methods of user authentication, accurate data use accounting, and interactive, dynamic methods of data access. Representatives from ICCG, NCBI, NIH, and other stakeholder groups are collaborating to discuss what types of initiatives may be necessary. Our goal is to create an environment of 'adaptive governance' to support the unique needs of our multi-faceted user community while maintaining the confidentiality of our patient participants.

2429W

Utility of the X chromosome pattern of inheritance: the identification of close relatives through direct-to-consumer (DTC) genetic testing. *K. Johnston.* Retired from HealthCare Partners Medical Group, 3565 Del Amo Blvd. Torrance California 90503.

Members of the public who are looking for their birth parents can now order autosomal tests directly from ancestry companies to aid in their search for cousins who carry identical by descent (IBD) DNA segments. Females who are looking for their birth fathers, and both males and females who are looking for their birth mothers can utilize the matching X chromosome IBD segments they receive simultaneously from these companies. Extensive pedigrees compiled by close cousins who are related by way of the X chromosome provide useful links. The inheritance pattern of the X chromosome is unique. When studying pedigrees, all lineages that show two males in a row (father to son) can be eliminated from the family tree. X chromosome tracking aids in the identification of the most recent common ancestor(s) between two closely related DNA cousins. In addition, extensive online genealogical records submitted by the public, especially those that include living relatives, facilitate in the identification of collateral lines. In some cases, paternity testing is no longer needed to confirm a lineage when close relatives (e.g. half-siblings) are contacted, then are willing to be tested and matched through one of the DTC ancestry companies or online third party tools. Unsuspecting birth fathers, sperm donors and parents who have put up a child for adoption can now be traced through online pedigrees provided by closely related genealogists with the help of public records and familial DNA testing. An example involving two recently matched second cousins will demonstrate the steps involved.

2430T

NIH's approach and expectations for genomic data sharing: The draft NIH Genomic Data Sharing Policy. *D. Paltoo[1], A. Bailey[2], C. Fomous[1], K. Langlais[1], E. Lueketmeier[1], T. Paine[1], R. Wise[2], L. Lyman Rodriguez[2].* 1) Genetics, Health, and Society Program, Office of Science Policy, Office of the Director, NIH, Bethesda, MD; 2) Division of Policy, Communications and Education, National Human Genome Research Institute, NIH, Bethesda, MD.

Since 2007, the Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) has governed the sharing of GWAS data from human participants in research supported by NIH funding. The GWAS Policy established the expectation that all such data be made available, with exceptions as warranted, for secondary research uses that are consistent with research participants' informed consent. Rapid advances in DNA sequencing and other high-throughput technologies have increased the volume, complexity, and types of data generated in genomic studies. To ensure the full value of genomic data in light of this growth, NIH has drafted the Policy for Genomic Data Sharing in NIH-Supported or NIH-Conducted Studies (GDS Policy), which updates and expands on the 2007 NIH GWAS Policy. The draft GDS Policy encompasses a broader range of data types (e.g., GWAS data and genomic, transcriptomic, or epigenomic data produced by array-based technologies, or data obtained from high-throughput sequencing technologies) from research in both human and nonhuman or model organisms, and updates expectations for data submission, data release, and human participant protections. As with the NIH GWAS Policy, the draft NIH GDS Policy is founded on the principle of maximizing public benefit by facilitating broad data sharing to advance the understanding of public health needs, while ensuring the responsible oversight of genomic data sharing. Through notices in the Federal Register and NIH Guide as well as two public webinars in the summer of 2013, NIH obtained public perspectives on all aspects of the draft GDS Policy, as well as on related data management topics such as the potential benefits and harms of broadly sharing genomic research data; scientific, technological, policy, or legal strategies or factors, that would minimize risks and should be considered when genomic data is shared through open access repositories; and models that NIH should consider to facilitate access to large-scale controlled-access data, enhance data utility, and protect participant interests. The comments were publicly posted on the NIH GDS website [http://gds.nih.gov] after the close of the public comment period. NIH expects this policy to be implemented in 2014. This presentation provides an overview of the draft NIH GDS Policy, the public consultation process and key public comments, and expectations for NIH GDS Policy implementation.

2431W

The Potential Power of Personal Genomics in Reducing Social Stereotypes: Attitudinal Study and Computer Animation of Results for 4,000 Japanese Respondents. *T. Kido[1], M. Swan[2].* 1) Rikengenesis, Tokyo, Japan; 2) DIYgenomics, Palo Alto, U.S.

This paper provides a flagship study of ethics and discrimination as individuals react to information about individual genetic differences. The study reports on the results of the analysis of social physiological statistics related to the impacts of the emergence of personal genome services. We were interested in determining how we might choose to respect others, not discriminate against others, and want to know differences among people. We would also like to understand the basic mechanisms underlying the reason for why discrimination emerges in our society. We conducted focus group discussions in a small lecture group in Tokyo University. Next, based on these discussions, we developed web-based survey systems with computer animations. We conducted 30 web-based interviews with 30 questions to more than 4000 Japanese individuals. A.Focus Group Discussions Focus group discussions on personal genome services revealed that a variety of opinions exist on both positive and negative sides. We classified these opinions, and asked more than 4000 Japanese people to vote on their favorite opinions. B.Web-based surveys and statistical investigations We created animation stories for symbolizing discrimination against people with a handicap and analyzed the responses. Our preliminary results that involved survey with more than 4000 volunteers revealed the following: 56.9% of people believed that society discriminates against people, with handicaps. 61.7% of people wished to have information about their genomes, even if it revealed a handicap. These preliminary results pose interesting questions about why a majority of people might believe in the existence of discrimination in our society, while they believe that they themselves are not discriminating. Social psychological research is expected to shed further light on these questions. This work demonstrates the importance of constructing frameworks for philosophical social psychological research with respect to new emerging information technologies. One of the biggest things we have seen emerge as people have their genetic data is that everyone realizes they have one 'handicap' or another - every person is at higher genetic risk than the average for one of the top 20 health conditions. This means the frame shifts from: A. Me (non-handicapped) vs. them (handicapped) to: B. We are all 'handicapped' (genetically predisposed) for some condition This has been one of the great destigmatizing influences of genomics.

2432T

Investigating the interplay of patients' hopes and expectations of Next Generation Sequencing. *K. Clift, A. Fiksdal, R. Topazian, A. Kumbamu, J. McCormick.* Mayo Clinic, Rochester, MN.

PURPOSE: The aim of this study is to describe the hopes and expectations of individuals engaged in the process of utilizing next generation sequencing (NGS) as a clinical tool. Participants include patients presenting to the Mayo Clinic Individualized Medicine Clinic (IM Clinic) who have a cancer that has failed standard treatments or have a diagnostic odyssey, their family members, and the clinicians involved in the care of the patient. METHODS: We conducted semi-structured interviews with 16 patients and their family members who were referred to the IM clinic throughout multiple points of the process of NGS. Audio recordings of the interviews were transcribed and analyzed using inductive narrative analysis. Future research activities will include more interviews and surveys of patients and their families entering the IM Clinic and throughout their journey as well as clinicians. RESULTS: The patients verbalized their understanding that the probabilities of finding something from NGS that will benefit them directly are low. However, patients maintained optimism despite having been advised by clinicians and genetic counselors that NGS is a relatively new process in the clinical realm and as such the possibility of it resulting in a beneficial outcome cannot be guaranteed. A prominent theme that emerged from the interview data was that the patients' hopes were nuanced by more 'realistic' expectations. To illustrate, one patient said that her hopes were to find out anything possible and her mother added: 'Our hopes are that...next week you will call and say we found a gene and we can fix it' to which the patient inserted a sarcastic aside that demonstrated her awareness of the likelihood of their hopes coming to fruition: 'And we're buying you Cinderella's castle and you get to live in it!' But despite their tempered understanding, many patients upheld what they knew to be unrealistic hopes and chose to move forward with NGS anyway. DISCUSSION: Patients' hopes and expectations can influence their decision-making in deciding whether to proceed with novel clinical testing. Understanding the interplay of patients' hopes and expectations may prove valuable to health care providers and genetic counselors as they guide individuals through the process of NGS. As NGS becomes increasingly integrated into clinical care, clinicians and researchers must continue to investigate the implications of using these technologies from a patient perspective.

2433W

Genetic privacy in the European Union - exploring the impact of the proposed Data Protection Regulations. *A. de Paor.* Centre Disability Law & Policy, Natl Univ Ireland, Galway, Galway, Galway, Ireland.

With advances in genetic science and technology, genetic information is becoming increasingly available and accessible. However, the deluge of genetic information raises many ethical and legal issues that may threaten advancing science, in the absence of appropriate regulation. One such issue is the violation of genetic privacy. Genetic information is a sensitive and inherently personal type of information that can reveal intimate details about an individual and an individual's family. In light of the sensitive nature of genetic information, the potential abuse is clear, as is the desire to protect such information from access and disclosure. Therefore, there are compelling reasons to maintain the privacy and confidentiality of genetic information. In the European Union (EU), the Data Protection Directive (introduced in 1995) provides a strong data protection framework that protects the privacy of personal information. However, there is currently no substantive provision for genetic data and no specific reference to this category of information, indicating a gap in the privacy protection for genetic data in the EU. Recent developments indicate an intention to include genetic data within the scope of EU data protection laws, with the European Commission proposing a new Regulation. With these new Regulations, the Commission aims to develop an updated data protection framework. It is committed to reform and modernise data protection legislation, in line with the realities of today's society, and changing norms. The draft Regulations identify 'genetic data' as a category of personal data designated for special protection. 'Genetic data' is defined broadly to include 'all data, of whatever type, concerning the characteristics of an individual that are inherited or acquired during early prenatal development', thus presumably incorporating all genetic data as well as family medical history and other genetic related health information. As regards protection of genetic privacy in third party contexts, these draft Regulations are welcomed as explicitly recognising genetic information as a category that deserves protection. This paper will explore the issue of genetic privacy in the EU. It will examine the current data protection framework, as well as the proposed reform of the EU data protection framework and the impact that these reforms will have on the protection of genetic data.

2434T

An evidence-based framework for incidental findings from exome sequencing in the pediatric setting. *E.T. DeChene[1,2], S. Mulchandani[2], M.C. Dulik[1,2], L.K. Conlin[2], J.L. Abrudan[1,2], B.A. Bernhardt[3], K. Izumi[1], S.E. Noon[1], R.E. Pyeritz[3], A. Santani[2], I.F. Slack[1], C.A. Stolle[2], A.B. Wilkens[1], I.D. Krantz[1], N.B. Spinner[2].* 1) Dept of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Dept of Pathology & Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Dept of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Clinicians and laboratories are considering how much secondary information to return to patients undergoing genome-wide tests. The use of exome and genome sequencing has led to an increase in discovery of secondary or incidental findings (IFs), which are clinically relevant genetic variants unrelated to the reason the test was ordered. The issues surrounding genetic IFs in pediatrics are particularly complex, because most children and adolescents cannot fully comprehend implications of potential IFs or independently consent to testing. However, ACMG recommends that laboratories search out and report certain medically actionable IFs, due to the potential impact on the child's or family's health. Funded by the NHGRI's Clinical Sequencing Exploratory Research Program, the Pediatric Genetic Sequencing (PediSeq) Project at The Children's Hospital of Philadelphia (CHOP) and University of Pennsylvania is exploring best methods for performing exome and genome sequencing and returning IFs in the pediatric setting. PediSeq's Oversight Committee (OC) carefully considered the result options offered to families, creating a balance in the principles of beneficence, non-maleficence and patient autonomy. Our current framework allows families to opt to receive results related to medically actionable (MA) childhood-onset disease, MA adult-onset disease, and carrier status for recessive disorders. The OC agreed that certain conditions are immediately MA (IMA) requiring immediate change in treatment or care based purely on genetic diagnosis (in the absence of symptoms). We believe that we are obligated to release IMA results regardless of a family's preferences, due to the likely impact on a child's long-term health. Initially categorization of IFs was established by OC consensus. Due to the subjectivity of this approach, we developed an evidence-based categorization method centered on gene/disease evidence; affected gender and age of onset; availability of treatment or management; and disease severity. Here we present our current IF framework and verify its utility by categorizing over 40 IFs identified in the CHOP CytoGenomics Laboratory. For example, Brugada and several childhood cancer syndromes were categorized as IMA. We propose that carefully categorizing IFs and offering families options for most IFs they receive, along with detailed pre-test counseling, is the best way to balance patient autonomy with our ethical and legal responsibilities to patient care.

2435W

ELSI as applied to non-invasive prenatal testing. *M.W. Leach.* University of Louisville, Louisville, KY.

This year is the 10th anniversary of the completion of the Human Genome Project. ELSI, or Ethical, Legal, and Social Implications Research Program, received a portion of the total project's funding. With the advances and uptake in non-invasive prenatal testing (NIPT), the same justifications for ELSI support funding of research and counseling resources to balance the ethical, legal, and social implications of NIPT. The development of prenatal testing for aneuploidy has received public funding through research grants. Since the introduction of NIPT in 2011, private and public health insurers are now including NIPT in its coverage of prenatal care. Under the Patient Protection and Affordable Care Act (PPACA), prenatal care and maternity care services will be required to be provided as no-cost preventive services and as essential health benefits by private insurers. The investment and coverage of NIPT, however, has been limited to the testing itself. Coverage and inclusion of NIPT as part of prenatal care are premised on professional medical organization guidelines and statements recognizing NIPT as another accepted prenatal testing option to be offered patients. These same guidelines and statements, though, emphasize the need for pre-test and post-test counseling, and the testing manufacturers themselves have stated the need to develop educational materials for patients and providers. The ELSI Research Program was established out of recognition that the Human Genome Project necessarily involved significant ethical, legal, and social implications. So, too, does NIPT. With the lack of risk to the mother coupled with the most accurate screening results yet, testing companies and clinicians are reporting increased uptake in NIPT and a corresponding decrease in uptake of invasive diagnostic testing. Prenatal testing for aneuploidy is premised on informed consent and patient autonomy. But, since its inception, studies have found challenges in achieving informed consent and respect for autonomy based on lack of understanding of prenatal test results and the tested-for conditions. To ensure prenatal testing for genetic conditions and its latest development, NIPT, are administered ethically, legally, and with respect to its social implications, corresponding public research funding and private insurance coverage is justified for research and counseling resources.

2436T

Research Policy of the Genome Science Project in Japan. *J. Minari*[1], *K. Kato*[1,2]. 1) Biomedical Ethics and Public Policy, Osaka University, Osaka, Japan; 2) Institute for Integrated Cell-Material Science (iCeMS), Kyoto University, Kyoto, Japan.

A new large-scale project, the Genome Science Project, started in Japan in 2010. It is a five year project funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The project aims to support various genome researchers in Japan who are funded by MEXT and selected through an annual open call; it does so by providing services of high-throughput DNA sequencing and high-grade information technology. A program in the project, Medical Genome Science Program (MGSP) handles human specimens and sequences the whole genome/exome to support the selected researchers, and thus naturally has to take into account various ethical, legal and social implications. For this, a special group, Research Unit for the ELSI of Genomics, has been established within the project. In starting the MGSP, we had to create a model informed consent form (ICF) template. The process necessarily involved the consideration of major ethical issues of personal genome research, such as returning research results and incidental findings. First, we surveyed existing informed consent documents, and then came up with a draft of the ICF template based on a consideration of the Japanese context. To further refine the draft of the ICF template, we held repeated discussions with executive genome researchers of the MGSP. Through the discussion, we achieved some consensus regarding the research policy of the ICF template. However, for a key issue~returning research results and incidental findings~clear consensus could not be achieved. On this issue, we confirmed that many genome researchers hold the following views: in Japan most personal genome studies aim toward an understanding of the causes and mechanisms of diseases; and the disclosure of research results for clinical purposes was usually beyond the scope of the original aims of research. As a general policy, we decided to leave the decision to each MGSP applicant. In this research activity, we crafted a research policy of personal genome analysis in Japan, in which the cooperation with genome researchers in the MGSP was invaluable for incorporating pragmatic aspects of research. The lessons learned and experiences gained in establishing our research policy can be expected to share a number of similarities with those in other personal genome research and in clinical practice in Japan, and they may serve as a starting point for their policy-making activities.

2437W

Effect of Using a Family History Tool on Communication with Family and Health Care Providers. *C. Wang*[1], *A. Sen*[2], *M. Plegue*[2], *M. Ruffin*[2], *S. O'Neill*[3], *W. Rubinstein*[4], *L. Acheson*[5] *for the Family Healthware Impact Trial group.* 1) Boston University School of Public Health, Boston, MA; 2) University of Michigan, Ann Arbor, MI; 3) Northwestern University, Chicago, IL; 4) National Institutes of Health, Bethesda, MD; 5) Case Western Reserve University, Cleveland, OH.

The family health history offers an ideal proxy to assess genomic risk and is the simplest applied genomic tool available. The Family Healthware™ Impact Trial set out to evaluate the clinical utility of a family history tool developed by the CDC. Primary care patients were randomized to receive usual care with generic prevention messages, or personalized familial risk and prevention messages based on self-recorded family history of common cancers, diabetes, heart disease, and stroke (O'Neill et al., 2009). This study examines the impact of Family Healthware™ on communication behaviors; specifically, communication with family members and health care providers about family health history. Analyses focused on those who had not talked with a family member (2254/3866, 58%) or health care provider (1894/3866, 72%) about their family history during the 6 months preceding study entry. GEE models were run, adjusting for practice clustering and controlling for age, gender and study site. At baseline, communication about family history with either family members or health care providers did not differ between intervention and control groups. At 6 month follow-up, participants randomized to the intervention were more likely to report having talked about their family history to family members compared to those randomized to usual care (42% vs. 37%, p=.052). Not surprising, women were more likely to talk with their family members about family history than men, OR: 1.18, CI (1.01, 1.38). Talking with health care providers about family health history at follow-up was more likely at sites that enrolled patients with upcoming medical appointments, but did not differ between intervention and control groups (44% vs. 41%, n.s.). Additional analyses among Family Healthware™ users showed that women at increased familial risk for heart disease were more likely to discuss their familial risk with family members, as were men at increased familial risk for heart disease or diabetes. There was also evidence of a dose effect among men such that those at elevated familial risk for a greater number of diseases were more likely to communicate with family members. In sum, Family Healthware™ prompted more communication about family history with family members but not health care providers. Future research should explore reasons for not communicating and identify approaches to encourage greater sharing of family history information, particularly with health care providers.

2438T

Mapping the next generation sequencing industry. *M.A. Curnutte*[1], *K.L. Frumovitz*[1], *J. Bollinger*[2], *G.H. Javitt*[3,4], *K.S. Carner*[3], *D. Kaufman*[2], *A.L. McGuire*[1]. 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 2) Genetics & Public Policy Center, Washington, DC; 3) Sidley Austin, LLP, Washington, DC; 4) Berman Institute of Bioethics at Johns Hopkins, Baltimore, MD.

The next generation sequencing industry is rapidly evolving, as companies quickly transition from research services to clinical care. New business models are emerging, with companies offering a variety of services that have policy and regulatory implications. It is important to understand the development of the industry, and to anticipate policy issues that might impede clinical translation. We conducted a comprehensive web-based analysis with the aim of identifying the key business models within the next generation sequencing industry. This includes private companies offering a range of services from sequencing platforms to data analysis and storage. We confirmed our findings with company representatives and experts in the fields of genetics and biotechnology to deepen our understanding of the industry. To date, we have focused on accurately capturing the companies' customer base and practices and services. We present the main business models and discuss the regulatory implications of each. Our preliminary findings show that there is little competition amongst technology providers and increased market expansion in the areas of data analysis, storage, and informatics. As these technologies are incorporated into clinical diagnosis, we question how regulators will respond, and whether the modification of existing legal frameworks will be applicable or adequate. For example, will CLIA regulation apply to the complex analytics necessary to interpret sequence data, and if so are existing regulations appropriately tailored to these new methodologies? Will FDA assert jurisdiction over NGS-based laboratory testing, as it has over 'laboratory developed tests'? We anticipate that NGS may enable the separation of the core phases of clinical laboratory testing~preanalytic, analytic, and postanalytic~ into separate domains performed by different entities, which in turn may necessitate new models of regulation.

2439W

The commercialization of non-invasive prenatal testing: will a private market drive effective clinical translation? *A. Agarwal*[1], *L. Sayres*[1], *M. Cho*[2], *R. Cook-Deegan*[3], *S. Chandrasekharan*[3]. 1) Duke University School of Medicine, Durham, NC; 2) Stanford University Center for Biomedical Ethics, Stanford, CA; 3) Duke University Institute for Genome Sciences and Policy, Durham, NC.

Since its 2001 US market introduction, non-invasive prenatal testing (NIPT) using cell-free fetal DNA found in maternal blood has provided the ability to analyze fetal genomes for genetic conditions without relying on commonly used invasive procedures. Unlike other prenatal tests, NIPT is being primarily driven by private industry, setting precedent for a new market landscape. The prenatal testing field, estimated to be worth up to 1.3 billion US dollars per year, provides strong commercialization incentive as a rapidly growing sector of molecular diagnostics. Already, four US-based companies are offering NIPT for chromosomal aneuploidies as early as ten weeks into pregnancy and plan to expand offerings in the future. Intellectual property rights are playing a key role in the commercialization of NIPT, and raise questions about positive and negative effects on implementation. To address this issue, we assembled a NIPT patent and business landscape, and identified over 100 US patents and applications containing at least one claim to the use of cell-free fetal DNA for prenatal testing. We found that the four major US companies offering NIPT have applied for or received patents, and many of these have been exclusively licensed. All four are currently involved in patent infringement lawsuits, and interference proceedings are ongoing. The effect on availability and future NIPT development will depend heavily on the outcome of this litigation, validity and scope of claims, and licensing strategies. It is possible that litigation related costs could increase prices, decreasing the cost-effectiveness of using NIPT and limit patient access. Also, companies may withhold collected data to secure a market advantage, potentially undermining quality assurance initiatives and decision-making, especially if information about test utility in specific clinical contexts is not freely available. On the other hand, patents may help secure investment to drive future innovation in NIPT technology. While long-term effects are challenging to predict with ongoing litigation, it is necessary for physicians, patients, and other stakeholders to consider the risks and benefits associated with the private sector driving clinical NIPT adoption. As NIPT begins to cover more genetic conditions and becomes a standard of care, it is critical to understand the legal, regulatory, and commercialization aspects of this technology to ensure effective implementation and clinical use.

2440T

Continuity of care of patients with inherited Genetic disorders. *H. Azimi[1,3], S. Ghavimi[2].* 1) Psychogenome, Nepean, Ontario, Canada; 2) Shahid Beheshti Medical University, Tehran, Iran; 3) All Saints Univeristy School of Medicine, Dominica.

Objective: Continuity of care of patients with inherited Genetic disorders. Methods: Survey of patients and physicians was completed by phone, email and in the hospital. Patients answered a series of questions based on the continuity of care they received from their physicians. Physicians in the survey answered a questioner based on how they felt they contributed to the patients overall health by knowing the knowledge of genetics necessary in order to have the best continuity of care for the patients. Results: Of 100 patients surveyed , 34 (34%) with MENIIB syndrome, 11 (11%) diagnosed with Diabetes Mellitus type 2 and 25 (25%) diagnosed with Hemochromatosis. 40 patients (40%) had mixed disorders and their opinion was coupled to make one group. Out of the physicians surveyed, 34 Endocrinologist agreed that having the knowledge of Genetic, and studying their patients diseases allowed them to have a better a relationship with the patients in order to deliver the best continuity of care for them. Primary care physicians surveyed all agreed that cross multi-care for patients in the continuity of care practicum allowed them to better treat and deal with their patient's diseases. All the physicians agreed that being up-to-date with the current genetics advancement they can have the best continuity of care for their patients. 100 Physicians participated in this survey. Conclusions: Knowing more about genetic disorders of our patients, will help to better build a relationship with them. This relationship can help the patient's better cope and live a more productive healthier life. Physicians working as primary care physicians need to be up to date with current genetics advancement, and they need to make sure that the patients they care and treat are aware of genetics basis of their diseases as per this survey indicated.

2441W

Communicating with biobank participants: preferences for receiving aggregate results and providing updates to researchers. *J. Mester[1,4], M. Mercer[2], A. Goldenberg[2], R. Moore[3], C. Eng[1,3,4], R. Sharp[1,2,5].* 1) Genomic Med Inst, Cleveland Clinic, Cleveland, OH; 2) Dept of Bioethics, Cleveland Clinic, Cleveland, OH; 3) Dept of Genetic and Genomic Sciences, Case Western Reserve Univ, Cleveland, OH; 4) Taussig Cancer Inst, Cleveland Clinic, Cleveland, OH; 5) Center for Genetic Research Ethics & Law, Case Western Reserve Univ, Cleveland, OH.

Biobanks collect biological materials and/or health information for ongoing and future biomedical research. Guidelines recommend return of individual research results under certain criteria but do not discuss aggregate results or study updates. Previous work has found that biobank participants desire updates, but preferences as to the extent or manner have yet to be explored. Thus, we surveyed participants in a long-standing protocol-driven research biobank. Eligible participants were drawn from an IRB-approved study of patients with personal/family history suggestive of Cowden syndrome, a poorly-recognized hereditary condition predisposing to breast, thyroid, and other cancers. Participants gave blood samples and access to medical records and had no other interactions with researchers. The biobank had 3,618 participants at time of sampling. Survey eligibility included enrollment ≥18 years and within the biobank's first 5 years, normal *PTEN* analysis, and contiguous United States address. Multivariate logistic regression analyses were performed to identify predictors of participant interest in internet vs. non-internet-based communication modes and communication modes allowing for participant/researcher interaction vs. one-way communication. Independent variables for the regressions included demographic characteristics and attitudinal variables logically or theoretically associated with desired modes of receiving general research updates. Variables were narrowed by independent Pearson correlations by cutoff $p<.2$. Variables with $p<.02$ were considered predictors of preferences. Surveys were returned from 840/1267 (66%) eligible subjects. Most (97%) wanted to receive general updates about the study with 92% wanting updates at least yearly. Participants wished to receive updates via paper (66%) or emailed (62%) newsletter with 95% selecting one of these options. 31% were not comfortable with any e-communication mode, with older, less educated, and lower-income persons strongly preferring offline approaches ($p<0.001$). Most (93%) had no concerns about receiving updates. 97% were willing to provide updates to researchers by post (71%) or email (50%). Plans to communicate results and updates should be made during study development and participants should be able to select if and how to receive them during informed consent. Study demographics may optimize communication modes. Funders should monetarily support the infrastructure needed for these efforts.

2442T

The European Network for Human Congenital Imprinting Disorders - a new COST Action (BM1208). *T. Eggermann[1], E.U. BM1208 members[2].* 1) RWTH Aachen, Aachen, Germany; 2) COST Action.

Imprinting disorders (IDs) are a group of rare congenital diseases affecting growth, development and metabolism with a lifelong impact on patients' quality of life. Despite their common underlying (epi)genetic aetiologies, IDs are usually studied separately by small research groups working in isolation, and the basic pathogenesis and long term clinical consequences of IDs remain largely unknown. Efforts to elucidate the aetiology of IDs are currently fragmented across Europe and standardisation of diagnostic and clinical management is lacking. This COST Action will, for the first time, draw together researchers of all eight known human IDs in an interdisciplinary pan-European and worldwide Network for Human Congenital IDs, working to advance understanding of the pathophysiology with the major aim of translating this knowledge to improvement of diagnostic and clinical management for the benefit of the patients and their families. The Action will harmonise a common ID classification system, develop guidelines for treatment through consensus, create standard operation procedures (SOPs) for diagnosis based on best current practice, coordinate databases held in different countries to make them compatible and useful as a springboard for collective research initiatives, identify new imprinting disorders and causative mechanisms through collaborative effort, educate researchers and stimulate translational exchange. The ID network consists of >30 groups from >15 countries. It will join forces and complement studies to reduce health care costs and increase the life quality of patients suffering from IDs. The Action has started in June 2013, but is still open for further partners.

2443W

The `eugenics' program and public health genomics in China and their implications towards East Asia. *K. Muto[1,2], B. Zhao[2], H. Hong[1,2].* 1) Dept Pub Policy, IMS, Univ Tokyo, Tokyo, Japan; 2) Graduate School of Frontier Sciences, Univ Tokyo, Tokyo, Japan.

Background: China has a high population growth rate, and thus faces great challenges in terms of population growth, health care and economic development. The State launched a family-planning policy in 1980, to control the size and raise 'the general quality of the population'. As part of this family-planning policy, a new 'eugenics' program was developed to ensure the health of mothers and infants, and to improve the quality of the newborn population. As a result of economic development and improvements in medical treatment, the Chinese infant mortality rate dropped from 32 to 12 out of every 1,000 births in one decade. The National Medium- and Long-term Program for Science and Technology Development (2006-2020) set as a priority the acquisition of key technologies to ensure that the country's population remains below 1.5 billion, and its birth defect rate below 3%. Purposes: To facilitate further discussions how both the new bioethics concept and the old disability concept have changed in China, we have looked back and analyzed the first criticisms of the Western science community with regard to the Chinese eugenics program of the 1990s and examine the future impact in East Asia. Methods: Using literature review methods in this study, we collected topical statements from academic societies, government documents, law, and guidelines. We also reviewed articles identified in PubMed and CNKI databases. Results: Western geneticists have fiercely criticized the program as an 'abuse of genetics' and a 'violation of human rights' (Morton, 1998). In contrast, some argue the law would help reduce births of physically or mentally abnormal babies. To look back at the history of the eugenics movement and its ensuing legislation, the term 'eugenics' is defined differently in China than it is in other nations. Discussions: Japan and Korea import non-invasive prenatal testing (NIPT) from the US, even though the two countries have different views and attitudes toward it. At this time, Chinese companies are developing cheaper NIPT through their own original methods. Ideally, NIPT developed in China should be in harmony with other Asian and Western countries, as a basis of a Chinese 'eugenics' program that is balanced ethically and scientifically.

2444T

The attitudes of patients with cystic fibrosis and their parents towards direct-to-consumer genetic testing. *S. Janssens[1], C. Binst[2], I. Mahieu[2], A. De Paepe[1], P. Borry[3].* 1) Center for Medical Genetics , University Hospital Ghent, Ghent, Oost-Vlaanderen, Belgium; 2) Master Life Sciences and Medicine, University of Ghent, Belgium; 3) Center for Biomedical Ethics and Law, University of Leuven, Belgium.

BACKGROUND: Knowledge of carrier status enables prospective parents to make informed reproductive decisions. Recent technological advances in molecular genetics facilitate large-scale population carrier screening. In this prospect reflection is necessary about the desirability to offer community-based (preconception) carrier screening in the healthcare system. At this moment, commercial companies have started to offer preconceptional carrier tests directly to consumers. METHODS: As part of a survey developed to evaluate attitudes towards preconceptional carrier screening for cystic fibrosis (CF), attitudes towards Direct to Consumer testing (DTC) were assessed. The study population was recruited from a register of 157 patients with CF who consult at least once in 3 months at the pneumology department at the University Hospital of Ghent, where one of the eight Belgian reference centers for CF is located. All the parents of CF patients under 18 years and all CF patients aged 16 years and older who attended the clinic in the period between August 13 and December 11 2012 were asked to fill in a questionnaire. An approval from the local research ethics committee was obtained. RESULTS: In total 134 questionnaires were distributed of which 112 were returned. Response rate was 86.7% for the parents and 79.7% for the CF patients aged 16 years and older. 78.3% of the patients with CF and 87.1% of the parents had never heard about testing for carrier status of genetic diseases through the internet. Only 21.7% of the patients and 12.9% of the parents were aware of this possibility. Of the total survey population 57.7% finds that people have the right to order such a test directly at a commercial company. However, 57.1% of respondents disagrees that someone should be able to obtain the results without medical supervision. Only 4.3% of the patients with CF and 9.4% of their parents would buy a carrier test for other conditions then CF through a DTC genetic testing company. DISCUSSION AND CONCLUSION: Awareness about DTC genetic testing for carrier status is low. Although a majority supports the right of individuals to access a carrier test directly, less than half of the respondents supports that individuals would receive test results without a physician. Only a minority would consider buying a carrier test from a DTC genetic testing company.

2445W

Genetic causal beliefs of morbidity: associations with health behaviors and outcome beliefs of behavior changes during two decades in the general population. *A.H. Haukkala[1], N. Hankonen[1], H. Konttinen[1], M. Perola[2], H. Kääriäinen[2], V. Salomaa[2].* 1) Social Reseeach, University of Helsinki, Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland.

Background: The role and meaning of genetic information has increased considerably during the last decades . We examined changes in causal beliefs for morbidity and their associations with health behaviors and control beliefs from 1982 to 2002. Methods: Five population-based risk factor surveys from 1982 to 2002, participants aged 25 to 64 years (n=37,503). Subjects were asked to choose the most important cause for morbidity from the list of ten alternatives. Outcome expectancies of health behaviour changes were assessed with 2 items. Health behaviors included self-reported physical inactivity, current smoking and obesity based on measured height and weight. Findings: Prevalence of genetic causal belief increased from 4% (1982) to 10%(1992) and remained at that level until 2002. Older age groups were more likely to choose genetic causes while there were no associations with gender or education. Lack of exercise and overweight increased while inappropriate diet and stress diminished as causal beliefs for morbidity during the study period. Smokers, physically inactive and obese respondents chose less often physical inactivity, smoking or diet as causes for morbidity and more likely genetic causes for morbidity. Genetic causal belief group had more pessimistic outcome expectations of health behavior changes. These associations remained similar during two decades. Discussion: The proportion of those who choose physical inactivity and overweight as the most important cause for morbidity has increased from 1982 to 2002. Despite increased public discussion about genomics the relative proportion of people with a genetic causal belief has remained low and associations with pessimistic expectancies on health behavior changes have remained the same during two decades.

2446T

A systematic approach to the development of evidence-based family history screening in pediatric primary care. *E. Edelman[1], B.K. Lin[2], N. Mikat-Stevens[3], L. Vasquez[4], K. Hughes[5], J. Scott[1].* 1) NCHPEG, Lutherville, MD; 2) March of Dimes, White Plains, NY; 3) American Academy of Pediatrics, Elk Grove Village, Il; 4) HRSA, Rockville, MD; 5) Harvard Partners, Boston, MA.

Background: While family health history (FHH) is recognized as an important risk assessment and management tool for primary care pediatric providers (PCPs), numerous barriers to its integration have been documented, including limited time, provider knowledge and confidence, and lack of guidelines and tools to support collection and management. Objective: To determine conditions for inclusion on a novel electronic FHH and clinical decision support (CDS) tool for PCPs. Methods: A multi-method approach was used to determine inclusion of conditions for a pediatric FHH tool. Semi-structured interviews focusing on FHH practices were conducted with diverse PCPs (4 pediatricians, 1 family medicine physician, 1 family medicine and 1 pediatric nurse practitioner) recruited from the AAP Genetics in Primary Care Institute. Interviews were analyzed using qualitative thematic analysis. The published literature regarding PCPs' use of FHH and the validity and utility for pediatric and adult conditions was reviewed. Practice guidelines and pediatric FHH forms (13 from pediatric practices and internet searches) were analyzed to determine FHH screening practices. Data on conditions were recorded in a database and analyzed using descriptive statistics. Practice findings and interview themes were summarized to inform condition selection. Criteria adapted from Yoon et al. (2003) were used to determine inclusion and focused on the clinical utility of screening in the general pediatric population. Results: Seventy-six conditions for possible inclusion on the panel were identified. Fifty conditions met criteria, most of which were common conditions such as asthma, intellectual disability, and mental illness. Conclusion: This is the first report to propose an evidence-based approach to determine conditions for pediatric FHH screening. This study presents a model for tool development that can be applied to other institutional or national efforts to develop FHH or genetic CDS tools for diverse clinics. These efforts are increasingly relevant to geneticists as they are called to provide guidance on FHH and genomic data standards in electronic health records and health information technology systems. Next steps include development of risk assessment algorithms and clinical decision support and identification of solutions for interoperability with electronic health records.

2447W

An Assessment of Perceived Medical and Psychosocial Needs of Families that have Children Affected by Duchenne Muscular Dystrophy in Madurai, India. *S. Kejriwal[1], K. Ormond[1], V.S. Arun[2], L. Stanislas[2], C. Siskind[1].* 1) Stanford University, 450 Serra Mall, Stanford, CA 94305; 2) Muscular Dystrophy Foundation India, 26-B, 1st Street, State Bank Colony-II, Bypass Road, Madurai - 625 010, Tamil Nadu, India.

Duchenne muscular dystrophy (DMD) is a hereditary progressive neuromuscular condition that affects approximately 1 in 3500 male children worldwide, and which leads to various medical, psychological and social needs for children and their families. Medical, social and cultural barriers in developing countries are likely to pose additional, unique needs for families. In order to determine these needs, we recorded, transcribed and analyzed 20 interviews with a total of 29 family members of children affected with DMD in Madurai, India. Participants were recruited based on variation in socioeconomic status, literacy, age and number of affected children, and number of deceased children. The main themes that evolved from the interviews centered on issues related to medical care, access and mobility, emotional difficulties and social support. Some aspects of these issues were specific to affected individuals and their families in India, whereas others were also experienced - albeit to a lesser degree - by their counterparts in developed countries. The former consisted of major inadequacies in healthcare, medical communication, psychosocial management and financial support. The latter consisted of underdeveloped solutions to mobility-related difficulties, a lengthy diagnostic process, and a lack of awareness amongst the medical community and society at large. In summary, our study revealed numerous unmet needs of families that have children affected by DMD in Madurai, India. Long-term solutions to unmet needs may include: 1) policy change to improve healthcare funding and mobility infrastructure, 2) development of ancillary services such as genetic counseling and social work, 3) spreading awareness amongst government authorities, the medical community and society at large, and 4) cross-talk between advocacy groups and health professionals. Relatively immediate solutions may include 1) guidance of families towards appropriate medical resources by advocacy groups, 2) increased investment in fundraising activities by advocacy groups to provide relief from financial strain, and 3) collaborations between muscular dystrophy advocacy groups and organizations that are invested in finding solutions to mobility needs of the disability community at large.

2448T

Reporting of incidental copy number variation (CNV) detected by chromosomal microarray analysis in `normal' parents/family members - A review of 3500 cases. *A. Patel, S.W. Cheung, P. Stankiewicz, A. Breman, S.R. Lalani, J. Smith, C. Shaw, S. Peacock, A. Braxton, L. Ellis, P. Ward, J.R. Lupski, A. Beaudet, W. Bi.* Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX.

When interrogating the human genome, it is sometimes the case that the most significant finding is unrelated to the referring diagnosis. Such incidental findings can pose challenges with regard to which results should be revealed, to whom and by whom. Such issues have begun to be addressed with the recent ACMG recommendations for reporting incidental findings in Clinical whole Exome and Genome sequencing. While some incidental findings are currently of uncertain clinical significance, others are medically actionable and reporting of these incidental copy number variations (CNVs) can have a significant impact on risk counseling for future pregnancies and other family members at risk, as well as enabling interventions to improve health outcomes. In the course of performing chromosomal microarray analysis (CMA) on 3500 apparently 'normal' parents/family members for interpretation of variants of unknown clinical significance in the proband, we incidentally found 28 (0.8%) clinically relevant CNVs including [1] mosaic or full chromosome aneuploides, [2] genomic disorders with incomplete penetrance, [3] carriers of X-linked Mendelian traits, [4] carriers of autosomal recessive disorders, [5] cancer susceptibility and [6] dominant adult onset disorders. Reporting of incidental findings is not new to medicine, so should reporting incidental findings in whole genome analysis be any different? Should the current ACMG recommendations for reporting incidental findings incorporate incidental CNVs?

2449W

The Hospital for Sick Children Genome Clinic: Developing and evaluating a pediatric model for individualized genomic medicine. *M.S. Meyn[1,2,3,4,5], S. Bowdin[2,3,5], N. Monfared[2], D. Merico[1,6], D.J. Stavropoulos[7,9], M. Girdea[1,8], R. Hayeems[3], T. Stockley[7,9], M. Szego[6,10,11,12], G.D. Bader[4,13], R.D. Cohn[1,2,3,5], C.R. Marshall[1,6,11], R. Zlotnik Shaul[5,10,14], M. Brudno[1,8], C. Shuman[1,3,4], P. Ray[1,2,4,7].* 1) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 3) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Paediatric Laboratory Medicine, Hospital for Sick Children, Toronto, ON, Canada; 8) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 9) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada; 10) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 11) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 12) Centre for Clinical Ethics; Providence Healthcare, St. Joseph's Health Centre and St. Michael's Hospital; Toronto, ON, Canada; 13) The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 14) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada.

The transformative potential of whole genome sequencing (WGS) lies in its use in individualized genomic medicine, where knowledge of an individual's genomic variants guides health care decisions throughout life. However, there are multiple challenges: Many disease genes have yet to be discovered; bioinformatic tools are still under development, the diagnostic advantages of WGS are largely unproven; the clinical utility and socioeconomic impact of identifying incidental medically actionable genomic variants (MAVs) in children are unknown; and existing counselling models are inadequate for clinical genomics.

To address these and other knowledge gaps we have created a paediatric 'Genome Clinic.' This multi-disciplinary research platform has several distinctive features, including: a) The project treats an individual's genome as a source of genetic information to be repeatedly queried over time in order to manage overall health, rather than a one time test performed to diagnose a pre-existing disorder. As a result, MAVs are a major focus of study rather than an inconvenient burden. b) Focusing on children offers the greatest potential for increased quality of life years. c) Project scope is broad - from development of new bioinformatics tools and discovery of new genes to health policy inquiries, assessment of models of clinical care and the ethics of consent. d) Project design emphasizes an interactive partnership between patients, parents and health care researchers in discovering the medical and psychological risks and benefits of genomic medicine. e) The Genome Clinic offers WES/WGS-based clinical research projects a 'safe harbor' for managing incidental findings.

Piloting our bioinformatics pipeline using a curated list of 2000+ disease genes found 5-13 MAVs and 5-18 carrier variants/individual requiring manual evaluation. We are enrolling 50+ children/year who are under investigation for a genetic disorder, along with their parents. Participants are counselled regarding WGS and given options of learning about specific classes of adult-onset MAVs and carrier variants. Variants relevant to the primary disorder are returned by the referring physician, while pharmocogenomic variants, MAVs, and carrier variants are disclosed to participants by Genome Clinic geneticists and genetic counsellors. Our systematic approach mitigates potential risks and maximizes benefits of this new clinical care model.

2450T

Frequency of ACMG recommended 57 gene incidental findings from whole exome sequencing in a cohort of 47 adult individuals. *J. Wynn[1], M.L. Cremona[1], J. Martinez[1], Y.H. Cheung[1], W.K. Chung[1,2].* 1) Department of Pediatrics, Columbia University Medical Center; 2) Department of Medicine, Columbia University Medical Center.

The American College of Medical Genetics (ACMG) recently recommended mandatory return of secondary findings for 57 genes in patients receiving clinical whole exome and whole genome sequencing (WES/WGS). Before implementation, the frequency of mutations in these genes in an unselected cohort should be determined to appreciate the impact of this policy on clinical practice. We completed WES on 47 adult individuals and analyzed results for secondary findings in the 57 ACMG genes. Variants in Human Gene Mutation Database were identified, and any variants with a minor allele frequency > 2% in the 1000 Genomes and Exome Variant Server were removed. The literature for all remaining variants was reviewed and only Class I variants using ACMG criteria were classified as mutations. Five participants had previously had negative genetic testing of the BRCA1 and BRCA2 genes. For the other 42 subjects, the indication for WES was unrelated to the conditions on the ACMG list. A published pathogenic mutation in one of the 57 genes was identified in 13% (6/47) of participants. We identified 3 (6.4%) participants with SCN5A mutations; 2 of the 3 participants had no personal or family history suggestive of Long QT or Brugada syndrome. One participant had a history of two relatives who passed away suddenly <50 years. We identified 2 participants (4.3%) with mutations in cardiomyopathy genes (MYBPC3 and MYH7). One participant had recently had an echocardiogram showing mild concentric left ventricular hypertrophy with an intraventricular septal wall thickness of 1.3cm (normal 0.6-1.1). The other participant was asymptomatic and had never had an echocardiogram. One subject had a mutation in the MSH6 gene. She had never been diagnosed with cancer but her mother had a history of colon and uterine cancer .We found a frequency of secondary findings in 13% of adult participants for conditions that are believed to be rare. Few participants had a personal or family history of the condition identified by the secondary finding. These findings suggest problems with the accuracy of mutation classifications in the literature and/or inflated estimates of the penetrance for these conditions in an unselected population. Our results indicate there will be a significant burden to the clinical community to return secondary findings for these 57 genes, and perhaps implementation of any policy to return secondary findings should await curation of the mutation databases and genetic literature.

2451W

Incidental Findings and the ACMG Guidelines: What is the Real Burden? *J.J. Johnston[1], D. Ng[1], S.G. Gonsalves[1], K.L. Lewis[1], D.N. Cooper[2], J. Berg[3], H.L. Rehm[4], L.G. Biesecker[1].* 1) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 2) Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK; 3) Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA; 4) Partners Healthcare Center for Personalized Genetic Medicine, Boston, Massachusetts, USA.

With the ever-increasing use of exome sequence analysis in both research and the clinic, the issue of how to handle incidental findings has become controversial. The American College of Medical Genetics and Genomics (ACMG) has published guidelines detailing a list of genes/diseases that should be analyzed in all clinical exomes, with clinically relevant results being returned to the patient. We set out to assess the feasibility of these guidelines using the ClinSeq® dataset. In all, 951 exomes were analyzed for missense, nonsense, splice site, and insertion/deletion variants in the ACMG-suggested list of 57 genes. A total of 1,340 unique variants were identified. Filtering on a minor allele frequency of <0.01 reduced the list to 1,214 variants. Clinical labs will need to rely on published literature in order to annotate relevant findings. The ACMG guidelines recommend the use of curated databases to identify previously annotated variants followed by thorough re-evaluation of evidence for causation. Current variation databases include the Human Gene Mutation Database (HGMD), individual Locus-Specific Databases (LSDBs) and ClinVar. A total of 372 identified variants were annotated in the subscription version of HGMD (HGMD Professional) and an additional 164 variants were annotated in available LSDBs. Of these 536 variants, 211 were predicted to be causative for a disease on the ACMG disease list in at least one database. Further manual curation reduced the list of returnable causative variants to 27. Four variants were identified in multiple individuals bringing the total number of variants to return to 33. Eleven individuals had variants in *LDLR* or *APOB*, probably representing a bias of the ClinSeq® cohort selected for cardiovascular disease. Identifying 33 variants in our cohort of 951 individuals suggests a burden of approximately 3.5%; removing the *LDLR* and *APOB* variants reduces that burden to 2.3%. However, as current databases vary quite widely in terms of both quality and completeness, the real burden lies in the annotation of these variants. Although the identification of variants in certain genes can contribute important health information to patients, the current state of the field is such that making sense of these variants is both challenging and time consuming. It is clear that improvements in the annotation of mutation database are urgently required to address this problem.

2452T

An assessment of the prevalence of rare nonsynonymous variants within the genes listed in the ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. *S.P. Strom[1], H. Lee[1], J.L. Deignan[1], K. Das[1], E. Vilain[2,3], W.W. Grody[1,2,3], S.F. Nelson[1,2,3].* 1) Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 2) Human Genetics, UCLA, Los Angeles, CA; 3) Pediatrics, UCLA, Los Angeles, CA.

BACKGROUND: The American College of Medical Genetics recently released guidelines which recommend the reporting of incidental findings (IF) when performing clinical exome or genome sequencing. Specifically, the ACMG has stated that laboratories performing this type of testing should report ‘known pathogenic’ and ‘expected pathogenic’ variants within a list of 57 genes, all but three of which act in an autosomal dominant fashion. In these guidelines, the ACMG has estimated that 1% of individuals will carry a known or expected pathogenic variant in any one of these 57 genes. However, as no clinical grade database currently exists to assess which variants in each of these genes are known to be pathogenic, variant interpretation falls to each individual laboratory. Although guidelines state ‘these recommendations should not be construed as an expectation that the laboratory comprehensively assess these genes for all variants,’ compliance without such analysis requires currently unavailable variant annotation. METHODS: We analyzed prior clinical exome sequencing results at UCLA from 226 de-identified individuals (either index or parental cases) to identify all high quality ($Q \geq 500$) potentially pathogenic variants in any of the 57 IF genes. We define potentially pathogenic as: nonsynonymous variants with a population allele frequency <1%. RESULTS AND CONCLUSIONS: Contrary to the a priori estimate in the ACMG guidelines, we identified at least one potentially pathogenic variant within an IF gene in 184 out of 226 individuals (81%). 36% of variants identified are located at an HGMD annotated locus. A total of 294 unique variant loci, or approximately 1.3 per individual on average, were identified. As the combined disease prevalence of all IF gene disorders is far less than 81%, the majority of these variants are likely not clinically significant. However, identifying the subset of variants that are potentially significant represents an additional interpretive burden for the laboratory. Until clinical grade variant databases exist for all of these 57 genes, laboratories may need to report all potentially pathogenic variants in the IF genes rather than only reporting a select few, if the ‘fiduciary duty to prevent harm’ aspired in the guidelines is to be met. For now, laboratories attempting to implement these guidelines should work with ordering physicians to coordinate an appropriate approach for discussing incidental findings with each patient or family.

2453W

Do preferences matter? Creating and assessing a novel preference-setting tool for the return of genomic research results. *P.L. Bacon[1], S.K. Savage[1], S.I. Ziniel[2,8,9], K.D. Christensen[5], N.L. Huntington[2,7], E.R. Weitzman[2,4,8], P.L. Taylor[2,4], R.C. Green[2,5,6], I.A. Holm[1,2,3].* 1) Program in Genomics, Boston Children's Hospital, Boston, MA; 2) Department of Pediatrics, Harvard Medical School, Boston, MA; 3) The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 4) Children's Hospital Informatics Program, Boston Children's Hospital, Boston, MA; 5) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States; 6) Partners Center for Personalized Genetic Medicine, Boston, MA; 7) Division of Developmental Medicine, Boston Children's Hospital, Boston, MA; 8) Division of Adolescent Medicine, Boston Children's Hospital, Boston, MA; 9) Program for Patient Safety and Quality, Boston Children's Hospital, Boston, MA.

Whether and how research participants receive individual research results (IRRs) from genomic studies has generated great concern, and many argue that participants should be allowed to designate their preferences for IRR return. The goal of this study was to develop an online preference-setting tool for the return of IRR and to determine if parents anticipate greater benefit and satisfaction from enrolling their child in biobanks where they could set preferences for IRR versus biobanks where they could not. The preference setting tool was informed by a large survey conducted at Boston Children’s Hospital (BCH) assessing parents’ interest in enrolling themselves and their children in a hypothetical genetic research repository and preferences for return of IRR. The tool was further refined through cognitive interviews with parents of BCH patients to understand what dimensions of a condition mattered most with parents as they decided if they were interested in receiving a genetic result. The resulting preference setting model is based on a) the preventability of the condition and b) the severity of the condition. Parents may also opt-out of any of four categories of results: mental illness/psychological conditions; developmental disorders/learning disabilities; childhood-onset degenerative conditions; and adult-onset conditions for which there are no interventions during childhood. We then developed an online survey in which participants were randomized to be enrolled into one of four hypothetical biobanks with different results return policies: 1) participant receives no IRRs; 2) participant receives all IRRs; 3) participant chooses to receive all or no IRRs (‘binary choice’); and 4) participant sets preferences for return of IRR using the online tool (‘granular choice’). We pilot tested the survey with parents of children seen at BCH. Preliminary data suggests that, when given a choice, the majority of parents elect to receive all possible IRRs about their children. Those parents who chose to receive all IRRs using the preference-setting tool rated the highest satisfaction of the 4 groups. In addition, the ‘granular choice’ group had the greatest proportion of parents stating that they would participate in a biobank that resembled the one they were randomized to. Preliminary findings will be confirmed in a forthcoming survey of 20,000 parents of patients at BCH.

2454T

Biobank Participants' Perspectives on Aggregate Result Return. *E. Bane[1], E. Ludman[2], J. Richards[2], G. Jarvik[3], S.M. Fullerton[4].* 1) Institute for Public Health Genetics, University of Washington, seattle, WA; 2) Group Health Research Institute, Seattle, WA; 3) Medical Genetics, University of Washington, Seattle, WA; 4) Bioethics and Humanities, University of Washington, Seattle, WA.

Introduction: Researchers and policy makers have recommended returning, where feasible, aggregate (or summary) results of research conducted using biospecimens to biobank participants (Beskow et al. 2012). There is, however, lack of consensus on best practices for aggregate return, particularly about how and when to communicate results; participants' perspectives on such issues would be valuable to incorporate into policy and practice. To date there have been few explorations of participant preferences with respect to the receipt of aggregate results and even fewer in the context of biobank participation that may involve genetic investigation. Here, we report a preliminary exploration of participant preferences and perspectives on the necessity for, and mode of, returning aggregate results of genomic biorepository research. **Methods**: Fifteen semi-structured interviews were conducted with participants in the Northwest Institute of Genetic Medicine (NWIGM) biorepository at the Group Health Cooperative and University of Washington in Seattle, Washington. NWIGM participants were interviewed by telephone about their perspectives on, and preferences for, return of results from research using human tissues. Interviews were recorded and de-identified transcripts subjected to a directed content analysis. **Results**: The participants interviewed reported a desire for individual results, and expressed a clear preference for allocating funds designated for return of summary results to additional research. In the event that aggregate results were to be returned, these participants preferred to receive them in the form of emails, letters, or websites rather than in-person interviews or phone calls. **Conclusions**: The results of this preliminary research indicate that while participants of genomic biorepository research are interested in research results, participants view individual results as more relevant than aggregate results, and are more invested in the objectives and fruition of the research itself than in receiving results per se. These results suggest individuals should be consulted regarding preferences prior to formation of overarching policy regarding aggregate return, and indicate that further research with larger and more representative cohorts is warranted.

2455W

Scripps Idiopathic Diseases of Man Study: The First 2 Years. *C.S. Bloss[1], S. Topol[1], B.F. Darst[1], A. Torkamani[1,2], N.J. Schork[1,2], E.J. Topol[1,2,3].* 1) Scripps Genomic Medicine, Scripps Translational Science Institute, La Jolla, CA; 2) Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA; 3) Scripps Clinic Medical Group, La Jolla, CA.

Introduction: The Scripps Idiopathic Diseases of Man (IDIOM) study aims to facilitate diagnosis and treatment of individuals with rare or idiopathic diseases through the use of whole genome sequencing (WGS) integrated with clinical assessment and multidisciplinary case review. We will describe the IDIOM study, including the operational protocol and descriptive statistics for the first, nearly 100, patient referrals. We will also present data on return of results for cases enrolled to date, including patient and provider preferences for the types of results returned from WGS, as well as perceptions of the return-of-results communication itself. Methods: IDIOM patient inquiries that pass initial screening are forwarded for review by a 12-person scientist-clinician panel that includes physicians, bioinformaticians, sequencing experts, ethicists, and nurses. Criteria include the presence of a grave condition that has remained undiagnosed despite extensive medical follow-up. For selected cases, WGS results are delivered back to the proband's referring physician, who is then responsible for returning the results to the patient. Patient and physician attitudes, preferences, and perceptions of the physician-patient return-of-results communication are assessed via semi-structured interviews and standardized questionnaires. Results: To date, 85 patient referrals for IDIOM have been received and screened, 30 have undergone 2nd tier review by our clinician-scientist panel, and 9 patients and their families have been enrolled. With respect to return of results, baseline data indicate that the majority of patients would prefer to receive sequencing results for the presenting condition along with secondary findings in multiple disease categories. Data from the completed return-of-results sessions suggest notable variation in patient satisfaction with communication of results by the referring physician. Case-by-case findings will be discussed in detail. Conclusion: The Scripps IDIOM study is a fully functioning program at the interface of clinical care and research, focused on the recruitment, screening, clinical evaluation, and WGS of individuals with undiagnosed diseases. We hope that results from this study will inform the effective, sustainable, and cost-effective use of WGS for rare/unknown disease diagnosis in community hospital settings by practicing physicians.

2456T

Should Secondary Findings From Whole Exome And Whole Genome Sequencing Be Released To Research Subjects? Our Ethical Responsibility. *C.A. Campbell[1,2], T. Bair[1], D. Kolbe[1], R.J.H. Smith[1,2].* 1) Iowa Institute of Human Genetics, Iowa City, IA, USA; 2) Department of Otolaryngology - Head and Neck Surgery, University of Iowa, Iowa City, IA, USA.

Whole exome (WES) and whole genome sequencing (WGS) are powerful tools that have facilitated the identification of multiple genetic causes of human disease. Genetic variants identified incidental to this primary discovery process are referred to as secondary findings. The American College of Medical Genetics (ACMG) has recently recommended that when WES and/or WGS are requested for clinical diagnostic purposes, the final variant report should include secondary findings in a defined set of 57 disease genes. These 'Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing' target genes primarily associated with hereditary cancer and cardiovascular disease since therapeutic intervention and/or screening make disease treatment and/or prevention possible. Analysis of data from the NHLBI Exome Sequencing Project on the Exome Variant Server (EVS) shows that there are 8,483 variants in European-American exomes in these genes. We will present the likelihood of identifying an ACMG-reportable variant in a research exome. As researchers, we must address this risk and the ACMG guidelines if we are to enroll patients with appropriate informed consent in genetic studies in which WES and/or WGS is planned. As the detection of variants depends on data quality, a summary of data quality for these genes using the SureSelectXT Human All Exon V4 (Agilent) capture and sequencing performed using paired-end chemistry on the HiSeq 2000 platform (Illumina, San Diego, CA) will be presented.

In the rapidly changing landscape of human genetics researchers are confronted with the dilemmas of; if, how, and when to give back research results to study participants. We will discuss the unique issues surrounding implementation of these guidelines in a research setting. In addition, the role of the genetic counselor in a research laboratory performing WES and/or WGS will be described. We propose select language in the informed consent document to deal with various reporting options.

2457W

`We don't know her history, her background': Adoptive parents' perspectives about sequencing results. *J. Crouch[1], J. Yu[2], A.G. Shankar[2], H.K. Tabor[1,2].* 1) Seattle Children's Research Institute, Seattle, WA; 2) University of Washington, Department of Pediatrics, Seattle, WA.

Pediatric exome and whole genome sequencing (ES/WGS) can provide parents with a wide range of genetic information about their children. Adoptive parents may especially benefit from ES/WGS results because they often lack information about their children's family health history. There are no data on adoptive parent perspectives about pediatric ES/WGS information, including what kinds of results they want, potential benefits and risks to their families, and how to respect the developing and future autonomy of their adoptive children. We conducted four focus groups with parents of adopted children (n=26) about WGS. Participants were predominantly female (73%) and European-American (77%), with a mean age of 49 years old (range 35-66). The majority of parents (65%) adopted at least one of their children internationally and 42% had two or more adopted children. Adoptive parents viewed WGS results as potentially beneficial for providing unknown medical information about their children; specifically, information that parents often know about their biological children such as family history of disease. Parents hoped that genetic information could help guide prevention, treatment and management of existing health concerns and diagnoses for their children. Genetic information about ancestry was important to many adoptive parents, particularly those whose children were adopted internationally, to provide a greater sense of identity in the absence of information about family and cultural origins. In fact, several parents described using and receiving targeted marketing from direct-to-consumer companies for genetic ancestry testing of their adopted children. Many parents cited concerns about their children's privacy, including potentially increased privacy risks for adopted children. Most parents wanted to defer disclosure of genetic information about late onset disease and reproductive risks in order to respect their children's future autonomy. These results suggest that adoptive parents want to receive WGS results in order to learn ancestry and medical information that biological parents already have, and that tailored policies and tools may be needed for offering and returning ES/WGS results in this population.

2458T

Human genetic researchers and biobank leaders support the return of high-risk, actionable research findings but face numerous impediments to responsible return of results. *R. Dvoskin[1], J. Bollinger[1], K. Kreger[1], A.A. Padon[1], K.L. Edwards[2], D. Kaufman[1].* 1) Genetics and Public Policy Center, Berman Institute of Bioethics, Johns Hopkins University, Washington, DC; 2) University of Washington, School of Public Health, Department of Epidemiology, Seattle, WA.

Background: Biobanks and genetic cohort studies are important tools in the search for genetic and environmental factors underlying a broad range of diseases. The data generated by these resources may include genetic information of potential significance to study participants. Whether, and which, individual research results should be returned are under heated debate. As policies and recommendations evolve, it is useful to consider the opinions and practices of the research and biobank communities. **Methods**: 59 U.S.-based genetic researchers and biobank leaders were interviewed about human subjects issues including the return of research results. Semi-structured interviews were transcribed, coded, and analyzed for themes. Reasons for and against returning results were summarized. **Results**: A majority of interviewees were not opposed in principle to the idea of returning some medically useful research results. Interviewees raised a number of arguments against their return, however, most of which fell into five categories: practical and logistical reasons; the uncertain quality and meaning of the data; excessive cost and personnel burdens; the risk of doing more harm than good; and the lack of guidance regarding researcher and biobank obligations. Two common reasons for returning results were a perceived duty to disclose information that could benefit someone's health, and participants' right to their own information. Additional reasons included participants' desire for results, improving trust in the research community, and rewarding participation. Many believed that any results returned should be actionable, validated, and indicate a high risk for disease. Researchers' opinions and decisions about returning results are also dependent on the relationship between an investigator and the research participants and whether participants are clinical patients or healthy volunteers. **Conclusions**: Concerns about uncertain or unconfirmed results and ambiguity about researchers' obligations may influence decisions to return results. In addition to providing guidance on what results to return and when, recommendations should consider the importance of contextual factors and address the practical, financial, and ethical challenges to disclosing research results. Researchers' support for the return only of well-understood genetic findings and for transparency about the quality and meaning of research data agree with public preferences observed in some previous studies.

2459W

Focus group exploration of return of research results. *W.A. Faucett[1], S.N. Fetterolf[2], L.H. Rogers[1], S.A. Martin[1], M.S. Williams[1], D.J. Carey[2].* 1) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 2) Geisinger Health System, Danville, PA.

A goal of the eMerge project was to develop methods to return to biobank participants clinically confirmed (CLIA) research results. Geisinger Health System currently has three active biobanks with 43,000+ participants including MyCode, a general population biobank; AAA, a biobank for abdominal aortic aneurysm families; and NASH, a biobank for bariatric surgery patients. Biobank consents allow for re-contact but none of the biobank consents currently allow research results to be placed in the electronic medical record. Geisinger is considering a model in which the expectation would be to return research results. The project goal was to determine the best approaches to return results. Biobank participants were contacted via opt-in letters to participate in focus groups held in their local clinic offices. Six patient focus groups in rural Pennsylvania towns have been held since November 2012 with a total of 93 participants (57% female, 43% male). Of these participants, most were 61-70 years old (~42%; >71: ~22%; 51-60: ~23%; 41-50: ~8%; 18-40: ~5%). Educational backgrounds varied(college degrees: ~33%; beyond high school: 28%; high school diploma/GED: ~34%; less than high school: 4%). Nearly half of all focus group participants have been Geisinger patients for at least 20 years (49.5%), and ~85% received all or most care from Geisinger. The group discussions included five areas related to results: pharmacogenomics; recessive carriers; increased risks for preventable or treatable diseases (e.g. colon cancer); increased risk for diseases that cannot be prevented or treated (e.g. Alzheimer's disease); and variants of uncertain significance (VUS). The overall consensus of focus group participants is that they would want to have all of their research results returned no matter the severity of the disease or clinical impact. Most felt they would want to know about a VUS. The vast majority of participants wanted the results returned to them and to their healthcare provider at the same time. They stated that educational materials should be developed for participants and healthcare providers as results are returned. Most participants approved of the results being documented in their Electronic Medical Records. Results of the participant focus groups will be used to develop surveys for all biobank participants on the return of research results and focus groups consisting of Geisinger healthcare providers who may potentially be returning these research results.

2460T

Pleiotropy and the potential return of (additional) incidental information with incidental result return. *S.M. Fullerton[1], J.M. Kocarnik[2].* 1) Bioethics and Humanities, University of Washington, Seattle, WA; 2) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.

Introduction: Recent discussion regarding the return of incidental findings generated in the course of whole genome sequencing has focused on returning only information that meets stringent thresholds of clinical validity and utility. Most recently, an American College of Medical Genetics and Genomics (ACMG) statement recommended that variants in 57 genes associated with 24 phenotypes are sufficiently important that they should be sought out and reported when any clinical sequencing is performed (Green et al. 2013). A potential problem with such recommendations, however, is that they do not address the issue of pleiotropy, where a single gene is associated with multiple phenotypes that may differ in their appropriateness for return. For example, one genotype-phenotype relationship may be clinically actionable while another might be of unproven validity. Since the genetic information is the same, returning information on the former also returns information on the latter. To estimate the potential extent of this issue, we sought to evaluate how many of the genes in the ACMG recommendation list might be associated with additional phenotypes. **Methods:** We searched the Online Mendelian Inheritance in Man catalog, a publicly accessible database of human genes and genetic disorders (OMIM.org). We performed a search on each of the 57 MIM genes listed by the 2013 ACMG recommendations, and recorded how many and what kind of MIM phenotypes were noted as associated with variants in each gene. **Results:** As of 5/31/2013, the majority (>75%) of these genes were linked to more than one MIM phenotype, with a few genes associated with greater than 10 phenotypes. However, the ACMG recommendations typically only identified one of these phenotypes as the rationale to search out and report variants in these genes as incidental findings. **Conclusion:** Most of the genes in the ACMG list of recommendations have known pleiotropic relationships with phenotypes beyond those listed, suggesting that additional information about genetic associations may be conveyed regardless of clinical utility. No guidance is provided for how to deal with this extraneous yet readily-obtainable information. Pleiotropy appears to be a pervasive issue that has been under-considered in current incidental findings guidelines. Future guidelines will need to consider the potential impact of such pleiotropic relationships on whether, and if so how best, to return individual genetic results.

2461W

Reasons associated with different levels of interest in receiving individual genetic research results: a public survey. *L. Jamal[1,2], J. Bollinger[1], R. Dvoskin[1], D. Kaufman[1].* 1) Genetics and Public Policy Center, Johns Hopkins University, Washington D.C; 2) Kennedy Krieger Institute, Baltimore, MD.

Objectives: To a) differentiate between the individual research results (IRRs) that participants would like to receive back from genomic research study and those they would be merely satisfied with, and b) determine the relative importance of clinical and personal utility (reasons why) for different demographic groups. **Background:** Despite guidelines for reporting IRRs to participants in genomic studies, whether and how to return them remains a matter of researcher and institutional review board (IRB) discretion. Few researchers have practical experience selecting results to return and deciding how to do so with limited time and resources. **Methods:** An online survey of a random national sample of 1,474 U.S. adults, fielded in both English and Spanish, measured public attitudes about receiving and using individual research results collected or discovered as part of a proposed national cohort study. **Results:** Given a choice, 55% of respondents wanted all of their individual research results, while 33% preferred only results pertaining to preventable or treatable ('actionable') conditions and 12% expressed no desire to receive IRRs. The same fraction wanting IRRs was observed across all demographic and racial groups. The most important reason participants wanted to receive their IRRs was clinical utility, but majorities of those wanting some and all results saw value in sharing the information with family members and gathering information on implicated diseases. Half of respondents who wanted all their results said so 'because they belong to' them, compared to 25% of those preferring only actionable information ($p = 0.05$). The most common reasons people would not want any results were lack of interest and a perception that returning results was not the purpose of research. Few significant demographic differences in attitudes were noted. **Conclusions:** Nearly half (45%) of the public does not want to receive all research results, given a choice. Participants who desire IRRs are primarily motivated by an interest in information with clinical utility. Personal utility was also an important factor, and an acute sense of fairness seems to be a key reason why some individuals want all results instead of just actionable ones.

2462T

User satisfaction with a web-based tool for self-guided management of results from ES/WGS. *S.M. Jamal[1], J. Yu[1], J. Crouch[2], A.A. Lemke[2], M.J. McMillin[1], A.G. Shankar[1], K.M. Dent[3], M.J. Bamshad[1,4,5], H.K. Tabor[1,2].* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Seattle Children's Research Institute, Seattle, WA; 3) Department of Pediatrics, University of Utah, Salt Lake City, UT; 4) Seattle Children's Hospital; 5) Department of Genome Sciences, University of Washington, Seattle, WA.

A major obstacle to taking full advantage of exome and genome sequencing (ES/WGS) data in a clinical setting is how to effectively manage incidental or secondary results. Use of traditional approaches (e.g., face-to-face interview with a provider) for each result that could be offered for return is simply untenable. Over the past several years, we have developed a return of result strategy based self-guided management of results using a web-based tool called My46 (http://www.my46.org). My46 enables users to set preferences for which results they want to receive, to change their preferences over time, to review standardized information about each trait for which a result is available, and to receive each result privately and conveniently. The usability of such a tool partly depends on whether individuals can successfully complete each task needed to manage results return and whether they are comfortable doing so. We tested the usability and satisfaction of setting ES/WGS result preferences with My46 in adults (n=47) with Mendelian or complex diseases enrolled in ES projects. Participants were primarily European American (94%), female (83%), and 26 to 81 years old (mean=44). The mean composite scores for a modified Computer System Usability Questionnaire were highly positive, on a scale of 1 to 5 with 5 being most favorable: usefulness=4.68 (3.68-5.00), information quality=4.58 (3.14-5.00) and interface quality=4.46 (3.00-5.00), with an overall score of 4.61 (3.68-5.00). The majority of participants (96%) indicated it was simple to use My46, 91% found the information provided in My46 easy to understand, and 81% indicated My46 had all the functions and capabilities they expected it to have. Most participants (75%) indicated they were satisfied with the use of My46 as a way to select preferences, thought My46 had the right kind of information to aid in decision making, and would recommend My46 to someone else for managing genetic results. A subset of individuals (n=21) were interviewed after setting their preferences, and all reported that the site was easy to use, simple, thorough and user-friendly. These individuals valued that setting their preferences using My46 was fast and efficient, though some expressed concern that use of confirmation 'pop-ups' was unnecessary. These results suggest that usability and satisfaction with My46 are high, and underscore the need for further development and testing of web-based approaches for ES/WGS results management.

2463W

The Industry Pharmacogenomics Working Group (I-PWG) Perspective on Providing Individual Research Results and Incidental Findings to Clinical Trial Research Participants. *S.K. Prucka[1], L.J. Arnold[2], J.E. Brandt[1], S. Gilardi[3], L.C. Harty[2], F. Hong[4], J.S. Malia[5], D.J. Pulford[6].* 1) Tailored Therapeutics Genetics, Eli Lilly and Company, Indianapolis, IN; 2) Pfizer Worldwide Inc., 445 Eastern Point Road, Mail Stop 8260-2143, Groton, CT; 3) Bristol Myers Squibb - Hopewell, Attn: Terrye DelMonte, Clinical Genetics Operations, Biomarker Technologies, Blg 3A-209A, 311 Pennington-Rocky Hill Rd, Pennington, NJ; 4) Amgen Inc., One Amgen Center Drive, M/S 38-3-B, Thousand Oaks, CA; 5) Purdue Pharma LP, One Stamford Forum, Stamford, CT; 6) GlaxoSmithKline R&D, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom.

The pharmaceutical industry has not been a prominent figure in recent discussions regarding the return of individual genetic research results and incidental findings. The Industry Pharmacogenomics Working Group (I-PWG) however, is well-positioned to comment on issues related to providing genetic research results that are generated from clinical trials. The I-PWG consists of representatives from 20 pharmaceutical companies engaged in pharmacogenomics research as part of product development. Industry-sponsored genetic research is typically directed at understanding the safety and efficacy of the therapeutic compound under development and researching the genetics of disease to better define subgroups that are the most likely to receive a benefit. This research progresses alongside the development of a therapeutic compound from pre-clinical work through clinical trials and often continues with post-marketing studies after product is launched on the market. The majority of this research is exploratory in nature and many years may pass before the clinical implications can be confirmed and validated. Without clear clinical relevance, the benefit of providing genetic research results to participants is unclear. Further, given the exploratory nature of this research, many of the assays employed may not yet have been fully validated or meet the local requirements for sharing with research participants. When results are provided, the Investigator at each clinical trial site is responsible for delivering this information. Investigators' varying levels of genetic literacy can make it challenging to communicate the information required for a thorough understanding of both the research opportunity and associated results, especially given that the focus of the Investigator/patient relationship is typically on the overall trial and not pharmacogenomic research opportunities. In cases where the research is conducted many years after the clinical trial is closed, communicating results may be difficult given the challenges of maintaining contact with Investigators and research participants (particularly for global trials). Here we plan to discuss the challenges and opportunities with the goal of contributing to the development of criteria to assess which genetic results are best suited to provide research participants, adding to this discussion the need to consider the feasibility of implementing these criteria when working in a global clinical trial environment.

2464T

Interest in different types of individual genome sequencing results among younger breast cancer patients. *M.R. Ray[1], J. Ivanovich[1], B.B. Biesecker[2], L.G. Dressler[3], R. Dresser[4], M.S. Goodman[1], P.J. Goodfellow[5], K.A. Kaphingst[1].* 1) Division of Public Health Sciences, Washington University School of Medicine, Saint Louis, MO; 2) Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD; 3) Mission HealthCare, Asheville, NC; 4) Washington University School of Law, Saint Louis, MO; 5) College of Medicine, Ohio State University, Columbus, OH.

Background: Critical communication challenges arise if individual results from genome sequencing are returned. Patients diagnosed with breast cancer at a young age are a key population for early application of genome sequencing to identify susceptibility alleles and variants affecting treatment response. Empirical data are needed regarding what genome sequencing results are of interest to these patients to inform ongoing policy discussions for return of results. Methods: We conducted 48 semi-structured individual interviews with women diagnosed with breast cancer at age 40 or younger to investigate their interest in different types of genome sequencing results. We stratified recruitment by family history of breast cancer and BRCA1/2 mutation status to examine differences across subgroups. Interviews focused on interest in return of individual results for six types of gene variants: affect risk for a preventable disease; affect treatment response; uncertain clinical significance; affect risk for a disease that cannot be prevented or treated; carrier status; and ancestry or physical traits. Descriptive statistics were used to examine types of variants of greatest and least interest. Qualitative thematic analysis of interview transcripts was conducted using NVivo. Results: Participants identified variants that affect risk for a preventable disease (81%) or affect treatment response (17%) as those of greatest interest, often due to clinical actionability: '... you can delay the onset of any real symptoms or issues...it's always gonna come back to quality of life.' Participants had the least interest in variants without a health meaning (56%) or that affect risk for an unpreventable or untreatable disease (27%). Lack of interest in the latter type often related to quality of life: 'If there's nothing I can do about it, I'd rather live as happy a life as possible.' Participants differed in their interest in variants of uncertain clinical significance, with some suggesting the information might have meaning in the future, while others thought it could cause unnecessary worry. Conclusions: Participants had the greatest interest in clinically actionable results; results that might negatively impact quality of life were of less interest. However, interest in different types of results, and perceptions of clinical and personal utility, varied. Policies for return of individual genome sequencing results should take into account differences in patient preferences.

2465W

Ethically Optimized Consents for Pediatric Whole Genome Sequencing. *R. Zlotnik Shaul[1, 2, 3], M.S. Meyn[2, 4, 8, 9, 10], M. Szego[3, 5, 6, 7], R. Hayeems[8], C. Shuman[8, 9, 10], N. Monfared[4], S. Bowdin[4, 5, 8, 10].* 1) Bioethics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Dept of Paediatrics, University of Toronto, Toronto, ON, Canada; 3) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 4) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 5) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 6) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 7) Centre for Clinical Ethics: Providence Healthcare, St. Joseph's Health Centre and St. Michael's Hospital: Toronto, ON, Canada; 8) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, The Hospital for Sick Children; 9) Program in Genetics & Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 10) Dept of Molecular Genetics, University of Toronto, ON, Canada.

Whole genome sequencing (WGS) is a transformative technology for gene discovery and clinical care, but the clinical use of WGS presents major challenges to existing ethical frameworks. Most agree that medically actionable genomic variants (MAVs) should be disclosed to research participants/parents, but there is no consensus on what constitutes a MAV or what obligation researchers and clinicians have to look for and report MAVs that are incidental to the research or diagnosis at hand. For WGS of children decision-making is especially complex, since clinically relevant variants identified in a child's genome may have health implications for relatives. There is a paucity of published data examining the perceived benefits and risks of WGS from the perspectives of patients/parents and health care providers. An examination of 43 pediatric research consent forms from across Canada highlighted the variability and lack of national/international guidelines on how incidental findings should be treated. Developing a consistent, ethically defensible, innovative and inquiry-based approach to informed consent and return of WGS results in the pediatric context is essential. The Hospital for Sick Children's Genome Clinic is a multidisciplinary research project for which we have designed a novel consent process for pediatric WGS. A morally innovative feature of the consent is the choice given to parents and/or capable patients about the disclosure of MAVs associated with adult-onset disorders and carrier status. Past genetic testing paradigms have generally not favored disclosing this information to parents since testing/disclosure was not considered to be in the best interest of the child. Aligned with emerging guidelines we have interpreted best interests more broadly: the presence of an incidental MAV (e.g. BRCA 1 or 2) in the patient has potential health implications for the parent and knowledge about parental disease-risks, has implications for the best interests of the child. Additional features of our consent process will be described as part of this presentation. To help evaluate our consent, we are using structured interviews and qualitative analysis to examine participant/parent perceptions/expectations regarding WGS and the consent process and clinician expectations/challenges regarding WGS. Our inquiry is the first to examine these issues in a Canadian pediatric population and will be used to develop evidence-based standards for informed consent for WGS.

2466T

Evaluation of ACMG Recommended Incidental Findings in Clinical Whole Exome Sequencing. *E. Haverfield[1], A. Daly[1], A. Fuller[1], N. Smaoui[1], K. Retterer[1], P. Vitazka[1], G. Richard[1], W. Chung[2], S. Bale[1].* 1) GeneDx, Inc., Gaithersburg, MD; 2) Columbia University Medical Center, New York, NY.

Whole exome sequencing (WES) has the potential to identify incidental findings, or genetic variants that can result in medical implications unrelated to the primary indication for testing. The American College of Medical Genetics and Genomics (ACMG) recently recommended reporting incidental findings in 57 genes for individuals undergoing WES. These 57 genes are primarily associated with an increased risk for cancer or cardiac disease. The ACMG estimated that ~ 1% of patients undergoing WES will have reportable incidental findings. To evaluate this, the frequency of incidental findings in these 57 genes was examined in 50 de-identified probands who had WES and did not have cardiac disease or cancer as a test indication. The 57 genes were assessed for sequence changes present at ≤2% frequency in the 1000 Genomes database, and further reviewed for variant frequency in sub-populations, evidence for pathogenicity in the literature, and consistency with the expected mutation spectrum of the gene. The initial screen yielded 345 variants, of which 318 variants were eliminated based on high frequency in at least one ethnic sub-population, leaving 27 variants requiring further assessment. Of those, ten mutations (one in each of 10 individuals) met the criteria outlined by the ACMG to be reportable as known pathogenic (KP) or expected pathogenic (EP), demonstrating that 20% of WES cases have a reportable incidental finding. Of the 10 mutations, two were in cancer-related genes, six in cardiac-related genes and one mutation each was identified in a gene associated with hypercholesterolemia and malignant hyperthermia. Identification of incidental findings in 20% of patients undergoing WES has significant implications for ordering clinicians, genetic counselors and the clinical laboratories providing testing. Clinicians who return incidental findings to probands and families have an increased responsibility for pre and post-test counseling as well as medical follow-up, while clinical laboratories must integrate additional bioinformatic and analytic resources into these evaluations. The costs associated with the additional efforts could be substantial. Going forward, it will be important to continue to evaluate the impact of reporting incidental findings on patients and providers.

2467W

Giving and receiving: comparing parents', paediatricians' and genetic health professionals' opinions about uncertain chromosomal microarray results. *E. Turbitt[1,2], D. Amor[1,2], J. Halliday[1,2], S. Metcalfe[1,2].* 1) Murdoch Childrens Research Institute, Melbourne, Australia; 2) The University of Melbourne, Melbourne, Australia.

Chromosomal microarray (CMA) testing for childhood investigations has been on the Australian Medical Benefits Schedule (the federal healthcare funding scheme) since 2010. This has enabled a range of clinicians to order the test, including non-specialists. Although CMAs improve diagnostic yields, they also increase the detection of variants of uncertain clinical significance. The opinions of parents, paediatricians and genetic health professionals regarding CMA result disclosure in a clinical setting have not been documented. Our aim was to investigate and compare the perspectives of these three groups.

Following 15 qualitative interviews with key informants, a questionnaire was designed and piloted with 11 parents. Recruitment of parents was then undertaken at community-based immunisation sessions, and recruitment of paediatricians and genetic health professionals was undertaken using email invitation via listservs.

Responses were received from 147 parents, 154 paediatricians and 52 genetic health professionals. Interim analysis indicates all health professionals preferred to disclose all CMA results, regardless of level of certainty relating to clinical significance. At most, 75% of parents reported definitely wanting to receive all result types, with the remainder being unsure (Kruskal-Wallis:p<0.001). Despite this widespread preference to disclose uncertain results, health professionals reported experiencing difficulty explaining them to parents. Parents perceived uncertain results as having less practical value (e.g. access to services/treatment) and predicted these results would cause more anxiety than a certain result.

This study offers novel insight into the opinions of parents and health professionals about the disclosure of uncertain CMA results in Australia. Results demonstrate a degree of discordance between health professionals and consumers, which should be considered in development of optimal national guidelines for reporting genomic information.

2468T

Development of a decision aid for the selection of incidental genome sequencing results. *Y. Bombard[1,2,3], K. Schrader[1], E. Glogowski[1], M. Salerno[1], S. Patil[2], M. Massie[4], R. Rau-Murthy[1], M. Corines[1], C. Manschreck[1], J. Vijai[6], Z. Stadler[1], S. Lipkin[5], K. Offit[1,5,6], M. Robson[1], (first and second authors have equal contributions).* 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) University of Toronto, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Ontario, Canada; 4) Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY; 5) Weill Cornell Medical College, New York NY; 6) Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY.

Background: Conceptual frameworks have been developed to guide the disclosure of incidental results from whole genome or exome sequencing (WG/ES), however limited tools exist to support patient decision-making about the receipt of large volumes of incidental data. We aimed to: (1) develop and evaluate a decision-aid categorizing incidental WG/ES results and (2) explore perceived utility and decision-making regarding the categories of results. Methods: We developed a decision-aid based on Berg et al's 'binning' scheme. Variants are categorized by their level of actionability and potential for distress. Bin 1 includes variants with established disease prevention or treatment guidelines (as the ACMG recent guidelines); Bin 2 includes clinically valid variants without established interventions and is further divided into 2a,b,c (refer to companion abstract by Schrader for binning methodology); Bin 3 includes carrier results; Bin 4 contains variants of unknown significance. We are pre-testing the decision-aid through interviews and focus groups with a convenience sample (n≈40) recruited from hospital waiting rooms. The decisional conflict scale assesses satisfaction and effective decision-making. Data analysis includes qualitative content analysis and descriptive statistics. Results: Preliminary results demonstrate wide variation in selection and perceived utility of the bins. Participants favored Bin 1 because results were actionable. They perceived Bin 2a unlikely to have major health/lifestyle impact. Bin 2b was difficult to understand because it consists of a heterogeneous group of variants (carrier, diagnostic and progressive results). Progressive diseases were perceived as more distressing, precluding selection of Bin 2b. Selection of Bin 2a and 2c was driven by motivations to employ risk-reducing behaviors and inform life plans. Selection of Bin 3 varied by attitudes towards reproductive intervention. Participants had mixed views about sharing results with relatives, doctors and among researchers. Conclusion: Preliminary results suggest that this decision-aid can foster informed decision-making, which involved trade-offs between potential distress with perceived clinical or lifestyle benefits, or informing life-plans. However perceptions of utility and 'distressing results' may diverge from clinical definitions and expectations. Future work will evaluate decisional-conflict and potential re-conceptualization of Bin 2.

2469W

Actionable incidental findings in the 1000 Genomes dataset.
E. Olfson[1], C.E. Cottrell[2], N.O. Davidson[3], N.O. Stitziel[4,5], L. Chen[1], S. Hartz[1], S. Koul[2], R. Nagarajan[2,6], N.L. Saccone[6], L.J. Bierut[1]. 1) Department of Psychiatry, Washington University School of Medicine, St Louis, MO; 2) Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO; 3) Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St Louis, MO; 4) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St Louis, MO; 5) Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO; 6) Department of Genetics, Washington University School of Medicine, St Louis, MO.

Purpose: Advances in genomic sequencing have increased the likelihood of uncovering incidental findings. The goal of this study was to provide a conservative estimate of the frequency of actionable incidental findings using the 1000 Genomes dataset and to present a multi-stage systematic approach for identifying these findings. Our approach not only focused on diseases that several specialists agree should be returned as incidental findings in clinical sequencing studies, but also incorporated literature review and expert opinion to provide a lower bound estimate of the prevalence of incidental findings. **Methods:** Candidate single nucleotide variants in 17 genes associated with 11 dominant actionable conditions were extracted from the 1000 Genomes Browser. These variants were filtered using the Human Gene Mutation Database and defined parameters, appraised through literature review, and examined by a clinical laboratory specialist and expert physician. Only variants that the clinical specialists agreed were pathogenic were identified as incidental findings. **Results:** From the 1000 Genomes Browser, 22 162 single nucleotide variants were extracted from the 17 genes, and filtering led to 119 candidate variants. Literature review refined this number to 11 variants, and clinical specialists agreed that 2 were likely pathogenic: *LDLR* p.Trp4* associated with familial hypercholesterolemia and *KCNH2* p.Leu552Ser associated with Long QT syndrome. For 3 variants, 2 in *APC* and 1 in *MLH1*, there was discordance between the pathogenicity predictions of the specialists: while the clinical laboratory specialist determined that they were variants of unknown significance or rare polymorphic variants, the expert physician identified these variants as probably pathogenic. **Conclusion:** Among individuals from diverse populations, at least 0.2% carry an actionable pathogenic variant, demonstrating that a non-trivial number of individuals will benefit from incidental finding return. As our knowledge about the medical actionability of diseases, genes, and specific variants evolves, the likelihood of uncovering these findings will increase. The time-consuming process of manual review, coupled with differences in the pathogenicity thresholds of specialists, indicate a need for streamlined classification allowing rapid identification of actionable incidental findings as well as reliable guidelines for variant evaluation to ease the translation of genomics into clinical care.

2470T

Binning framework for a decision aid for the selection of incidental genome sequencing results.
K. Schrader[1], Y. Bombard[1,3,4], E. Glogowski[1], M. Salerno[1], S. Patil[3], M. Massie[5], R. Rau-Murthy[1], M. Corines[1], C. Manschreck[1], J. Vijai[2], Z. Stadler[1], S. Lipkin[6], K. Offit[1,2,6], M. Robson[1], (first and second authors have equal contributions). 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY; 4) University of Toronto, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON; 5) Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY; 6) Weill Cornell Medical College, New York, NY.

Trepidation of disclosure of incidental genomic results stems from a perceived difficulty in adequately counseling patients about the broad scope of available findings and implications. To simplify this process, Berg [2011,2013] pioneered the concept of 'binning' incidental results based on medical actionability (Bin1) or, in it's absence, the potential for psychological harm with disclosure (Bin2a-c). We have adapted these bins and propose further modifications to attempt to improve patient understanding and decision-making. Berg's bins acted as a guiding framework for our decision aid, and were adapted by consensus of our working group of medical and cancer geneticists, genetic counselors, and a public health genomics researcher. Adaptations include: adding Bin3, which allows patients to learn healthy carrier status (for diseases on New York State's newborn screening panel) without the rest of Bin2b; moving genes identified as medically actionable by the ACMG [2013] to Bin1, recommending but not mandating Bin1 disclosure; limiting Bin2a SNPs to those used by Coriell Personalized Medicine Collaborative, limiting the number of pharmacogenomic SNPs, and limiting Bin2c to few progressive diseases with which the lay population is generally familiar. Compared to Berg's bins, Bin2b still contains disease genes ineligible for the other bins, and Bin2c contains selected genes with the highest subjective risk for psychological harm. Bin4 still contains uninterpretable data not available for disclosure. Determination of medically actionable genes is an iterative process and concerted efforts have developed frameworks for doing this in a reproducible manner. Determining which genes may cause psychological harm and comprise divisions of Bin2 is decidedly more subjective. Adapted bins are pretested in the decision aid (refer to companion abstract Bombard et al). Preliminary findings suggest patients have difficulty accepting the broad implications of Bin2b, thus we propose remodelling using an alternate framework to separate genes into three inherent divisions 1) Diagnostic: suggesting static, congenital conditions that are potentially diagnosable with further clinical investigation 2) Carrier: with no anticipated phenotype and 3) Predictive: suggesting risk for future disease eg. progressive degenerative disorders. Patient data will guide whether additional parsing (or merging with other bins) may enable patients to make more informed decisions.

2471W

How Interested are Parents in Genome Screening for Their Newborns?

R.C. Green[1,2], S.S. Kalia[1], H.L. Levy[3], I.A. Holm[4], S.E. Waisbren[3,5]. 1) Department of Medicine, Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Partners Center for Personalized Genetic Medicine, Boston, MA; 3) Division of Genetics, Boston Children's Hospital, Boston, MA; 4) Division of Genetics, Program in Genomics, and the Manton Center for Orphan Diseases Research, Boston Children's Hospital, Boston, MA; 5) Department of Psychiatry, Boston Children's Hospital, Boston, MA.

Background: Parents of newborns are highly stressed while in the hospital after delivery, yet this would be the most efficient time to enroll such parents into empirical studies of newborn genome screening (NGS). Concerns about enrollment at this time include whether discussing this topic would cause parents to question or refuse state-mandated newborn screening (NBS), and whether parental interest in NGS research would be the same if queried several months later, after the post-partum period. Methods: We surveyed parents of healthy neonates on the post-partum unit at Brigham and Women's Hospital about their interest in participating in research where their infant would receive NGS. Parents rated their interest on a 5-point Likert scale from 'not at all interested' to 'extremely interested'. We are currently re-contacting these parents and have so far surveyed 66 parents at 3-6 months after delivery with the same questions. Results: Parents were systematically approached while on the post-partum unit and 87.3% agreed to participate. Among 216 parents queried about their interest in NGS, none questioned or declined state-mandated NBS. Responses were dichotomized to 'not at all' or 'a little' interested in NGS, versus 'somewhat,' 'very,' or 'extremely' interested. In the hospital, 182 (84.3%) parents were at least 'somewhat' interested in NGS, and 101 (46.8%) were 'very' or 'extremely' interested. Level of interest was not significantly associated with parents' age, sex, race, education, income, or number of biological children. Among the 66 parents surveyed again after 3-6 months, 53 (80.3%) were at least 'somewhat interested' in NGS in the hospital vs. 55 (83.3%) at follow-up. However, the intensity of interest declined slightly over time; 32 (48.5%) were 'very' or 'extremely' interested in the hospital vs. 27 (40.9%) at follow-up. At 3-6 months, 58 parents (87.9%) remained equally interested in NGS, while 3 (4.5%) became less interested, and 5 (7.6%) became more interested. Conclusions: These preliminary data suggest that parents of healthy neonates can be surveyed about NGS in the hospital without threatening compliance on conventional NBS. Interest in research that would involve NGS is high among parents of newborns in the hospital, with nearly half reporting that they would be 'very' or 'extremely' interested in such research. Interest in NGS appears to remain high 6 months later, though the intensity of the interest may decline slightly over time.

2472T

Psychological responses to genetic risk disclosure among individuals at imminent risk for Alzheimer's disease and their study partners: findings from the REVEAL Study.

K.D. Christensen[1], J.S. Roberts[2], J.H. Karlawish[3], T.O. Obisesan[4], L.B. Waterston[1], L.A. Cupples[5], W.R. Uhlmann[6], E. McCarty Wood[7], R.C. Green[1,8] on behalf of the REVEAL Study Group. 1) Dept Medicine, Div Genetics, Brigham & Women's Hosp/Harvard Med, Boston, MA; 2) Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, MI; 3) Department of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Medicine, Howard University, Washington, DC; 5) Departments of Biostatistics & Epidemiology, Boston University School of Public Health, Boston, MA; 6) Departments of Internal Medicine & Human Genetics, University of Michigan, Ann Arbor, MI; 7) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 8) Partners HealthCare Center for Personalized Genetic Medicine, Boston, MA.

Background: Studies of genetic risk assessments for common disease consistently show that individuals are not adversely affected by results. However, few studies have enrolled populations at imminent risk for disease onset or examined the impact on care partners. **Methods**: The REVEAL Study enrolled patients with mild cognitive impairment (MCI) and their cognitively normal study partners, typically a spouse or adult child. Depending on randomization, genetic counselors disclosed 3-year Alzheimer's disease risk estimates (range: 8-57%) based on age and MCI diagnoses alone or in conjunction with *APOE* genotypes. Primary outcomes included general anxiety, per the State-Trait Anxiety Inventory (STAI) short form; and general depression, per the Geriatric Depression Rating Scale (GDRS). Test-related distress was assessed secondarily using the Impact of Events Scale (IES). Individuals with baseline scores suggesting moderate anxiety or depression were excluded from participation. **Results**: Among 61 patients assessed to date (mean age 72; 46% male; 18% Black), mood scores 6 weeks post-disclosure were well below cutoffs for concern (cutoffs: STAI=17, GDRS=8, IES=26) for both the genotyped group (STAI=11.7, GDRS=1.9, IES=11.4) and control group (STAI=11.9, GDRS=2.7, IES=15.7). Changes from baseline did not differ by randomization arm or genotype. However, IES scores 3-days post-disclosure were greater among ε4-positive patients than control (Δ=5.9, p=.03) and ε4-negative patients (Δ=7.3, p=.01), although differences were not observed at the 6-week follow-up. Among study partners, post-disclosure STAI and GDRS scores were also well below cutoffs for the genotyped group (STAI=12.3, GDRS=1.3, IES=11.4) and control group (STAI=11.6, GDRS=1.6, IES=11.4). Changes from baseline in study partners were not observed by randomization arm, but IES scores of partners of ε4-positive patients were greater than partners of ε4-negative patients 3-days post-disclosure (Δ=8.4, p=.01). Differences in IES scores among study partners were not observed at the 6-week follow-up. **Conclusions**: Genetic risk information about imminent onset of a common disease did not pose significant short-term psychological risks. Test-related distress immediately following disclosure was greater among participants with increased genetic risk and their study partners, but effects were transient. Results correspond with prior work on healthy adults, and extend those findings to potential caregivers.

2473W

The impact of genetic risk disclosure for Alzheimer's disease: Findings from the REVEAL Study APOE ε4 homozygotes. *L.B. Waterston[1], J.H. Karlawish[2], J.S. Roberts[3], C.A. Chen[4], K.D. Christensen[1], R.C. Green[1, 5] for the REVEAL Study Group.* 1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Department of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, MI; 4) Boston University School of Public Health, Boston, MA; 5) Partners Healthcare Center for Personalized Genetic Medicine, Boston, MA.

Background: Prior work has shown that individuals at increased risk for Alzheimer's disease (AD) because they carry at least one copy of the APOE ε4 allele experience greater test-related distress, perceive greater AD susceptibility, and make more changes to health behaviors than ε4 negative (ε4Neg) individuals. Analyses to date have not differentiated between ε4 homozygotes (ε4HM) and ε4 heterozygotes (ε4HT), however. Methods: Cognitively normal adults (mean age 57; 64% female; 84% White, 15% African American; 88% with family history of AD) across three randomized trials learned their APOE genotypes and AD risk estimates (range: 6-77%) based on their genotype, gender, ethnicity, and family history. Primary outcomes measured at 6 weeks, 6 months and one year post-disclosure were symptoms of anxiety, depression, and test-specific distress. Also assessed at one year were perceived AD risk (proportions endorsing 'high' or 'very high'), self-reported behavior changes (diet, exercise, taking medications or vitamins) and overall impact of risk disclosure. Results: Of 648 genotyped participants, 399 were ε4Neg (62%), 221 were ε4HT (34%), and 28 were ε4HM (4%). ε4Neg, ε4HT and ε4HM's did not differ in serial depression or anxiety measures (all $p>0.05$). Test-specific distress was higher among ε4HT and HM's at all time points (all $p<0.0001$), but ε4HT and ε4HM's did not differ on this measure (all $p>0.05$). Differences in percentage of ε4HT and ε4HM's perceiving their AD risk as high were not evident (64% versus 72%; $p>0.05$), but both groups perceived their AD risk as higher than ε4Neg's ($p<0.001$). ε4HM's were more likely than ε4HT's and ε4Neg's to report 12-month changes in diet (61% versus 34% and 27%), exercise (58% versus 30% and 28%) and taking medications or vitamins (58% versus 38% and 27%) (all $p<0.001$). ε4Neg's were more likely than other groups to rate the overall impact as positive (74%, $p<0.001$), but ε4HT's and ε4HM's did not differ in their percentages rating the overall impact as positive (50% vs. 50%). Conclusions: Disclosure of ε4 homozygote genotype was not associated with increased perceived risk of AD, depression, anxiety, test-specific distress, or overall impact ratings compared to being a heterozygote; but it was correlated with greater likelihood of engaging in putative AD risk reducing behaviors. Such findings can inform future efforts to disclose APOE genotype to ε4 homozygotes in research trials, consumer settings or clinical practice.

2474T

A historical overview and a reflection on bioethical and ELSI activities for basic researches in human genetics and genomics in Japan. *N. Yamamoto, K. Kato.* Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Suita, Osaka, Japan.

Over 20 years have passed since the human genome project started. Including genetic-recombination era, it is more than half a century. Though it may come across as being a short period compared to the history of many scientific disciplines, the world of biology significantly changed with these new genetics and genomics. Furthermore, this new science gave the great impacts to bioethical and ELSI activities that appeared in that almost same period in Japan. We first overview a history of bioethical and ELSI activities, focusing on the following four kinds of activities; the bioethics committees of ministries, the ethics review boards in the institutes, the specialist groups or committees for ELSI in the large-scale research projects, and the bioethicists in the academic sector. From these viewpoints, it was shown that the people or groups have created new ways or changed old ones to respond the requirements conducting the research appropriately. For example, the fundamental rights of the research participants, such as the donor autonomy and privacy protection, became guaranteed in the written research guidelines since around 2000 by the efforts of the ministries bioethics committees. However, comparing these activities and present science and social situations, it was also revealed that there are a few remaining issues that have been considered less, therefore have not been fully introduced in the formal systems, such as the right of the populations, the right of the people other than a direct relationship to the donors of DNA, the long-term storage and reuse of the data, and the ethical status of human genome itself. In addition, we looked back the activities of another category of actors, that is, the scientists engaging basic research in molecular genetics and genomics. Some of the scientists' concerns are different from the ones described above, because their research are not directly associated with newly born issues in bioethics nor ELSI. What they are concerned was more about the governance or the national/societal policy of the science. What we would like to show in this study is that in Japan there are many bioethical and ELSI related activities and they have different beginning and process of development. We believe that we can learn many lessons from these historical analyses for the better design of future ELSI activities in the field of genetic and genomics.

2475W

Result of the bioethics questionnaire survey of Nepali university students on genetic testings. *H. NUMABE[1,2], R. POKHAREL[3].* 1) Department of Genetic Counselling, Graduate School of Humanities and Sciences, Ochanomizu University, Bunkyo-ku, Tokyo, Japan; 2) Department of Paediatrics, Kyoto University Hospital, Kyoto, Japan; 3) Department of Orthopaedics, Institute of Medicine, Tribhuvan University Teaching Hospital, Kathmandu, Nepal.

In Asian countries such as Japan, Korea, and China, genetic testing has been carried out using similar methods with accuracy comparable to western countries. However, the implementation of testing is influenced by various factors such as population policy, economic situation, and the medical system of each government, as well as the bioethical view of each individual. Especially, culture plays a substantial role in genetic medicine and reproductive medicine. It is thus imperative to take cultural influence into consideration when giving genetic counseling and genetic testing to patients and clients in Japan and in the other Asian countries. We made the questionnaire survey of Japanese university students on genetic diagnoses, and reported the result at the ASHG Meeting in 2008. The questionnaire was consist of properties of the student, considerations to spiritual matters, acceptability of an artificial abortion, the time of the beginning of human life, and considerations to various genetic testing. We made the similar questionnaire survey in Nepal in 2012. We make an interim report. In Nepal, 70 medical students (65 Hindu, 2 atheists, 1 Buddhist, 1 Muslim, 1 Jewish person) answered our questionnaires. Fifty students recognized that the human life begins from the fertilization. Acceptances of genetic testing were as follows: Clinical genetic testing 91%, [Postnatal] Predictive Genetic Diagnosis 74%, Genetic Carrier detection 84%, Prenatal diagnosis 82%, Preimplantation genetic diagnosis (PGD) 81%. We also compare the results of Japanese and Nepali.

2476T

Attitudes and concerns of educated scientists undergoing whole exome sequencing. *A. Fiksdal, N. Lindor, K. Johnson, K. Hunt, K. Mensink, J. McCormick.* Mayo Clinic, Rochester, MN.

Background: Next-generation sequencing technologies have advanced rapidly over the past decade, resulting in a dramatic decrease in the cost of whole genome and whole exome sequencing. Although these technologies are quickly being incorporated in the practice of individualized medicine, very little research has been conducted evaluating the process of sequencing from a patient perspective. Methods: Mayo Clinic is currently conducting a study in which 10 educated scientists with no history of a specific disease or medical condition are receiving whole exome sequencing (WES) at no cost. We initially invited 40 individuals to participate in a lottery to offer genetic counseling and testing to 10 participants. Those who declined or did not respond to the invitation were not included in the lottery. Participants attended two genetic counseling sessions prior to sequencing. They also completed three surveys assessing attitudes and experiences related to the WES process before, between, and after the genetic counseling sessions. We analyzed the surveys using standard qualitative methodology. After participants receive their WES results, they will participate in an additional genetic counseling session and focus group. We will also complete interviews with those who were not chosen by the lottery. Results: Participants were generally comfortable with the WES process. Only one participant reported concerns regarding confidentiality of data. The same participant also reported concerns about colleagues learning their genomic information. Additionally, no participants reported being concerned about how they might react to particular genetic information. Participants reported more concerns in the context of how WES may affect their families. In survey 3, 3/10 participants agreed with the statement 'I am concerned about how a family member might react about a particular finding if I were to share it with them.' These participants cited the 'impact on their [family members'] health and psychological state', actionability of results, and specific concerns voiced by family members as factors that influenced their views. Conclusion: Although the majority of our sample did not report significant concerns regarding the WES process, issues surrounding the returning and sharing of results emerged as an area that required careful consideration. Such issues will be examined in greater depth in future surveys and focus groups after participants receive their results.

2477F

Parental decisions following prenatal diagnosis of chromosomal abnormalities around Nagoya, Japan. *N. Suzumori[1], K. Kumagai[1], S. Goto[1], A. Nakamura[2], S. Saitoh[1], M. Sugiura-Ogasawara[1].* 1) Nagoya City University, Nagoya, Japan; 2) Chiba University of Commerce, Chiba, Japan.

Indications for amniocentesis at 15-18 weeks of gestation include advanced maternal age, maternal serum screening results, ultrasonographic findings for congenital malformations or markers of aneuploidy, and chromosomal abnormality in a previous pregnancy. Cytogenetic karyotype analysis by amniocentesis is highly accurate as compared to cell free fetal DNA testing, which has become clinically available in a few countries for women at increased risk of fetal aneuploidy. When an abnormal karyotype is identified by prenatal diagnosis, parents are faced with decision options that include termination versus continuation of the pregnancy. Decisions concerning prenatal testing and termination of pregnancy in case of affected fetuses are complex and may be influenced by a variety of factors, such as the country's health system, its abortion laws, as well as social and cultural backgrounds Parental decision-making to terminate or continue a pregnancy was studied after prenatal diagnosis of a chromosome aneuploidy among a sample of patients around Nagoya, Japan. The 1051 amniocentesis cases at 15-18 weeks of gestation were analyzed. The 60 cases of chromosomal anomalies with aneuploidies were diagnosed by conventional cytogenetic analysis. Of the 45 diagnoses of autosomal chromosome aneuploidies, pregnancy was terminated in 93.3%. Of the 15 cases diagnosed with sex chromosome aneuploidy, pregnancy was terminated in 46.7%. The differences of parental decisions with respect to maternal age, gestational week at diagnosis, number of pregnancies per individual and existing number of children were not significant in patients diagnosed with autosomal or sex chromosome aneuploidy. The present study indicates that when diagnosed with a chromosome aneuploidy in which a severe prognosis was expected, most of the couples decided to terminate the pregnancy in Japan.

2478F

Parental causal attributions of OCD and implications for genetic counseling: An exploratory study. *H.J. Andrighetti[1], A. Semaka[2], S.E. Stewart[2], C. Shuman[1], D. Chitayat[1], R. Hayeems[1], J.C. Austin[2].* 1) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada.

Introduction: Studies assessing etiological understanding and interest in genetic counseling have been conducted among patients with complex psychiatric illnesses including schizophrenia and bipolar disorder. Obsessive compulsive disorder (OCD) is an anxiety disorder with primarily pediatric onset and well-documented unique impacts on family functioning. Limited research has been done on etiological understanding and the potential role of psychiatric genetic counseling among families affected by OCD. The purpose of this study is to explore parents' etiological understanding of OCD, and the potential value of psychiatric genetic counseling in this population.

Methods: Parents of children diagnosed with OCD were recruited through the BC Children's Hospital OCD Clinic and Translational Research Program. We administered qualitative semi-structured telephone interviews consisting of open-ended questions exploring participants' experiences with their child's OCD, causal attributions of OCD, and perceptions of two genetic counseling scenarios. Interviews were audio-recorded, transcribed, and analyzed thematically using qualitative content analysis and elements of grounded theory.

Results: Interviews with 13 parents of 13 different children with OCD revealed that their experiences with OCD are complicated by the evolution of the condition and treatments over time, stigmatization of child and parents, and significant impacts on the child's life and family functioning. Parental causal attributions for OCD were diverse, but often involved elements of self-blame. Psychiatric genetic counseling was perceived as useful for decreasing parental blame and guilt and empowering parents to make well-informed decisions.

Conclusions: These data provide insight into the education and counseling needs of this population, marking the first step towards provision of effective and comprehensive genetic counseling services for families affected by OCD. These results are similar to findings from research on other complex psychiatric illnesses and can be used to inform future studies and practice guidelines for other anxiety-related and/or pediatric-onset psychiatric disorders.

2479F

An exploration of families' experiences regarding a comorbid diagnosis of neurofibromatosis type 1 (NF1) and autism spectrum disorder (ASD) in their child: guiding screening and disclosure practices. *L. Baret[1,2], R. Hayeems[1,2], P. Parkin[1,2], C. Shuman[1,2], M. Carter[1,2], P. Kannu[1,2], D. Chitayat[1,2], A. Shugar[1,2].* 1) The Hospital for Sick Children, Toronto, Ontario, Canada; 2) University of Toronto, Toronto, Ontario, Canada.

Introduction: Recent studies have identified an increased prevalence of autistic symptomatology in the NF1 population. Given these findings, it has been suggested that screening and diagnostic practices for young children with NF1 be expanded to encompass screening for ASD. Little is known about the attitudes of key stakeholders (parents of children with NF1 and ASD) regarding screening implementation.

Objective: We aim to explore the experiences of parents of children diagnosed with both NF1 and ASD and to probe their attitudes towards the potential implementation of early universal ASD screening for children with NF1.

Methods: We conducted a comprehensive retrospective chart review ascertained via the NF1 pediatric database at the Hospital for Sick Children. The chart review identified 22 children from 21 families with clinically confirmed diagnoses of both NF1 and ASD. With informed consent, we conducted 16 open-ended, semi-structured interviews with parents (73% response rate). Transcripts were analyzed for thematic patterns using qualitative content analysis.

Results: Most parents expressed relief after receiving an ASD diagnosis. Different meanings were conferred to the two diagnoses. The ASD diagnosis was perceived to be more closely tied the childs identity, thereby conferring a greater impact on the parental psychosocial construct. The ASD diagnosis had both intrinsic and instrumental value for families as it also allowed them to access appropriate resources. In comparison, the meaning attributed to the NF1 diagnosis was reported as less consequential. Interestingly, perceptions shifted if severe NF1 complications were experienced. Parents acknowledge the challenges associated with potential universal ASD screening; they fear a double label could provoke anxiety and present an additional barrier to adapting to NF1. Yet they almost unanimously support early ASD screening.

Conclusion: Implementing ASD screening for children with NF1 is likely to be beneficial for families if screening guidelines also address the psychosocial and medical impacts of an early comorbid diagnosis.

2480F

Perspectives of Adolescents Regarding their Genetic Counseling Experience: A Qualitative Study. *A. Pichini[1], K. Sappleton[1], M. Kaufman[1], C. Shuman[1], D. Chitayat[1,2], R. Babul-Hirji[1].* 1) Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Prenatal Diagnosis and Medical Genetics, Mount Sinai Hospital, Toronto, Ontario, Canada.

Adolescence is a period of development that involves creating a sense of individuality, self-concept, sexual identity, independence, relationships, personal values and goals. This stage of adjustment can be complicated by having a genetic condition, and genetic counseling can play a crucial role in providing information and support to this adolescent patient population. We sought to investigate the experiences and perspectives of adolescents with a genetic condition regarding their genetic counseling interaction. Information gleaned from this study will contribute to the development of an adolescent-focused framework to enhance the provision of genetic counseling. This study utilized a qualitative exploratory approach, specifically the Interpretive Description methodology. Eleven semi-structured interviews were conducted in person and by telephone, with adolescents diagnosed with a genetic condition who received genetic counseling between the ages of 12 and 18 years at The Hospital for Sick Children. Transcripts were analyzed thematically using qualitative content analysis. Preliminary analysis indicates the following themes: (i) adolescents view genetic counseling as distinctly important for understanding the underlying cause of their condition; (ii) genetic counseling plays a crucial role in helping adolescents appreciate the significance of their clinical management or surveillance; (iii) visual aids, particularly the pedigree, are helpful to conceptualize how the genetic condition segregates in the family; (iv) adolescents are interested in learning about recurrence risks for future children even if family planning is not of immediate relevance; (v) adolescents prefer having parents present during the session, except during discussions about prognosis and life expectancy. To our knowledge, this study is the first to explore the perspectives of adolescents with respect to their genetic counseling experience. These findings can inform emerging genetic counseling approaches unique to this patient population, and facilitate in the transition process from pediatric to adult care within patient and family-centered contexts.

2481F

Reproductive genetic counseling - The most suitable timing? -.
N. Takeshita, A. Takashima, Y. Yasuda, T. Adachi, I. Sasaki, K. Yokokawa, M. Manrai, T. Ichinose, A. Ishida, M. Urita, A. Yokoyama, T. Kinoshita. Department Obstetrics & Gynecology, Toho University, Medical Center Sakura Hospital 564-1 Shizu, Sakura, Chiba, Japan 285-8741.

[Introduction] Thorough genetic counseling is essential for preimplantation genetic diagnosis and prenatal diagnosis. However, there are more than a few cases in which the time available for the diagnosis is limited. Here, we discuss and report the importance of timely counseling based on our experiences in two prenatally diagnosed cases. [Case] (1) A 38-year-old unipara. An urgent hysterotomy was performed due to polyhydramnios and non-reassuring fetal status at the previous delivery. The newborn child died at one-day old due to respiratory failure and was diagnosed clinically having congenital myotonic dystrophy. This result made us to suspect the same disease in the mother and the mother was followed up at our department of internal medicine. Two years later, she visited our hospital with a history of secondary amenorrhea and was diagnosed as being pregnant. On learning about her unexpected pregnancy, she was confused, as she had been told that a natural pregnancy would be difficult and she did not wish to have the baby. She was already 15-weeks pregnant. (2) A 31-years old unipara. Her father was diagnosed as having adrenoleukodystrophy and being treated for it. She delivered a girl one and a half years previously. Because she was suspected to be a carrier based on the biochemical findings, she visited our hospital to seek counseling for her next pregnancy. While she and her husband were receiving counseling, her pregnancy was detected. [Results] Because in both cases, both the husband and wife sought a prenatal diagnosis, we set up an appointment for villi collection urgently at another institution and the gene analysis was performed. Prenatal diagnosis could be obtained in both cases. [Conclusion] In reproductive genetic counseling, some explanations about the next pregnancy is provided, however, the understanding usually becomes vague with time. In addition, we should keep in mind that a pregnancy is not always planned. It goes without saying that the independence of the client must be respected in genetic counseling, however, it is also important to establish a continuous communication and support system in reproductive genetic counseling in cases with time limitation.

2482F

Whole exome sequencing: Assessing what patients want to know.
K. Hitch[1], G. Joseph[2], J. Guiltinan[1], J. Kianmahd[3], J. Youngblom[1], A. Blanco[2]. 1) California State University, Stanislaus, Turlock, CA; 2) University of California, San Francisco, San Francisco, CA; 3) University of California, Los Angeles, Los Angeles, CA.

Whole exome sequencing (WES) uses next generation sequencing to provide information on nearly all functional, protein-coding regions in an individual's genome. Given the vast amount of information and incidental findings that can be generated from this technology with potentially significant health implications for patients and their family members, it is imperative to understand the patients' perspectives on genomic sequencing. To explore this topic, semi-structured interviews were conducted with 19 adult patients who consented to WES in the research setting. All participants were previously diagnosed with Lynch syndrome based on high microsatellite instability in their tumors and absence of one or more mismatch repair proteins based on immunohistochemistry results, but received uninformative negative results through traditional molecular testing methods. Nearly all participants believed that the benefits of receiving all possible WES results outweighed the potential risks. The majority of participants conveyed the sentiment that relative to coping with a cancer diagnosis, information generated from WES would be manageable. Participants' experience with Lynch syndrome influenced their notions of genetic determinism, their tolerance for uncertain results, and their family communication plans. Participants preferred to receive WES results in person from a genetic counselor or medical geneticist so that an expert could explain the meaning and implications of the potentially large quantity and range of results. Participants were keen on individual choice and empowerment, as demonstrated by their desire for direct access to WES results and transparency between clinicians and patients. These results underscore the need to study the needs and preferences for return of genomic sequencing results with various populations in order to effectively communicate the possible implications of this new technology and appropriately return results. The opinions of these patients revealed several pretest implications when consenting patients to genomic sequencing. Clinicians should provide direct and clear information about the results patients do or do not have control over choosing to receive. To ensure adequate anticipatory guidance and informed consent, laboratory consent forms should include the differing types of possible WES results and clinicians should divulge to patients the possible implications that could result from receiving WES data in detail.

2483F

Parents' understanding of uncertain chromosomal microarray results: a search for meaning.
L.A. Kiedrowski[1, 2], K.M. Owens[3], B.M. Yashar[2], J.L. Schuette[3]. 1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI.

Chromosomal microarray analysis (CMA) is utilized as first-tier testing for unexplained anomalies and developmental delay. This technology has improved diagnosis rates, but results classified as variants of uncertain significance (VUS) challenge both clinicians and parents. We developed a semi-structured telephone interview guide to explore parents' knowledge and interpretation of genetic test results and the impact of results and genetic counseling for CMA VUS. We interviewed parents (n=14) who received genetic counseling for their children's CMA VUS. Interview transcripts were analyzed for cross-cutting themes. Several thematic areas were identified, including comprehension of testing, scientific uncertainty, impact of results, expectations, and patient satisfaction. Participants demonstrated a wide range of recall and personal interpretation regarding whether test results provided a causal explanation for their children's health issues. Participants also maintained contradictory interpretations, i.e. describing the result as an answer but not an explanation for illness, or vice versa. Despite varied interpretation, parents consistently applied meaning by incorporating results into a personal context. Many expressed benefits to having a result, including obtaining further services and personal validation. A wide variety of emotional impacts (e.g. relief, loss, and guilt) were also noted; parents described an adaptation/coping process similar to that which occurs after a positive genetic test result. Recall of terminology, including `VUS' and precise CMA abnormality, was poor. However, the majority demonstrated an understanding of the concept of scientific uncertainty. Evidence of retention of relevant concepts without specific terminology warrants further consideration of the language and content that is truly necessary in genetic counseling for this type of finding. Reassuringly, all participants recalled and were interested in recommended follow-up with the genetics clinic, but had misconceptions as to the specific reasons for such a return visit and what this process entails. These results provide insight into the patient-and-family experience of genetic counseling for uncertain genomic findings. They emphasize the importance of exploring uncertainty in CMA VUS results with parents and highlight a number of areas for potential attention or improvement in the clinical encounter.

2484F

Comparing uptake of testing and psychosocial impact in pregnant and non-pregnant women offered carrier screening for fragile X syndrome.

S. Metcalfe[1,2], M. Martyn[1], J. Emery[3], J. Halliday[1,2], S. Donath[1,2], J. Cohen[4], the FaXeS study team. 1) Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Australia; 2) Dept Paediatrics, The University of Melbourne, Melbourne, Australia; 3) Dept General Practice, The University of Western Australia, Perth, Australia; 4) Fragile X Alliance and Centre for Developmental Disability Health Victoria, Monash University, Melbourne, Australia.

Offering population carrier screening to women for FXS identifies those at increased risk of having an affected child, and provides women with information about their own health risk. Concerns have been expressed around educational and counselling difficulties associated with the complex genetic and health information inherent in FXS carrier screening. Hence we examined decision-making, knowledge and attitudes in non-pregnant and pregnant women in the general population offered carrier screening for FXS. Women were approached through general practice, obstetric or ultrasound clinics, and received written information and telephone pre-test counselling with consent. At home, women decided about testing and completed a questionnaire (Q1), which they returned in the mail with their buccal sample as appropriate. A premutation (PM) or grey zone (GZ) result was discussed by telephone and women offered genetic counselling; test-negative results were mailed. All women received a second questionnaire (Q2). A sub-group of women were interviewed either at the time of making the decision or after completing Q2; program evaluation interviews were conducted with clinic staff. 1237 women initially consented: 702 non-pregnant and 535 pregnant women. Excluding 80 women who subsequently withdrew from the study, 85% and 81% returned Q1 while 76% and 70% returned Q2 respectively. 71% non-pregnant and 59% pregnant women chose testing; 0.4% (n=3) received a PM and 2.0% (n=15) a GZ result. Overall, 85% had good knowledge (≥7/10 correct) at Q1. 77% non-pregnant and 68% pregnant women had positive attitudes towards the test. Pregnant women were less depressed and less stressed than non-pregnant women on the Depression Anxiety Stress Scale, with no differences between tested and non-tested. Scores on this scale were in the normal range for all but a few women. Women who chose not to be tested had significantly higher decisional conflict than those who were tested, but scores were in the normal range. The majority of women offered FXS carrier screening had good understanding despite the complex nature of the information, with minimal psychosocial impact. These may be related to the extensive prior field-testing of the written information and the pre-test counselling embedded within this study, important elements for consideration in screening programs. Overall, women supported the availability of being offered screening, although testing before pregnancy was preferred.

2485F

A study in contrasts: The effect of personal genomic testing on perceived risk of melanoma and lung cancer in the PGen Study.

D.A. Carere[1], P. Kraft[1], C.A. Chen[2], L.A. Cupples[3], T. Moreno[4], J. Mountain[5], J.S. Roberts[6], R.C. Green[7], the PGen Study Group. 1) Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115; 2) Data Coordinating Center, Boston University School of Public Health, 801 Massachusetts Avenue, Crosstown 3rd floor, Boston, MA 02118; 3) Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston MA 02118; 4) Pathway Genomics Corporation, 4045 Sorrento Valley Blvd., San Diego CA, 92121; 5) 23andMe, Inc. 1390 Shorebird Way, Mountain View, CA 94043; 6) Department of Health Behavior and Health Education, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109; 7) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 41 Avenue Louis Pasteur, Suite 301, Boston, MA 02115.

Background: In the Impact of Personal Genomics (PGen) study, we collaborated with Pathway Genomics and 23andMe to follow new customers for 6 months, linking individual genetic results with longitudinal survey data. Here we evaluate the impact of genetic results on perceived risk (PR) of two cancers with primarily non-genetic risk factors: lung cancer (PR_L) and melanoma (PR_M). We predicted that an elevated risk genetic result would be associated with an increase in PR and that the effect would be modified by baseline PR. Methods: PR was measured on a 5-point scale from *much lower than average* (1) to *much higher than average* (5) at baseline (pre-results) and 6 months post-results. Genetic results were based on common variants from genome-wide association studies and dichotomized as *elevated risk* or *average/low risk.* Linear regression for change in PR (6 months minus baseline) was performed, adjusting for baseline PR, age, sex, education, income, race, and company. Analyses stratified by baseline PR, and interaction terms between baseline PR and genetic result, were evaluated. **Results:** Among 890 participants for whom data on lung cancer risk was available, the proportion rating themselves at above average risk was constant (12.9% at baseline; 11.4% at 6 months). Genetic result was significantly associated with change in PR_L (p<.0001). In those who received an *elevated risk* result, PR_L increased by an average 0.40 points on the 5-point scale; PR_L was unchanged in those with an *average/low risk* result. The effect of genetic result was greater in those who reported an above average risk at baseline than those who did not, but the difference was not significant. Data on melanoma risk was available for 509 participants, and the proportion rating themselves at above average risk for melanoma fell significantly from 48.5% at baseline to 35.2% at 6 months (p<.0001). Genetic result was significantly associated with change in PR_M (p<.0001), with no effect modification by baseline PR_M. PR_M was unchanged in those with an *elevated risk* result, but decreased by an average 0.61 points in those with an *average/low risk* result. **Conclusions:** The effect of genetic result on PR_L and PR_M is significant but modest. For lung cancer, an *elevated risk* result has a greater influence on PR than an *average/low risk* result, while the opposite is true for melanoma. This finding could be explained by consumers' non-genetic risk factor exposure status, and further analysis is warranted.

2486F

Genetic testing for cardiovascular risk can promote better control over controlled risk factors during one year follow up in women. *O.A. Makeeva, V.V. Markova.* Inst of Medical Genetics, Tomsk, Russian Federation.

Aim of the study was to test a hypothesis that genetic testing for disease predisposition can have bigger impact on controlling existing controlled risk factors for future disease than regular medical recommendation on improving life style. **Methods and samples**: To test the hypothesis two samples of relatively healthy overweight young adults were recruited during routine annual medical examination. First sample (N=95), 70% males, mean age 31±6 years, mean body mass index (BMI) 28.3 ±2.9, undergo medical consultation in respect to reduce their body mass as the obesity is a risk factor for a number of diseases, including cardiovascular disease (CVD). Second sample (N=103), 71% males, mean age 31±6, mean body mass index 29.3±3.5 undergo similar recommendation on body mass lowering and were offered to undergo genetic testing in respect to cardiovascular risk assessment. After informed consent obtained, blood samples were withdrawn and genotyping for several well know risk alleles completed. In a month period subjects were invited for genetic consultation and their risk for CVD communicated. In one-year period subjects from both samples were contacted to invite for the follow up visit. The information about their weight and BMI was collected. **Results**: Weight difference (difference between weight measures at the first and follow-up visits) was analyzed for two samples. Both samples reduced their weight to some extent. Mean weight difference for the group without genotyping was (-1.51±0.35) kg and for the group which undergo genetic risk assessment it was (-2.29±0.45) kg which was not significantly different. When men and women were analyzed separately, it was revealed that in women weight reduction was significantly higher in the subgroup underwent genetic testing (-4.22±0.80 vs -0.82±0.55) kg, p=0.0006, while in men it did not differ (-1.50±0.53 vs -1.81±0.43) kg. In subjects who were genetically tested this association did not depend on a risk estimate (high, low, average risk of CVD). Risk estimates distribution did not differ between men and women. **Conclusion**: Women tend to pay more attention for doctor's recommendation in respect to body mass control when this was followed by genetic risk assessment, but risk estimate itself was not the factor, which had influence. Men generally tend to respond to doctor's recommendation for BMI reduction and genetic testing for the future health problems had no impact.

2487F

Disease Status and Genetic Testing Among Consumers of Two Personal Genomics Companies: Findings from the PGen Study. *S.F. Meisel[1], J. Wardle[1], J. Mountain[2], T. Moreno[3], S.S. Kalia[4], J.S. Roberts[5], R.C. Green[4] for the PGen Study group.* 1) Health Behaviour Research Centre, University College London, London, United Kingdom; 2) 23andMe, Inc; 3) Pathway Genomics; 4) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; 5) Department of Health Behavior & Health Education, University of Michigan School of Public Health.

Background: There is a widespread assumption that risk prediction is a major driver of consumer interest in genetic testing. However, an alternative motivation might be to find out whether existing diseases have a genetic etiology; a phenomenon that has not yet been studied. **Methods:**This analysis examined interest in genetic testing in relation to disease status using data from the PGen (Impact of Personal Genomics) Study, a prospective study in collaboration with two genetic testing companies, 23andMe and Pathway Genomics.Baseline data on current diagnoses, demographics, and family history (FH) was used to examine whether having a particular diagnosis predicted interest in receiving a genetic test result for that condition.Analyses were carried out for 12 conditions included in the genetic profile (Arthritis, Asthma, Heart Disease (HD), High Cholesterol, Diabetes, Obesity, Breast Cancer, Skin Cancer, Multiple Sclerosis, Ulcerative Colitis (UC), Macular Degeneration and Bipolar Disorder). For each condition, we tested whether having a diagnosis was associated with being more likely to express `high' interest in receiving genetic test feedback for the same condition. **Results:** 1801 consumers responded to the baseline survey.Their mean age was 48 years (SD 15.6 years), they were predominantly white (89.4%), just over half were women (60.5%), and nearly half were at least college educated (47.1%). About 80% of them had a diagnosis of at least one of the 12 conditions.The proportion of consumers who expressed high interest in the test results ranged from 28% (UC) to 68% (HD). For each condition, logistic regression analysis revealed that after adjusting for demographic and personal variables and the company offering the test, having a diagnosis of that condition was a strongly significant predictor of interest, with odds ratios (OR) ranging from 1.7 (HD) to 17.9 (Bipolar disorder). FH for each condition had ORs from 1.6 (Arthritis) to 2.4 (Breast Cancer), and effects were considerably smaller for age, sex, ethnicity and education, ranging from 0.6 (Arthritis) to 1.7 (Diabetes). **Conclusions:**These findings suggest that some consumers might seek genetic testing not only for its predictive value, but also to explore the etiology of a condition they already had.This motivation has so far not been described in the debate about clinical utility of consumer based genetic testing, but should be considered as an important factor influencing decisions to seek testing.

2488F

Pathways to and through genetic testing and cancer aged 18-25. *L.M. Hoskins[1], A. Werner-Lin[2].* 1) Clinical Genetics Branch, National Cancer Institute, National Institutes of Health; 2) School of Social Work, New York University.

Background: Much of the extant literature on the psychosocial aspects of BRCA1/2 mutation testing aggregates mutation carriers of all ages in recruitment, analysis, and interpretation. This strategy does not adequately address the needs of the youngest genetic testing consumers, i.e., women aged 18-25. Despite low absolute cancer risk estimates before age 30, BRCA1/2 mutation-positive women aged 18-25 feel vulnerable to a cancer diagnosis but find themselves in a management quandary because the clinical utility of screening and prevention options are not yet well defined for such young carriers. This presentation aims to demonstrate the unique developmental, relational and policy influences, as well as the challenges, experienced by very young BRCA mutation-positive women as they complete genetic testing and initiate cancer risk management. Methods: This study integrated qualitative data from three independent investigations of BRCA1/2-positive women recruited through cancer risk clinics, hospital-based research centers, and online organizations. We present three representative cases, selected from this larger study of 32 BRCA1/2 mutation-positive women aged 21-25 who tested positive for a BRCA1/2 gene mutation between 2 and 60 months prior to data collection. Findings: The first case describes the maturation of a young woman whose family participated in a national cancer registry. The second addresses a young woman who completed genetic testing after learning that her unaffected father was a mutation carrier. The third case highlights a young woman parentally bereaved in childhood who presented for genetic testing due to intense family pressure. Together, these cases suggest that BRCA1/2-positive women aged 18-25 are challenged to reconcile their burgeoning independence from their families with risk-related support needs. Conclusions: Young women who carry deleterious mutations in BRCA1 and BRCA2 face not only increased short-term and lifetime risk of developing cancer, but also a unique set of challenges related to managing cancer risk during an already demanding phase of life. Loved ones acting in ways meant to care for these young women may inadvertently apply pressure, convoluting family support dynamics and autonomous decision-making. Ongoing support from competent healthcare professionals will enable these young women to remain informed and receive objective counsel about their risk- management decisions.

2489F

Moving Up Mastectomies: Emerging Adulthood, Surveillance Fatigue and the Affordable Care Act of 2010. *A. Werner-Lin[1], L. Hoskins[2].* 1) Silver School of Social Work, New York University, New York, NY; 2) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

Significance: Increasingly, women aged 18-25 pursue genetic testing to identify a BRCA gene mutation, despite low absolute cancer risk and the lack of evidence-based risk management options during these years. Complicating risk management decisions are a family context that clouds autonomous election of risk management strategies with prolonged grief and uncertainty. Further, the Affordable Care Act created a public policy context that provides insurance coverage to dependents until they turn 26, enabling them to have genetic services covered. Methods: This study integrated qualitative data from three independent investigations of BRCA1/2-positive women recruited through cancer risk clinics, hospital research centers, and online organizations. Data were originally subsumed as a part of each study cohort. 26 participants aged 18-25 at data collection were re-analyzed. In 2011, 6 BRCA1/2 mutation-positive women aged 21-25 were recruited at the meeting of a web-based advocacy group. All 32 women tested positive for a BRCA1/2 gene mutation between 2 and 60 months prior to data collection. Investigators used grounded theory and interpretive description to conduct both within and cross-study analysis. Findings: Women from all three cohorts expressed the need for clarity in recommendations for screening and prevention before age 25, especially with consideration of early and regular exposure to mammography-based radiation and hormones used in birth control. Parents strongly influenced testing decisions; participants frequently used family providers, which complicated issues of privacy and disclosure. Five participants completed risk-reducing bilateral mastectomy or had one scheduled, four of these reported the timing of the procedures was due to their ability to have care covered by a parent's robust insurance policy, and to the desire to minimize the repeated anxiety surrounding regular breast screening, characterized as `surveillance fatigue.' All reported lifestyle change to support healthy living. Implications: While renegotiating independence from their families, BRCA1/2 mutation positive emerging adult women may be in need of expert guidance from, and prolonged engagement with, health and mental health providers who can address their medical needs and unique psychosocial challenges. The Affordable Care Act may inadvertently endorse rapid risk reducing surgery, earlier than anticipated, before adult dependents `age out' of a parent's insurance plan.

2490F

High satisfaction and low distress after diagnostic whole-exome sequencing in adults. *N. Hoogerbrugge[1], A.S. Sie[1], W.A.G. Zelst-Stams[1], M.R. Nelen[1], H.G. Yntema[1], L. Spruijt[1], J.A. Veltman[1], I. Feenstra[1], J.B. Prins[2].* 1) Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Medical Psychology, Radboud University Medical Center Nijmegen, Netherlands.

Background: Our Human Genetics department is one of the first offering diagnostic two-step whole-exome sequencing (WES). This study aimed to evaluate acceptance, risk perception and distress of WES in adults. **Patients/Methods**: Between August 15th 2011 and June 20th 2012, 213 patients were offered diagnostic WES for colorectal/kidney cancer <40 years (n=51), deafness (n=60), blindness (n=41) or movement disorders (n=61), 177 (83%) accepted. Baseline questionnaires including risk perception and heredity-specific distress (IES) were sent to 141 adults with normal intellectual ability, 111 (79%) were returned. Follow-up after initial WES-results from disease-related gene sets, including satisfaction, is ongoing (currently n=89). **Results**: Baseline responders were diagnosed with: 26% cancer, 27% deafness, 12% blindness, 35% movement disorders. Median age was 49 [22-79], 50% women, 83% with a positive family history. At follow-up, diagnostic suggestive mutations were found in 30%. Nearly all responders (92%) were satisfied with WES. Heredity risk perceptions were similar in baseline and follow-up: believing heredity caused their disease (76% and 69%) and expecting WES to find a genetic cause (49% and 43%). However, significantly fewer patients expected incidental findings (29% versus 18%, p=0.01) and more believed chance likely caused their disease (47% versus 59%, p=0.052). Heredity-specific distress was reported in only 18% versus 17%. **Conclusion**: Acceptance was high in the group offered diagnostic WES after genetic counseling. Adults with normal intellectual ability reported low distress. After initial WES-results, they were highly satisfied. At follow-up heredity risk perceptions were unchanged and fewer patients expected incidental findings, while perceptions of random causality increased.

2491F

A Study of Perception of Health Problems in Patients with Prader-Willi Syndrome by Their Caregivers and the Caregivers' Health Care Behaviors. *W. Khunin, P. Tanpowpong, D. Wattanasirichaigoon.* Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Background: Prader-Willi syndrome (PWS) is a rare genetic disorder characterized by infantile hypotonia, lethargy, feeding difficulty, poor weight gain, followed by excessive eating, morbid obesity, obstructive sleep apnea, endocrinopathies, cognitive impairment, and multiple behavioral problems later in life. Treatments include multidisciplinary medical approach as well as behavioral and environmental modification by caregivers at home, which require excessive understanding and patience. The present study is aimed to explore perception of health problems in patients with PWS by their caregivers and the caregivers' health care behaviors. Methods: This is a descriptive study. Eligible subjects were primary caregivers of patients with confirmed PWS. The caregivers answered questionnaire by mail, or by phone or face-to-face interview. The questionnaire questions comprise demographic data of the caregivers and the children, health problems of the patients, and the caregiver's health care behavior which include regularity of medical follow-up and the major reason of loss follow-up; promotion of the children to take part in social activities; registration for disability benefit card; urgent needs for help; impact to the family members by having a child with PWS. We also asked the effect of having patients with PWS onto their family planning, acceptance for prenatal diagnosis, and choosing option for termination of pregnancy if the fetus was found affected prenatally. Results and Conclusion: Thirty-one caregivers participated the study, including 29 mothers and 2 fathers. The age ranges from 2 to 25 years (mean age 9 years). In regards to health problems, the frequency of health problems that caregivers perceived are as followed: obesity 24/31, sleep problem 13/14, short stature 11/14, undescended testes 8/9, obcessive-compulsive 13/29, developmental delay 21/31, regular follow-up 25/31, register for disability benefit card 19/31, having patient with PWS effect their physical health 15/31 and psychosocial/emotional health 29/31. In term of acceptance for prenatal diagnosis (PND) for the next pregnancy, 28/31 accepted PND and 15/28 would op termination of pregnancy. In conclusion, the study add the value of caregivers' perception which will be essential for tailoring patient care and for helping the caregiver to cope with stress and get support for themselves as well.

2492F

Is it `just like any other test?': Parents experiences with array cGH in pediatrics. *R. Babul-Hirji, N. Hoang, R. Hayeems, R. Weksberg, C. Shuman.* Division of Clinical & Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada.

Background: Understanding families' perceptions of the value and complexity of array cGH testing will enhance its delivery in the clinical setting. This information will also serve to inform the implementation of next-generation sequencing in pediatrics. **Methods:** Sixty-nine families who pursued array cGH testing for children with congenital anomalies +/- delay were invited to participate in semi-structured interviews. Fourteen families accepted, 19 declined, and 44 have not responded. Of the 13 interviews conducted to date, parents represented diverse ethnic backgrounds and the majority were well educated. The age of the child at time of array cGH ranged from 3 days to 21 years with an average age of 6.8 years. Six had an uncertain result, 4 had a pathogenic result, and 3 had a benign result. Qualitative analytic techniques were consistent with interpretive description. **Results:** Preliminary analysis indicates the following themes: (i) microarray test is not readily distinguished from other genetic tests; (ii) uncertainty is experienced regardless of result type and responses to this uncertainty include both hope and frustration; (iii) families articulate a personal utility of array results even if they do not perceive specific utility for medical management, access to support services, and family planning. **Conclusion:** Our finding that families do not distinguish among types of genetic tests should be considered in the context of pre-test genetic counseling. While emerging frameworks for genome-based testing emphasize extensive counseling about the nature and possible outcomes of testing, parents appear to focus on the rationale for testing and the potential for additional diagnostic information. Finally, emerging genetic counseling approaches should be informed by our finding that receiving uncertain information does not appear to compound the intrinsic uncertainties associated with caring for a child with a rare or undiagnosed genetic condition, but this merits further research.

2493F

Impact of a brief theory-driven intervention on family communication about cancer family history in a diverse urban clinic. *J. Bodurtha[1], D. Bowen[3], J. Borzelleca[2], R. Corona[2], M. Gyure[2], A. Krist[2], A. Maibauer[2], D. McClish[2], V. Rodriguez[2], J. Quillin[2].* 1) Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Virginia Commonwealth University, Richmond, VA; 3) School of Public Health, Boston University, Boston, MA.

It is recognized that the public generally understands the importance of family health history (FH), but its translation from a topic of family discussion to a driver of preventive care is suboptimal. Women, as the family `kinkeepers' are positioned to gather FH and share cancer (CA) prevention information. Family discussion of FH can open doors to genetic understanding and impact patient-centered care that incorporates differing risks based on FH. The KinFact (R01 CA140959) intervention was based in the Coordinated Management of Meaning (communication is a process of coordinating actions with others and making/managing meanings) and Daly's 6-step communication process (Prepare, Ask relatives - How much they know? How much do they want to know? Share information, Respond to feelings, Plan and follow-through). We tested in a randomized trial whether a 20-minute individual CaGene pedigree-informed educational intervention in an urban academic women's health clinic could impact women's subsequent gathering and sharing of CA FH information. Control participants received an informational CA brochure. 490 adult women not selected for CA risk were randomized and followed at 1, 6, and 14 mos. Mean age was 33.4 yrs, 59% were African-American, ~1/3 were uninsured, and 1/3 were pregnant. 74% of women had at least one 1st or 2ndo relative with a CA FH. There were no baseline differences in demographics or + CA FH. In a preliminary analysis more intervention than control women gathered FH information by 14 mos (47% vs. 36%, p=0.036). Controlling for race, the REAL-G genetic literacy measure, and CA FH, the intervention effect remained (p=0.0176). Also more intervention than control women shared CA information with relatives by 14 mos (44% vs. 33%, p=0.0162). Controlling for race, genetic literacy and cancer FH, the intervention effect remained (p= 0.0168). In addition the intervention was equally effective for women who were African-Americans or of other races, as well as of varying genetic literacy or CA FH. The KinFact intervention undertaken in a population at risk for disparities in genetic services may be useful in enhancing family discussions of CA risk information as an important step in promoting patient- and family-centered outcomes and adaptable and scalable in other settings. Educating women to enhance their `kinkeeping' skills may allow them to partner more effectively with their families and providers in discussing CA risk and prevention.

2494F

Cancer Genetics Referral Patterns of Physicians and Patient Socio-Demographics. *J. Cohn, W. Blazey, S. Koehler, B. Laurent, V. Chan, M. Jung, D. Tegay, B. Krishnamachari.* NYIT College of Osteopathic Medicine, Old Westbury, NY.

BACKGROUND: Genetic testing can aid with the selection of cancer screening and preventive health services in high-risk patients. Studies demonstrate that physicians are more likely to order genetic testing based on patient inquiry rather than on recommended guidelines. Literature also shows that there are racial disparities in patient use of genetic services that do not appear to be explained by differences in socioeconomic factors. The association between patient use of genetic services and physician behaviors is not well studied. METHODS: The goal of this study was to investigate racial and socio-economic disparities in cancer genetic risk assessment and referral patterns in clinical practice. A survey was administered to physicians (predominantly primary care physicians) evaluating physicians' awareness of the racial and socioeconomic details of their patients, their genetic risk assessment and referral patterns. RESULTS: 139 physicians completed the survey. 26.6% (n=37) of the physicians reported conducting frequent genetic risk assessment and 44.6% (n=62) reported frequently referring patients for genetic services. 13.8% (n=19) reported that at least 25% of their patients were African American, and 26.8% (n= 37) did not know what percentage of their patients were African American. Additionally, 35.8% (n=59) did not know what percentage of their patients were Latino/Hispanic and 34.9% (n=58) did not know what percentage of their patients were classified as living at poverty levels. Those who conducted frequent risk assessment were more likely to state that they would refer a non-English speaker for genetic testing regardless of language barriers (OR=8.34, 95% CI: 1.08,64.69, p<0.05). Those who frequently referred patients for genetic services were more likely to state they would refer a Medicaid patient for genetic testing even when they thought the testing costs would not be covered (OR=2.44 , 95% CI: 1.20, 4.99, p<0.05). CONCLUSIONS: Most research has focused on disparities in genetic services from the perspective of patient preferences and behaviors. In this study, about a third of physicians surveyed reported not knowing the socio-demographic breakdown of their patient populations and may not be conscious that some of their patients come from socio-demographic groups less likely to pursue genetic services on their own. Further research must focus on clinical practice and physician behaviors.

2495F

Prenatal chromosome SNP microarray analysis: Genetic counseling issues and dilemmas. *H. Cabral[1], S. Schwartz[1], J. Tepperberg[1], I. Gadi[1], V. Jaswaney[1], R. Burnside[1], K. Phillips[1], E. Keitges[2], H. Risheg[2], V. Potluri[3], R. Pasion[1], B. Rush[1], H. Taylor[1], L. Kline[1], J. Shafer[1], P. Papenhausen[1].* 1) Laboratory Corporation of America, Reseach Triangle Park, NC, United States; 2) LabCorp/Dynacare, Seattle, WA, United States; 3) LabCorp/Dyna-Gene, Houston, TX, United States.

Prenatal microarray analysis has been employed for a number of years, but its popularity has grown exponentially over the past two years, especially with the publication of the NICHD study. For ~2 ½ years we have utilized an Affymetrix SNP microarray platform to study over 8,000 prenatal patients. This work has highlighted some of the counseling issues that have become important with this analysis that is not part of traditional counseling for cytogenetic analysis. This includes the possible detection of uniparental disomy (UPD), suggestion of consanguinity and the detection of small copy number gains and losses and determination as variants of unknown significance (VOUS), pathogenic changes or recessive disorders. Most problematic are the small copy number changes detected and the lack of follow-up by families to better delineate whether these changes are normal variants or pathogenic changes. When abnormalities were detected as VOUS, there was family follow-up in ~ 79.7% of cases, but when the change was thought to be pathogenic, only 57.1% had follow-up. For patients referred because of AMA, anxiety or a positive maternal serum screen, there was 66.7% family follow-up after the detection of an abnormality, but only 59.1% family follow-up in patients ascertained due to ultrasound abnormalities. Patients with suspected UPD had family follow-up 50% of the time while pregnancies with structural chromosomal abnormalities subsequently studied by the array had parental studies in 45% of the cases. This work has highlighted a number of counseling issues including: (1) Need for pre-test counseling delineating the ramifications of the potential findings; (2) Importance for families to indicate the potential relationships among the parents prior to the information being detected in the analysis; (3) Lack of follow-up when potential UPD is identified, especially as >1/3 of suspected cases are not confirmed; (4) Importance of parental follow-up to differentiate VOUS from pathogenic abnormalities when a copy-number change is identified (5) Difficulty involved when normal carrier parents are identified in cases of new microdeletion syndromes associated with autism and other abnormalities; (6) Realization for the need to include microarray technology into Genetic Counseling programs to help prepare new graduates for the complexities involved; and (7) Critical need for trained laboratory genetic counselors to assist in the dissemination of the results to providers.

2496F

Offspring Risk Perceptions: Adolescents and Young Adults with Congenital Heart Disease Agree with their Parents (and both are wrong!). *S.M. Fitzgerald-Butt[1, 4], K.M. Fry[1], A.N. Zaidi[2,4], C.A. Gerhardt[3,4], V. Garg[1,2,4], K.L. McBride[1,4].* 1) Center for Cardiovascular and Pulmonary Research, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) The Heart Center, Nationwide Children's Hospital, Columbus, OH; 3) Center for BioBehavioral Health, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 4) Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, OH.

As individuals with congenital heart disease (CHD) now typically survive into adulthood, the risk of CHD in their offspring is a pertinent concern. Parents are the typical source of medical information through childhood but adolescents and young adults (AYA) must have sufficient knowledge to assume responsibility for their medical care. Therefore, we examined offspring risk perceptions using both categorical (below average, average, above average) and continuous (0-100%) measures, as well as their associations with demographic characteristics, knowledge of heart defect name, CHD diagnosis complexity, CHD perceived severity and general genetic knowledge in AYA and their parents. Participants included 196 AYA, 15-25 years old with structural CHD (mean age=19.0 years, 54% male, 85% Caucasian) and 179 parents (mean age=47 years, 34% male 94% Caucasian) who were recruited from an outpatient cardiology clinic (85% AYA consent rate). All participants were asked the name of their/their child's CHD and completed measures of demographics, genetic knowledge, perceived CHD severity, and perceived risk for an offspring with CHD. CHD complexity was rated as simple, moderate, or great. Categorical perceptions of risk were similar in AYA and parents with only approximately a third of participants endorsing higher risk (34% AYA, 35% parents) while two-thirds endorsed average or lower risk. The perception of being in the high risk category was associated with higher genetic knowledge (p<.001, p=.018) and higher perceived CHD severity in a hypothetical baby (p=.039, p=.001) in both AYA and parents, while Caucasian race (p=.008) and higher median household income (p=.001) were associated in just the AYA and perceived future heart defect severity (p=.016) was associated in just the parents. The ratings of risk on a continuous scale were extremely variable and remarkably high among both the AYA and parents, ranging from 0-100% with a mean of 36.8% (*SD*=24.3%) and 34.1% (*SD*=23.5%), respectively, with a mode of 50% and increased selection of 25% and 75% in both groups. The majority of AYA and parents have an inaccurate categorical risk perception and their continuous risk perception may represent poor numeracy skills and either lack of or inaccurate knowledge of risk. These results highlight the need to provide accurate offspring risk information to both AYA and parents, possibly while also providing additional genetic and numeracy education.

2497F

Emerging changes in genetic counseling and reproductive decision making based on AGG mapping of fragile X carriers. *A.G. Hadd[1], A. Glicksman[2], J. Coppinger[1], N. Ersalesi[2], J. McCarver[1], E. Blatt[1], W.T. Brown[2], J. Skeen[1], G.J. Latham[1], S.L. Nolin[2].* 1) Research and Technology Development, Asuragen, Inc, Austin, TX; 2) NYS Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

Fragile X carriers of intermediate and premutation alleles are increasingly identified as the result of broader guidelines for fragile X testing, increased participation in genetic screening and improvements in PCR technology. Consequently, more women are identified who have ambiguous risks for expansion to a full mutation in their offspring. Recent studies have demonstrated that the presence of AGG interruptions correlates with substantially reduced risk for transmitting a full fragile X mutation. We report the risks for full mutation expansions based on AGG analysis from over 2500 samples and 1000 FMR1 transmissions and highlight case studies of how these individualized risks changed genetic counseling and reproductive planning. Logistic regression was used to determine the predictors of allele instability (any change in repeat length) and risk for expansion to a full mutation (>200 CGG repeats). Consistent with prior data, the risk of expansion was associated with the total maternal repeat length and number of interspersed AGGs. A regression curve was used to establish risks and 95% confidence intervals for each repeat length and number of AGGs. In a parallel study, demographic and ordering information, prior fragile X testing results, and post-AGG analysis outcomes were obtained from a subset of 87 consecutive clinical samples received for AGG testing. The majority of providers ordering testing were from Infertility/Reproductive clinics, and indicated that patients use AGG testing to make decisions about IVF.In these 87 samples, the majority of carriers had reduced post-test risks for transmitting full fragile X mutations. Five premutation carriers, however, had increased post-test risk, such as a woman with 68 CGG but no AGG whose a priori risk of expansion to a full mutation increased from approximately 5% based on repeat length alone to approximately 27%. These findings demonstrate the ability to individualize the risk for transmitting a full mutation expansion. The refined risk estimates have implications for genetic counseling and provide important information to patients regarding reproductive decisions. These findings suggest ascertainment of AGG status should be considered as part of the work-up of women of reproductive age who have FMR1 alleles in the 45-90 CGG range.

2498F

A Two-Stage Approach to Genetic Risk Prediction of Breast and Ovarian Cancer. *P.M. Atienza[1,2], J. Chipman[3], K. Hughes[4], C.I. Amos[5], B. Arun[6], G. Parmigiani[3,7], S. Biswas[8].* 1) School of Public Health - Biostatistics Department, University of North Texas Health Science Center, Fort Worth, TX; 2) Alcon Research, Ltd., Fort Worth, TX; 3) Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA; 4) Massachusetts General Hospital, Boston, MA; 5) Department of Community and Family Medicine, Geisel College of Medicine, Dartmouth College, Hanover, NH; 6) Department of Breast Medical Oncology and Clinical Cancer Genetics, University of Texas M.D. Anderson Cancer Center, Houston, TX; 7) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 8) Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX.

BRCAPRO is widely used in genetic counseling to identify patients at high risk for hereditary breast or ovarian cancer. However, it requires extensive information on the patient's family history, which is a hindrance for its use in many health care settings. Recently, we proposed simplified versions of BRCAPRO for settings that do not require exhaustive genetic counseling (Breast Cancer Research and Treatment, 139:571-579, 2013). As simpler versions lead to loss in accuracy as compared to BRCAPRO, here we propose a two-stage approach to balance the tradeoff between the amount of information used and the accuracy achieved in assessing the risks. In the first stage, only a limited amount of family history will be analyzed using simpler versions of BRCAPRO. If the risk at this stage is sufficiently high, BRCAPRO, with exhaustive family information, will be used in the second stage to potentially achieve more accurate predictions. We consider four first stage tools: (1) BRCAPROLYTE uses information on relatives up to the second degree but for those affected with breast and/or ovarian cancer only; (2) BRCAPROLYTE-Plus additionally imputes the ages of unaffected relatives based on an external data source; (3) BRCAPROLYTE-Simple imputes the family structure and the ages of unaffected relatives up to the second degree; (4) BRCAPRO-1Degree uses information on only the first-degree affected relatives. The second stage is BRCAPRO for all these tools. We develop a methodology to evaluate the overall performance of a two-stage approach. In particular, we compute sensitivity, specificity, and predictive values, and area under the ROC curve (AUC) of a two-stage approach by considering different combinations of the first and second stage cutoffs. We also evaluate the clinical implications of using a two-stage approach. We use 2,713 probands from seven sites of the Cancer Genetics Network and MD Anderson Cancer Center to evaluate our approach. The proposed two-stage approach has comparable discrimination and calibration as BRCAPRO if the latter is applied to all probands. The overall AUCs are 0.782 (BRCAPROLYTE), 0.783 (BRCAPROLYTE-Plus), 0.782 (BRCAPROLYTE-Simple), and 0.775 (BRCAPRO-1Degree) while the AUC for BRCAPRO is 0.783. Also, the two-stage approach substantially reduces the genetic counseling and testing burden. Thus, we conclude that this approach can be adapted for genetic risk prediction of breast or ovarian cancer on a large scale.

2499F

Clinical implications of Variants of Unknown Significance in Chromosomal Microarrays in pregnancies in Israel. *H. Yonath[1, 2], S. Rienstein[1], J. Shamash[1], M. Berkenstadt([1], M. Dicastro[1], S. Eisenberg- Brazilai[1], N. Goldstein[1], M. Frydman[1,2], E. Pras[1,2].* 1) Inst Human Gen, Sheba Med Ctr, Ramat Gan, Israel; 2) Sackler Medical School, Tel-Aviv University, Tel-Avivi, Israel.

Background: One of the main concerns in finding a Variant of Unknown Significance (VOUS) in Chromosomal Microarray (CMA) during pregnancy is the implication it would have on the decision to continue or terminate it. Many couples in Israel have a low tolerance for abnormalities detected during pregnancy and prefer to terminate a pregnancy if an abnormality is found. The goal of this study was to evaluate the decision made by the parents regarding pregnancies in which a VOUS was found, in a large genetic institute in Israel. Methods: 305 prenatal CMA cases were evaluated. The arrays were oligonucleotide specially targeted prenatal arrays (Agilent arrays designed by Medical Genetics Laboratory at Baylor College of Medicine). The significance of the variants was determined based on local guidelines that are very similar to the ones described by Wapner et al (1). Parental CMA were performed in order to determine whether the finding is inherited or de novo. Genetic counseling was provided to the couples. Results: VOUS were found in 36 cases (12%), About half of them were inherited. And the indications were mainly abnormal US and parental concern. None of these pregnancies were terminated. Conclusions: Though there is a high concern that VOUS in CMA would lead to the termination of pregnancy, our results show that if there is no clinical data that suggests that the abnormality is pathogenic, and with the proper genetic counseling, the results do not lead to termination of the pregnancy. References: (1) Chromosomal microarray versus karyotyping for prenatal diagnosis.Wapner RJ et al. N Engl J Med. 2012.

2500F

A gift to the children - genetic testing at the end of life. *M. Mikhaelian[1], K. Stears[2].* 1) InVitae Corporation, San Francisco, CA; 2) St. Vincent's Hospital, Billings, MT.

INTRODUCTION: Recent publications discuss targeted testing for patients dying of ovarian cancer. Less is known about end of life testing for a patient with a complex history that does not directly point to a specific genetic disorder. Here we present the case of a patient who was seen for genetic counseling in order to learn what genetic mutations she may pass onto her descendants. The patient was consented for and tested for 262 genetic conditions by InVitae Corporation. CASE REPORT: The patient is a 81-year-old terminally ill woman with a history of peripheral vascular disease, pulmonary emboli, stroke, melanoma, breast cancer, calcium deposits, spinal stenosis, hyperlipidemia, hypertension and peripheral artery disease. She has three children. She has a family history of vascular disease, liver disease and cancer in first degree relatives and sudden cardiac death in a second degree relative. She did not have any previous genetic testing. The patient was interested in testing to provide carrier information for her descendants. Using the InVitae Assay, the patient tested for 262 conditions, including cardiac arrhythmias, a 50 gene cancer panel, clotting disorders and multiple other inherited conditions. A homozygous mutation for HFE related hereditary Hemochromatosis was found. The results prompted an additional Hemochromatosis evaluation, which was negative. DISCUSSION: The testing provided relevant information to the patient's children about the risk of Hemochromatosis and the reduced relative risk of inherited arrhythmias, clotting disorders and cancer predisposition. Broad genetic testing can be used as a supplement to DNA banking in a patient with a complex personal and family history. Issues around obtaining informed consent and genetic counseling about results are discussed. A description of the genetic counselor's personal experience with this case is also detailed.

2501W

Using the Cleveland Clinic Score to predict for germline *PTEN* mutations in the analytical algorithm of Cowden syndrome: a cost effectiveness study. *J. NGEOW[1], J. Mester[1], C. Eng[1, 2, 3].* 1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Oh; 2) Taussig Cancer Institute, Cleveland Clinic, Cleveland, Oh; 3) Department Of Genetics And Genome Sciences, And Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Oh.

Background: Cowden syndrome (CS) patients with germline *PTEN* mutations are at increased risk of benign and malignant features. Our group has developed an online risk calculator, the Cleveland Clinic Score (CC Score) to determine the patient's likelihood of harboring a germline *PTEN* mutation. The higher the CC score, the higher the probability of an underlying germline *PTEN* mutation. CC Score of 10 corresponds to an estimated 3% pathogenic *PTEN* germline mutation risk. Issues that affect clinical testing include sensitivity and specificity of the test, the benefit to the patients and the possible negative ramifications of the results, and the cost. The aim of this study was to assess the cost-effectiveness of using different CC Score thresholds to improve the yield of the diagnostic algorithm of CS. **Methods**: Data from an existing multicenter prospective study in which 3541 probands satisfying relaxed CS clinical criteria were used. *PTEN* mutation scanning, including promoter and large deletion analysis, was performed for all subjects. A decision model was developed to estimate the number of mutation carriers and the incremental costs of alternative case-finding methods for detecting *PTEN* mutation carriers. Strategy 1 involved *PTEN* testing of all individuals meeting relaxed CS clinical criteria. Strategy 2 involved *PTEN* mutation testing only in individuals with CC Scores ≥ 10. Incremental costs of using different CC Score thresholds (15, ≥20) for detecting *PTEN* mutation carriers were additionally evaluated. For each strategy, the number of individuals tested, mutations detected, missed, false positive results were computed. One-way sensitivity analysis was performed to assess robustness of estimations. *Results*: Pathogenic *PTEN* mutations were identified in 250 individuals (7.1%). Sensitivity for detecting *PTEN* mutation carriers using CC Score cutoff of 10 (strategy 2) was 83% and specificity was 71%, missing 49 patients with germline *PTEN* mutations. Average cost per *PTEN* mutation detected using Strategy 1 was $21246 compared with $8455 for Strategy 2. The incremental cost per additional mutation detected using Strategy 1 (all patients tested) was $73714 **Conclusions**: Use of clinical risk calculators such as the CC Score is a cost-effective pre-screening method in the selection of patients for *PTEN* germline analysis when CS is suspected.

2502W

Genetic tests evolution in the Genomic era. Is cost evaluation a relevant factor in health care planning? *D. Coviello[1], C. Lanza[1], A. Seri[1], M. Parodi[1], P. Casale[2], A. Fabbri[2], S. Casati[3], M. Esposto[2].* 1) Laboratory of Human Genetics, E.O. Ospedali Galliera, Genoa, Italy; 2) Management Control Unit, E.O. Ospedali Galliera, Genoa, Italy; 3) UNIAMO, Italian Patient Federation on Rare Diseases, Venice, Italy.

In the last thirty years, medical genetics has improved molecular diagnosis according to the evolution of gene technology, but it is really in the last five years that the new revolution of next generation sequencing (NGS) has open a new window not only on research but also on diagnostic testing. Laboratories performing routine genetic testing started their activity using mainly home made protocols derived from research studies. Now the scientific framework and the laboratory activities are completely changed: the number of known disease genes has increased, automation is applied also to genetic laboratories, certification and accreditation procedures of genetic services are in place, information and awareness among public and all stakeholders are crucial. Purpose of the study: both public National Health systems (NHS) and private laboratories or companies need to deal with costs of genetic testing performed with traditional methodologies or with the new strategies derived by the introduction of NGS. Since it has not been clearly defined how many genomic tests have enough clinical utility, the investigation of their costs could be a way to establish a correct public health policy. Methods used: Activity-based costing (ABC) is a methodology that identifies activities and assigns the cost of each activity taking into account resources to all products and services according to the actual consumption by each. This methodology has been applied to the steps of diagnostic processes of our genetic laboratory. Summary of results: The analysis of the activities has identified a number of indicators to assess the workload for every professional who participates in the process of diagnosis. A second part of the analysis has evaluated the proportion of material used for each activity (test or portion of the process). All these parameters have been incorporated into a software able to split all the lab costs (personnel, material and general costs) for each test provided. This tool can be used in several ways: it can be useful in comparing the costs among different laboratories, in comparing the performance of the same laboratory in subsequent years or to make a priority list of genetic/genomic tests to provide, taking into account costs/benefits data.

2503W

Innovation in genomic medicine to realize the bioeconomy in Mexico. *S. March[1,2], F. Valdez-Ortega[1,2], G. Soberon[2,3], J. Frenk[2,4], G. Jimenez-Sanchez[1,2,4].* 1) Global Biotech Consulting Group (GBC Group), Mexico; 2) Genómica y Bioeconomía AC, Mexico; 3) Mexican Health Foundation, Mexico; 4) Harvard School of Public Health, Boston, MA.

A bioeconomy is one based on the use of research and innovation in the biological sciences to create economic activity and public benefit. It is a large and rapidly growing segment of the world economy that provides substantial public benefit. The growth of the bioeconomy is due in large part to the development of genomics. In Mexico, genomic sciences are vigorously developing with a robust infrastructure and training programs (Jimenez-Sanchez G, el al. Design and implementation of a platform for genomic medicine in Mexico. In: Genomics and health in the developing world.(D Kumar Ed, Oxford University Press, 2012). To stimulate strategic synergies for innovation in genomics, we established Genomica y Bioeconomia, a non-for-profit organization led by experienced leaders in genomics, health policy, economy, business development, legal and social sciences (www.genomicaybioeconomia.org). Our initial results include a comparative analysis of knowledge-based economic initiatives in emerging economies including China, India, and Brazil. This led to specific recommendations as to how an emerging economy like Mexico can orient major genomic programs into innovation and economic development (Jimenez-Sanchez G, et al. Genomics and Bioeconomy: A window of opportunity for economic growth in Mexico. COLMEX-COLNAL, 2012). Genomic medicine is one of the key areas where previous work and novel developments in science and industry make it necessary to establish the grounds for implementation of genomic applications into medical practice. This is the case for pharmacogenomics, where scientific evidence and recommendations by regulatory agencies make it important to develop cost-benefit analyses for some of the most commonly used tests. We initiated an economic evaluation of the Warfarin sensitive test in the context of net present values for Mexico. Initial evidence indicates that this test could be used in the public health arena. Based on current technologies and the knowledge of genetic variation in the Mexican populations, we established a join effort with the Mexican Government to develop a regulatory framework for pharmacogenomics and the infrastructure required at the COFEPRIS (FDA equivalent) laboratory facilities. Results from this initiative include a nation-wide network to establish synergies between academy, industry and government to develop innovation in genomics oriented to public health challenges of the Mexican population.*Supported in part by CONACYT.

2504W

Awareness of genetic breast cancer epidemic in Iceland. *V. Stefansdottir[1], J.J. Jonsson[1,2,3], O.Th. Johannsson[4].* 1) Medical Genetics, The National Univ. Hosp. of Iceland, Reykjavik, Reykjavik, Iceland; 2) Dept. of Biochemistry and Molecular Biology, University of Iceland; 3) The Genetical Committee of the University of Iceland; 4) Dept. of Medical Oncology, The National Univ. Hosp. of Iceland.

Following Angelina's Jolie statement on her BRCA1 status, awareness of possible BRCA mutation in families was considerably raised. Less than a day after the statement, the CEO of deCode Genetics announced that from its research, the company had data on all BRCA2 mutation carriers in Iceland. Identifying them in their coded databases would be an easy task if permission to do so from the authorities would be granted. Also that it was a duty to the population to inform all at risk about their genetic status. A `genetic storm' broke out. Over the next few days over 200 people called or emailed our cancer genetic counseling clinic (CGC) for information about the BRCA genes and genetic counseling. And the media caught on. In Iceland, a country with 320.000 inhabitants, two founder BRCA mutations are known; the BRCA1 c.5074G>A, and the BRCA2 c.771_775del5. The prevalence of the BRCA1 mutation is unknown. The prevalence of the BRCA2 mutation is thought to be 0.6%. Cancer genetic counseling is done by the aid of an electronic genealogy database for all Icelanders back to 1850 at least and The Icelandic Cancer Registry, a population based nationwide cancer registry going back to 1911 for breast cancers and 1952 for other cancers. Use of the database and registry, based on presumed consent, enables for very fast and accurate workup and risk assessment. The CGC unit at the National Landspitali University Hospital has been in operation since December 2006. One genetic counselor and one part-time oncologist work at the unit. From December 2006-January 2013, 700 individuals have come for counseling. the majority of BRCA2 families are thought to have been identified and all BRCA1 families. As families can be traced very far back and the initial BRCA2 carrier has been identified from the 16th century, one question is always: `What part of Iceland is your family from? When a BRCA mutation is found, the carrier gets an information letter about the mutation and the genetic counseling service to distribute to his family. We describe how a very small CGC unit coped with a very big workload and complex counseling issues in a very short time.

2505W

My46: a genetic counseling extender. *K.M. Dent[1], S.M. Jamal[2], J.H. Yu[2], H.K. Tabor[2,3], M.J. Bamshad[2,4].* 1) Department of Pediatrics, University of Utah, Salt Lake City, UT; 2) Department of Pediatrics, University of Washington, Seattle, WA; 3) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA; 4) Genome Sciences, University of Washington, Seattle, WA.

My46 (http://www.my46.org) is an innovative web-based tool designed to enable self-guided management of genetic testing results. This online tool empowers individuals and families to make decisions about what genetic test results might be of utility to them and their children, and why they might want or not want such results. A primary focus of My46 is to educate individuals about genetics, risk of transmission, and medical conditions thereby providing genetic counselors more time to focus counseling sessions on interpreting genetic results and exploring their impact. To facilitate this aim, My46 has developed an open-access `Learning Center' that includes standardized summaries of a wide range of medical conditions written in patient-friendly language (`Trait Profiles'); key considerations to think about before receiving genetic testing results (`What you should know about genetic testing'); an introduction to human genetics with an extensive glossary of terms; and links to external resources for families and healthcare professionals (`More Resources'). The Trait Profiles can be searched by trait name (e.g., sickle cell disease) and are organized into seven categories (disease risk, carrier status, medication response, genetic syndromes, metabolic disorders, newborn screening conditions, and ancestry). The Trait Profiles include information on characteristics of the condition, diagnosis/testing, management/surveillance, genetic counseling, special consideration, resources with direct links to GeneReviewsTM, and key references. All Trait Profiles can be printed as pdf documents. Each Trait Profile is written by a genetic counselor, edited by the My46 team for readability, and reviewed by a clinical geneticist. The goal is to write a Trait Profile for each condition for which genetic testing is available. To date, the My46 team has received over 120 draft Trait Profiles, written by genetic counselors from around the world. As a result, the Trait Profiles broadly capture the expertise of genetic counselors in a single repository, greatly benefiting vast numbers of patients. Collectively, the My46 Learning Center is a valuable resource to the genetic counseling community and will enhance the delivery of genetic counseling services.

2506W

Engaging health professionals in evaluations of emerging genomic technologies. *C. Catley[1], J. Little[1, 2], S. Nicholl[1], H. Etchegary[3], J.C. Carroll[4], D. Castle[5], L. Lemyre[6], B.K. Potter[1], B.J. Wilson[1], CIHR Emerging Team in Genomics in Screening.* 1) Department of Epidemiology and Community Medicine, University of Ottawa, ON, Canada; 2) Canada Research Chair in Human Genome Epidemiology; 3) Clinical Epidemiology, Memorial University Newfoundland, St John's, NL, Canada; 4) Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 5) ESRC Innogen Centre, University of Edinburgh, Edinburgh, UK; 6) School of Psychology & Institute of Public Health, University of Ottawa, ON, Canada.

Background: The successful integration of emerging genomics technologies into mainstream health care requires their acceptance and adoption by professionals across many disciplines and specialties. It is well established that interventions of known effectiveness are not used as intended, and that ineffective or harmful interventions continue to be used inappropriately. We need insight into factors which influence professionals' own evaluations of genomic technologies, to guide effective and responsible implementation strategies. A major threshold in the adoption of new knowledge (and practices) is that of salience - perceived relevance to individual practice. We are conducting a pilot study of a method for engaging health professional audiences in evaluations of emerging genomic technologies which takes this into account. **Objectives:** To (a) understand how health professionals frame their understanding of emerging genomic technologies and (b) identify specific factors that may facilitate and challenge the appropriate adoption of these technologies in professional medical and nursing practice. **Methods:** A structured interactive workshop combines a largely didactic component (describing the range of genomic approaches in health care) and an interactive component which presents hypothetical case studies of different potential applications of genomics in routine practice. Participants engage in moderated discussion, and are encouraged to capture personal reactions to the content through ongoing written narrative. Non-participant observers capture discussion points in real time. Survey data are captured before, during, and post-workshop. A mixed methods approach is taken for data analysis. **Results:** We have conducted workshops with physicians and nursing professionals in primary care and a range of specialties. Results summarizing participants' evaluations of the utility of genomic approaches, and shifts in key attitudes compared with pre-workshop baseline, will be presented. The necessity of prompting participants' evaluation of the salience of genomic approaches generally, as a prerequisite for meaningful reflection on specific applications, will be discussed. We will also discuss the challenge of developing a workshop format that is efficient in its use of participants' time, and the relationship between professional engagement and continuing professional education exercises.

2507W

Uptake of a web-based patient-entered cancer family history collection tool. *M. Doerr[1], S. Griffith[2], C. Eng[1].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH.

Background: MyFamilyHealthHistory (MyFHH) is a family history collection tool developed by Cleveland Clinic (CC) for use in oncology settings. MyFHH is a CC quality improvement initiative to increase the efficiency of cancer family history collection without introducing care disparity. We have previously shown patients who complete MyFHH are more likely to attend cancer genetics clinic appointments (93% vs. 63% attendance in those not completing). We thus investigate determinants of MyFHH uptake. Methods: Between Aug 2009 and Sept 2012, 1161 patients scheduled appointments triggering MyFHH invitation. Cancer-specific family history is obtained for first and second degree relatives. Univariable/multivariable analysis of the association between completion status and personal history of neoplasm, sex, age and socioeconomic status (SES) was performed. Demographic data was extracted from the electronic medical record; median household income by ZIP Census Tabulation Area was used as a proxy for SES. Results: 359 (31%) completed MyFHH. Of 1161 invited, 877 (76%) had a personal diagnosis of neoplasm, 1002 (84%) were female, and 994 (87%) were age <65 yrs. In univariable analysis, we did not find evidence of a difference in the odds of completing MyFHH based on personal diagnosis of neoplasm (p= 0.353). Men showed a reduction in the odds of completing MyFHH as compared to women (OR 0.69; 95%CI 0.46, 1.04), although the difference between the groups was not statistically significant (p=0.097). Notably, individuals age ≥65yrs were significantly less likely to complete MyFHH as compared with those age <65yrs (OR 0.47; 95%CI 0.31, 0.71; p<0.001). There was no difference in the odds of completing MyFHH based on SES (p=0.513). On multivariable analysis, after adjusting for personal diagnosis of neoplasm, sex, and SES, individuals age ≥65yrs remained significantly less likely to complete MyFHH as compared to those age <65yrs (OR: 0.48; 95%CI: 0.32, 0.72; p<0.001). Conclusions: MyFHH uptake does not vary by personal diagnosis of cancer, sex, or SES. Our study identified patients age ≥65yrs as less likely to use MyFHH than those <65yrs. Future efforts to facilitate patients ≥65yrs use of MyFHH could lead to improved cancer genetics clinic attendance.

2508W

Perspectives on Universal Screening for Lynch Syndrome in a Managed Care Setting. *J.V. Davis, T. Kauffman, J. Reiss, C. McGinley, K. Arnold, M. Gilmore, K.AB. Goddard.* Science Programs, The Center for Health Research, Kaiser Permanente, Portland, OR.

Background: The Hereditary Non-polyposis Colorectal Cancer (HNPCC) study is a clinical trial conducted in an integrated health-care system~Kaiser Permanente Northwest (KPNW). Our goal is to determine whether universal screening for Lynch Syndrome among all newly diagnosed colorectal cancer cases increases the use of genetic counseling, compared with the current practice of physician-based referrals or self-referrals to medical genetics. In this phase of the trial, we interviewed health plan leaders and providers to better understand requirements for successful implementation of universal screening. Methods: We investigated factors that might hinder or facilitate implementation of universal screening for Lynch Syndrome in all newly diagnosed colorectal cancer patients. We conducted fifteen semi-structured interviews with KPNW leaders and key staff in pathology, oncology, medical genetics, gynecology, surgery, and laboratory services. Using NVivo software, we applied thematic analysis to the interview audio transcripts. Results: Informants identified 10 departments that should be represented in planning and implementation decisions. Key decisions to be made include: patient selection criteria, consent protocols, choice of Microsatellite Instability (MSI) or Immunohistochemistry (IHC) testing~or both~for initial testing, laboratory selection, and whether to include patients with endometrial cancer in the screened population. These decisions will influence several potential barriers and issues. Justification of screening decisions based on quantitative benefit analysis will be necessary, as will tracking mechanisms for orders, results, and surveillance follow-up. Training and workload burden are perceived as potential concerns in some departments. Clinicians consistently supported systematic testing, suggesting a broader belief within the organization that it is `the right thing to do.' Conclusion: Successful implementation of effective systematic screening for Lynch Syndrome is contingent on favorable benefit analysis and buy-in and coordination of many departments. Program development should include assessment of alternate screening strategies. Barriers must be addressed to achieve a process that is both well integrated and cost effective.

2509W

The awareness and need of genetic counseling service in Korea. *H. Kim*[1,2]. 1) Medical Genetics, Ajou Univ Med Sch, Seoul, South Korea; 2) Chairman, Korean Foundation for Rare Disease(KFRD), Seoul, South Korea.

Genetic counseling service which is well recognized as an integral part of clinical genetics service deals with diagnosis and management of genetic condition as well as genetic information presentation and family support in the developed countries of world. Korean Health care system is known for providing one of the most cost-effective services in the world. It is covered by the uniform national health insurance policy for which most people in Korea are mandatory policy holders. The necessity of genetic counseling service has been recently recognized by Korean medical communities, however, genetic counseling as an integral part of medical service is yet to be delivered to patients and their families in need. KFRD has held educational workshops and seminars on Genetic Counseling in eight university hospitals during 10 months period from Sep. 2011 to June 2012 to in order to educate and inform pts. and their families of rare disease, what is "genetic counseling service", and with full understanding of accurate "medical and genetic" information on the dis. involved, how can genetic counseling help pts. and their families to cope appropriately with consequence of the disease. Survey questionnaire were administered to 1663 attendees to evaluate the awareness and need of genetic counseling service. Among 1000 respondents, 79% reported that "previously never heard about genetic counseling service", 83% found that participation in workshop &/or seminar was helpful to comprehend the meaning of "genetic counseling" for them, and 98% indicated that "genetic counseling service" would be helpful to overcome with the condition for pts. and their families of rare disease. Based upon the result of the study, it is necessary to provide further educational workshops and seminars to increase public awareness on genetic counseling. And it is clear that the need of genetic counseling service for pts. and their families of rare disease should be met as an integral part of medical services earliest possible for better management of dis. and quality of life. Furthermore recognition and understanding of the fact that the scope and role of genetic counseling is expanding in post genomic era of personalized medicine for delivery of quality health care, will lead to the efforts to overcome obstacles in providing genetic counseling service in korean health care system.

2510W

Modelling of downstream counselling impact of ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. *L. Burnett*[1,2,3], *L.C. Ding*[1], *D. Chesher*[1,2], *R. Lew*[2,3], *A. Proos*[1,2]. 1) PaLMS Pathology North, NSW Health, Royal North Shore Hospital, St Leonards, Sydney, NSW, 2065, Australia; 2) Sydney Medical School - Northern, Royal North Shore Hospital E25, University of Sydney, Sydney, NSW, 2065, Australia; 3) Department of Obstetrics and Gynaecology, QEII Research Institute for Mothers and Infants, The University of Sydney, NSW, 2006 Australia.

Purpose To predict the clinical impact of expanded testing with Massively Parallel DNA Sequencing using the ACMG Recommendations for Clinical Reporting of Incidental Findings.

Design Mathematical modelling of gene and variant frequencies.

Methods A diagnostic panel of clinically significant genetic conditions was simulated, based on the American College of Medical Genetics and Genomics (ACMG) Recommendations. The number of patients with significant variants in one or more conditions that would be detected using the screening panel was calculated from known (or best estimate) prevalence. Where a range of variant prevalence data was available, we selected the lowest and highest values, and calculated the most-likely value as the geometric mean, while where only a single datum was available, we selected half and twice this prevalence as the lower and higher limits. Assuming all tested disorders were inherited independently, the probability calculation used was $1- [(1-p1)*(1-p2)*...(1-p24)]$, with p1 to p24 as the carrier frequencies of the 24 proposed ACMG conditions. Calculations were repeated separately for the lower and higher limits. The proportion of individuals screened who would require supplementary consultation and genetic counselling was calculated.

Results The proposed ACMG recommended screening panel would require supplementary consultation and genetic counselling for 2.7% (range 1.6%-6.9%) of screened individuals.

Conclusion Reporting of incidental findings arising from massively parallel sequencing will require increased clinical follow-up and genetic counselling referrals. The number of such individuals will vary with the particular population being screened as well as the choice of genes included in the screening panel. This will have implications for health care workforce and infrastructure planning. New models of genetic education and counselling may be required to cope with these projected demands.

2511W

Factors that drive pediatric medical management following chromosome microarray analysis. *R. Hayeems*[1,2], *N. Hoang*[1], *S. Chenier*[4], *J. Stavropoulos*[1], *S. Pu*[1], *S. Wodak*[1], *R. Babul-Hirji*[1], *J. Davies*[3], *L. Velsher*[3], *J. Aw*[3], *R. Weksberg*[1], *C. Shuman*[1]. 1) Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada; 2) Institute of Health Policy Management and Evaluation, Univ Toronto, Toronto, ON, Canada; 3) Genetics, Medcan Clinic, Toronto, Ontario, Canada; 4) Département de pédiatrie, Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Quebec, Canada.

Background: Understanding management practices triggered by microarray analysis will inform clinical utility and resource planning. We explored the association between patient clinical characteristics and medical recommendations that follow pediatric microarray. **Method:** Of 1664 microarrays ordered in a pediatric tertiary care hospital in 2009/2010, 945 (56.8%) met inclusion criteria for this historical cohort study. Of these, 281/945 had a reportable microarray result (29.7%) and 664/945 had a benign result (70.3%). All cases with a reportable result and a random sample of benign cases were reviewed. The primary outcome sought was new medical recommendations following microarray result reporting. **Results:** Medical recommendations were provided for 328/451 cases (72.7%). Binomial log link regression analyses of relative risk (RR) indicate that pathogenic microarray results are associated with increased management recommendations compared to benign results [RR=1.36, 95% CI (1.17, 1.59)] but only a slight increase among those with uncertain results compared to benign results [RR=1.22, 95% CI (1.07,1.40)]. In addition, age <12 months is associated with increased management recommendations compared to children >5 years [RR=1.44, 95% CI (1.27, 1.64)]. The presence of multiple congenital anomalies (MCA) alone is associated with slightly more management recommendations than developmental delay (DD) alone [RR=1.37, 95% CI (1.02, 1.84)]. In a two-factor model assessing the effects of result type and phenotype, the effect of pathogenic results is retained [RR=1.34, 95% CI (1.15,1.55)] but the effect of MCA is marginal [RR=1.36, 95% CI (1.02,1.82)]. In a two-factor model assessing the effects of result type and age, the effects of result type and age are retained [RRpath=1.27, 95% CI (1.14,1.43); RRage=1.42, 95% CI (1.25,1.60)]. **Conclusions:** Of the variables considered, pathogenic results and age are the strongest drivers of new medical recommendations in a pediatric cohort. Uncertain microarray results only drive medical management marginally. These finding warrants further study with respect to establishing clinical utility.

2512W

The NINDS Repository Biomarkers Discovery Collection: A Public Resource of Biomaterials for Neurodegenerative Disease Research. *C. Tarn*[1], *M. Self*[1], *K. Gwinn*[2], *M. Sutherland*[2], *C. Pérez*[1], *W. Muhammad*[1], *G. Balaburski*[1], *M. Frasier*[3], *L. Vincent*[3], *R. Corriveau*[2]. 1) The NINDS Repository, Coriell Institute for Medical Research, Camden, NJ; 2) National Institute for Neurological Disorders and Stroke-NIH, Bethesda, MD; 3) Michael J Fox Foundation, New York, NY.

Neurological disorders present a massive challenge to healthcare systems globally. Identification of novel and reliable biomarkers that would allow for more efficient diagnosis, monitoring of disease onset and progression, and assessment of response to therapeutics, holds immense promise for improving clinical outcomes in individuals affected with disorders such as Parkinson's Disease and Huntington's Disease. The National Institute of Neurological Disorders and Stroke (NINDS) Repository at the Coriell Institute, also known as the NINDS Human Genetics Resource Center, has an overall mission of accelerating discovery of causes and risks for neurological disease by sharing biomaterials and de-identified clinical data. As a centralized facility for storage, processing, and distribution of biofluids (cerebrospinal fluid, plasma, serum, whole-blood, urine) and nucleic acid (DNA and RNA), the NINDS Repository serves as an integral component in the effort to identify and validate biomarkers of neurological disease. The establishment of large collections of biological samples obtained longitudinally from both affected and neurologically healthy individuals should prove invaluable for furthering investigation of biochemical markers via transcriptomic, proteomic, or metabolomic approaches. Currently the NINDS Repository collects samples under multiple NINDS sponsored biomarker initiatives including the Parkinson's Disease Biomarkers Program (PDBP) and the Neurobiological Predictors of Huntington's Disease (PREDICT-HD), as well as a jointly sponsored study (BioFIND) on Parkinson's Disease in collaboration with the Michael J. Fox Foundation. The NINDS Repository aims to: (i) ensure that samples collected for biomarker discovery are of premier quality by collaboratively establishing unified standards of sample collection; (ii) provide rapid feedback to clinical sites regarding sample appearance and quality; (iii) maintain secure, high quality sample storage conditions with real-time monitoring and recording systems; (iv) perform standardized laboratory processing and quality assurance using validated operating procedures. The NINDS Repository thus provides a vital resource for research designed to discover and validate biomarkers of neurological disorders. Biomarker discovery samples are available upon request either directly from the NINDS Repository web catalog (http://ccr.coriell.org/NINDS), or via NIH-sponsored resources with links to the online catalog.

2513W

Measuring treatment preferences of parents of children with Duchenne muscular dystrophy using best-worst scaling. *H.L. Peay[1], I. Hollin[2], H. Sheffer[1], J.F.P. Bridges[2].* 1) Parent Project Muscular Dystrophy, Richmond, VA; 2) Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Purpose: Duchenne muscular dystrophy (DMD) is a progressive disorder with no approved treatment. Individuals lose ambulation at ~8-12 yrs and die in the 3rd decade. There are no data on priorities for DMD treatments. We quantified the preferences of parents/guardians for potential DMD treatments' benefits and risks. Methods: Engagement with clinicians, industry, and parents yielded 6 treatment attributes (muscle function, lifespan, amount of post-approval data, nausea, risk of bleeds, arrhythmia risk) with 3 levels each. Attributes/levels were reasonable based on treatments under trial. Best-worst scaling (BWS) case 2 was used in a survey of U.S. parents of children with DMD. Participants were presented 18 benefit/risk combinations based on an orthogonal main-effects design. In each profile, respondents chose the best and worst aspect. Under assumptions of sequential best-worst and homogeneity across respondents, a choice model was estimated using McFadden's conditional logistic regression. Parameters were effects-coded (ensuring levels were anchored around zero) and relative attribute importance calculated based on overall utility differences across the attribute's levels. Results: Of 119 parents/guardians, the average age was 43.7 (SD 7.7) and predominately Caucasian (92%), married (90%), biological mothers (67%). Relatively, the most important attribute was preservation of muscle function~accounting for 27% of the variation across model attributes, and comprising the most and least preferred levels (stopping the progression of weakness and no effect on weakness, respectively). This was followed by: risks of arrhythmia and bleeds, each accounting for ~21% of the variation; changes in life span (17%); nausea (10%); and additional years of post-approval data (4%). Conclusion: This study presents a unique and robust approach to prioritizing treatment benefits/risks. Though presented with hypothetical treatments, participants exhibited preferences and non-random choices. Stopping or slowing weakness progression accounted for the largest proportion of variation in attributes. Muscle function was prioritized over lifespan and risks. Our data suggest side effects/risks could be compensated for by stopping weakness progression. However, the experiment revealed no cumulative benefit profile that offset the most severe risk profile, suggesting a limit to risk tolerance. Future research will focus on preferences heterogeneity and latent-classes.

2514W

A review of the Welsh breast cancer screening programme for women with a family history of breast cancer at a moderately increased risk. *S.L.M. Nisbet[1], L. Murray[2], K. Pegington[2], M.T. Rogers[1], K. Gower-Thomas[2,3], A.J. Murray[1].* 1) Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff, South Glamorgan, United Kingdom; 2) Breast Test Wales, 18 Cathedral Road, Cardiff, CF11 9LJ, UK; 3) Royal Glamorgan Hospital, Ynysmaerdy, Llantrisant, Pontyclun, CF72 8XR, UK.

In Wales, women with a family history of breast cancer are referred to the All Wales Medical Genetics Service (AWMGS) for risk assessment, genetic counselling and discussion of screening options. Those women who are assessed to be at an increased for breast cancer are offered regular mammographic screening through a centrally-coordinated, national screening programme, which is run by Breast Test Wales (BTW). BTW is the organisation in Wales that delivers the NHS Breast Screening Programme (NHSBSP), offering 3-yearly mammography to all women in the general population aged 50 to 70 years. BTW took over the provision of all family history screening in Wales in 2001, in order to provide a more equitable service and to facilitate better data collection and quality assurance. Unlike the NHSBSP which provides screening on mobile mammography units, the family history screening is delivered in the three regional BTW centres in South East Wales (Cardiff), Mid & West Wales (Swansea) and North Wales (Llandudno). Since AWMGS established a cancer genetics service in 1999 our understanding of familial breast cancer risk assessment has improved significantly, and several different risk assessment algorithms and software packages have been developed. Following the recent publication of the updated guidance on Familial Breast Cancer by the National Institute of Health and Clinical Excellence (NICE), we commenced a review of all women previously assessed as being at a moderately increased risk, who are currently enrolled in the BTW Family History Screening Programme to ensure women have been given an accurate risk assessment and are not being screened inappropriately Women were informed about the review when attending for their annual mammogram and advised that they would be contacted if we felt their risk or screening requirements had changed. Every woman's family history has been reviewed and their risk recalculated using the Boadicea web application developed by the Centre for Cancer Genetic Epidemiology at the University of Cambridge. This model has been well validated and is one of the risk assessment models endorsed by NICE. We will present the results of the pilot study conducted in the South East Wales region, which is now being extended across Wales. We will discuss the implications for the future assessment of women at an increased risk for breast cancer in Wales, which resulted from this review and some of the difficulties associated with a large scale review.

2515W

Documenting family history of cancer and referral for genetic counseling and testing in screening families at risk of hereditary breast or colorectal cancer. *I. Shapira, K. Cheng, V. John, K. Sultan, R. Sharma, D. Gokalp Yasar, N. Nyatanga, N. Eisler, J. Cho, E. Taioli.* Hofstra North Shore LIJ School of Medicine, Lake Success, NY.

Background Over 30% of colorectal cancer (CRC) is diagnosed at stage 3 due to low screening. Family history is the most important tool to identify hereditary cancer syndromes. Screening in families at hereditary risk reduces cancer morbidity and death. Prior studies demonstrate a low rate of family history documentation and low referral rates for genetic counseling and genetic testing. Aims 1) Assess the documentation of family history in different settings: hospital CRC versus office (breast cancer-BrCa-) in syndromes with similar penetrance such as Lynch and hereditary breast ovarian cancer syndrome (HBOC). 2) Define barriers to family history documentation. Methods: We evaluated the documentation of family histories in in charts of 630 consecutive colorectal cancer patients admitted for initial surgery in hospital and 295 Br Ca patients seen in office breast their initial oncology consultations between 2009-2011. Statistical analysis was performed for each group of patients using Normal method with standard error of proportion of 0.261 for 95% confidence interval (normal value 1.96) Of the 630 CRC patients, 237/630 (37%) had cancer family history recorded for the 1st degree relatives and 41/ 630 (7%) had cancer family history recorded for their 2nd degree relatives. The Amsterdam II criteria for referral to genetic cancer counseling and testing requires documentation of family history of 3 generations (first and second degree relatives and the proband): only 7%[95%CI 6.7%-7.7%] of the entire cohort of CRC patients, and only 6% of the colorectal cancer patients diagnosed under the age of 50 years old satisfied the Amsterdam II criteria. Of BrCa patients seen in office: 213/ 295 (71% [95%CI 47%-87%]) had 3-generations pedigrees documented in their family history had documentation of family history and 82/295, (29%) did not have three generation pedigrees documented; Conclusions: Appropriate referral for genetic counseling and genetic testing requires a complete and accurate documentation of family history. Significant differences were seen between the breast cancer charts and colorectal cancer charts, with greater accuracy of family history documentation and higher referral rates among breast cancer patients. To obtain improvement in the identification and management of patients at high risk and their family members, significant improvements in family history documentation are needed. Education is part of the answer.

2516W

Needs assessment of individuals with 22q11.2 Deletion Syndrome transitioning from pediatric to adult health care settings. *W.L.A. Fung*[1,2,3,4], *E. Leung*[5], *A.S. Bassett*[1,2,3]. 1) Department of Psychiatry, University of Toronto Faculty of Medicine, Toronto, Ontario, Canada; 2) The Dalglish Family Hearts and Minds Clinic for 22q11.2 Deletion Syndrome, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada; 3) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 4) Departments of Psychiatry and Clinical Genetics, North York General Hospital, Toronto, Ontario, Canada; 5) Office of Continuing Education and Professional Development, University of Toronto Faculty of Medicine, Toronto, Ontario, Canada.

Background: The multi-systemic clinical manifestations of 22q11.2 deletion syndrome (22q11.2DS) make the provision of comprehensive care for patients challenging. A coordinated, multidisciplinary team approach in the provision of care is recommended in Practice Guidelines. While several comprehensive 22q11.2DS centers have been established for children - and the world's first such center for adults recently established in Toronto, Canada - the provision of such coordinated care had been identified as a particular challenge for 22q11.2DS patients transitioning from pediatric to adult health care setting. We sought to assess the needs of 22q11.2DS patients transitioning from pediatric to adult health care setting. Methods: A multi-pronged approach was utilized in assessing the needs of these patients. This included paper surveys of 22q11.2DS patients and families, phone interviews of 22q11.2DS patients and families, as well as electronic surveys of professionals involved in the care of these patients and families. Questions of both quantitative and qualitative nature were used. Results: Considerable differences were identified between 22q11.2DS patients and families in terms of their perspectives on the patients' 1) knowledge of 22q11.2DS; 2) ability to transition to adult health care setting and adult life overall; 3) social and emotional well-being. In general, patients rated themselves as having more knowledge regarding 22q11.2DS and greater social and emotional well-being, compared to their families' ratings. Patients also expressed a higher level of confidence in their ability to transition to adulthood than their families' rating. The key challenges identified by these families included psychological, emotional and behavioral issues associated with 22q11.2DS, the availability of social supports to patients, and the negative effects of these symptoms on the patients' support systems. Findings from the professional stakeholder surveys will be presented at the annual meeting. Conclusion: To our knowledge, this is the first needs assessment conducted on 22q11.2DS patients transitioning from pediatric to adult health care setting utilizing a multi-pronged approach. These findings will help inform the development of an evidence-based transition program to enhance the care of these patients and their families.

2517W

The DSD-Translational Research Network, a national research and clinical network to improve health for people with Disorders of Sex Development. *E. Delot*[1], *D.E. Sandberg*[2], *E. Vilain*[1]. 1) Human Genetics, UCLA, David Geffen Sch Med, Los Angeles, CA; 2) Mott Children's Center, University of Michigan, Ann Arbor.

Disorders of Sex Development (DSD) are congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical. For families, the birth of a child with a DSD, with the uncertainty it brings regarding the child's gender and future psychosexual development, is believed to be extraordinarily stressful. Clinical care in DSD has been severely hampered by a fragmented research agenda, leaving fundamental gaps in knowledge of DSD pathology or treatment outcomes. Our long-term goal is to establish an environment in which clinical care of persons affected by DSD is evidence-based and guided by research that identifies factors impeding or enhancing opportunities for a positive quality of life across the lifespan. With the support of the NICHD, we created the DSD-TRN, to establish best practices by: 1) Building a sustainable infrastructure for translational research, including: - an interactive registry that will maintain standardized data from clinical sites, provide actionable information to clinical care teams and patients/families to support diagnosis and treatment management, and foster rapid translation of new evidence into clinical practice; - a collaborative network of researchers, clinicians, and patient/family advocates that will drive the research agenda and objectively monitor the impact of translational research at the bedside. 2) Standardizing radiological, biochemical, histological evaluations, descriptions of genital phenotype and post-surgical appearance and function. 3) Identifying biological and social factors associated with variability in psychosocial, psychosexual, and quality of life outcomes in patients with DSD. 4) Identifying novel pathophysiological mechanisms & improving the molecular diagnosis of DSD. The DSD-TRN network currently includes 5 academic centers (UCLA, U. Michigan, UCSF, Seattle Children's Hospital, Cincinnati Children's Hospital) with multidisciplinary teams (urology, genetics, endocrinology, psychosocial, etc.), multiple consultants (ethics, cancer, cost-effectiveness,...) and an Advisory Board with representatives of the major DSD patient advocacy groups. Deliverables already in place include a new exome-based genetic testing platform that allowed identification of novel variants in DSD patients, clinical use data collection forms integrated with electronic medical records, information brochures for patients and a registry of patient data to support future evidence-based clinical practice.

2518W

Whole genome sequencing vs. family history: physician perceptions of clinical utility. *J.L. Vassy[1,2,3], R.C. Green[1,6], J. Krier[5], D. Lautenbach[6], K.D. Christensen[6], M.A. Giovanni[7], M.F. Murray[7], A.L. McGuire[4] for The MedSeq Project.* 1) Department of Medicine, Harvard Medical School, Boston, MA, USA; 2) Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, USA; 3) Section of General Internal Medicine, VA Boston Healthcare System, Boston, MA, USA; 4) Baylor College of Medicine Center for Ethics and Health Policy, Houston, TX, USA; 5) Medical Genetics Training Program, Harvard Medical School, Boston, MA, USA; 6) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; 7) Geisinger Health System, Danville, PA, USA.

Background The uptake of whole genome sequencing (WGS) into patient care will depend in large part on whether physicians perceive it to have clinical utility. It is informative to compare the perceived utility of WGS to the current benchmark of genetic risk assessment in clinical care: family history (FmHx). Standard medical practice includes at least a cursory FmHx assessment, while for most physicians WGS remains an experimental technology. We hypothesized that physicians would find FmHx more clinically useful than WGS now but that they expect the utility of WGS to increase in the future. **Methods** We have enrolled primary care physicians and cardiologists to participate with their patients in the MedSeq Project: a randomized trial of WGS in clinical care. Before enrolling patients, we surveyed physicians about their perceived clinical utility for FmHx and WGS. We asked them to rate on a scale of 1-10 (`Not at all Useful' to Extremely Useful') how useful they thought the study's FmHx and WGS reports would be for `managing [their] patients' health' at two times: now and in the future. We categorized responses as not useful (1-5) or useful (6-10) for descriptive analyses. We used paired t-tests to compare the reported utility ratings of FmHx vs. WGS. **Results** Of 17 physicians (mean age 52 years), 9 (53%) were women, and 5 (29%) were of non-white race. Nine (53%) were primary care physicians, and 11 (65%) reported no genetics training beyond the usual medical school curriculum. The majority of physicians said FmHx would have clinical utility both now and in the future (n=15, 88%, for both responses). In contrast, only 4 (24%) said that WGS would have utility now. However, 15 (88%) predicted that WGS would have future utility for their patients' care. Physicians rated FmHx to have higher utility than WGS now (mean responses 7.6 vs. 4.2, $p<0.001$), but they rated FmHx and WGS to have similar utility for the future (mean responses 7.9 vs. 7.5, $p=0.49$). **Conclusions** Physician participants of the MedSeq Project rated FmHx to have greater utility than WGS for the present-day management of their patients' health, but they expected this utility gap to close in the future. Whether and how quickly this gap closes will likely depend on how the clinical genomics community addresses barriers to integrating WGS into patient care.

2519W

Conceptualizing family history taking in clinical pediatric practice. *L. Tessier[1], B.K. Potter[1], J. Brehaut[2], P. Chakraborty[3,4], J.C. Caroll[5], J. Allanson[4], J. Little[1,6], D. Castle[7], B.J. Wilson[1], CIHR Emerging Team in Genomics in Screening.* 1) Department of Epidemiology & Community Medicine, University of Ottawa, ON, Canada; 2) Clinical Epidemiology Program, Ottawa Hospital Research Institute, ON, Canada; 3) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 4) Department of Pediatrics, University of Ottawa, ON, Canada; 5) Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 6) Canada Research Chair in Human Genome Epidemiology; 7) ESRC Innogen Centre, University of Edinburgh, Edinburgh, UK.

Introduction: Family history (FH) is regarded as offering useful information in many clinical settings, for identifying potential genetic disease but also for risk assessment for many complex disorders. However, very little is known about how it is applied in pediatrics. Understanding how and why pediatricians collect FH data, and how they use it, would be helpful in evaluating its potential clinical utility. Recognizing that patterns of clinical practice reflect attitudes, knowledge, and many contextual factors, we wanted to develop a conceptual model of FH taking in pediatrics to guide further research. With a view to developing evidence-based tools for pediatricians, this model will be used to inform qualitative interviews with practitioners, and subsequently in developing a survey for pediatricians working in different health care settings. **Objectives**: To develop a conceptual framework of FH taking in pediatric practice as the basis for implementation research. **Methods**: We used as a starting point the Theoretical Domains Framework (TDF), which is being used in health services and implementation research. The TDF groups 33 psychological theories of behavior, with 14 identified domains (behavior change mediators) which can be applied to understanding specific professional behaviors. Using the TDF promotes a broad, comprehensive view of a topic, and its linkage to known psychological theories facilitates the informed development of interventions such as FH tools. We conducted literature reviews and consulted with experts in pediatrics, family medicine, cognitive psychology, and health services research to map each of the 14 TDF domains to specific aspects of FH taking as a `behavior'. **Results**: The following TDF domains appear relevant to exploring FH taking by pediatricians (terminology reflects original description of the framework): `knowledge'; `skills'; `social/professional role and identity'; `beliefs about capabilities'; `optimism'; `beliefs about consequences'; `reinforcement'; `intentions'; `goals'; `memory, attention and decision processes'; `environmental context and resources'; `social influences'; `emotions'; `behavioral regulation'. Each domain maps to ≥1 published theory. We will present detailed explanations of each domain, and clinical examples to illustrate how the framework can be applied to understanding how pediatricians take FH, and their attitudes towards this as a component of practice.

2520W

Genetic counsellors' preferences for Preimplantation Genetic Diagnosis: Designing a Discrete Choice Experiment. *E. Goh[1, 2], W. Ungar[2], D. Marshall[3], F. Miller[1]*. 1) Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; 2) Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, ON, Canada; 3) Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada.

Preimplantation genetic diagnosis (PGD) permits couples at high risk of a genetic condition to test an embryo for it prior to pregnancy. In the absence of explicit Canadian public policy regarding conditions of use, the decisions of front line clinicians (genetic counsellors, GC) guide clinical practice.

Discrete choice experiments (DCE) are a valid method for quantifying preferences and for measuring trade-offs between the characteristics of alternatives. Our qualitative research included 2 focus groups of practicing GC in Toronto, using a semi-structured focus group guide to identify factors that are important to GC in recommending public PGD coverage to inform our DCE. The first focus group explored what GC deemed relevant using open probes. These criteria were combined with a literature review as probes in the second focus group to generate a list of common coverage criteria. GC ranked the criteria individually, drawing on their perception of importance. Field notes and rankings were descriptively analyzed to identify insights surrounding coverage criteria. Pre-testing of the draft DCE survey was completed using a cognitive interview protocol with 7 GC.

The following attributes were identified as relevant for inclusion: i) PGD indication (childhood-onset condition, adult-onset condition, and adult-onset cancer predisposition), ii) risk of the genetic condition (50%, 25% and 1%), iii) fertility status of couple (infertile or fertile), iv) family history (have children at least one of who is affected, have children none of whom are affected, or have no children) and v) number of *in vitro* fertilization cycles to be funded (1 cycle, 3 cycles or 6 cycles). Based on pre-testing, PGD indication appeared to be the most important factor in deciding between the alternatives.

This qualitative study identified factors deemed important to GC in informing public PGD coverage. Most, but not all of the factors were highlighted in the literature. A DCE was designed based on these attributes. A pilot study of the quantitative phase is currently being undertaken (results expected July 2013, N=20). The results of the DCE can be used to quantify trade-offs amongst these factors and quantify national GC preferences for PGD, which can help inform public policy.

2521W

Identification and Characterization of Genetic Services in Peru. *M.R. Cornejo-Olivas[1,2], M.R. Velit-Salazar[1,3], T. Avalos-Cruz, MD[4], M.M. Duenas-Roque, MD[5], M. Inca-Martinez[1,6], A.P. Mora-Alferez, MD[7,8], S.M. Siccha-Arancibia[3], M. Brunner-Sciarra, MS, PhD[3], P. Mazzetti, MD, MBA[1,6], M.I. De Michelena, MD, PhD[3,8], C. Matos-Miranda, MD, MsC[3,9]*. 1) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurologicas, Lima, Peru; 2) Northern Pacific Global Health Research Fellows Training Consortium, Bethesda, US; 3) Universidad Peruana Cayetano Heredia, Lima, Peru; 4) Ministerio Publico, Lima, Peru; 5) Hospital Nacional Edgardo Rebagliati Martins, Lima, Peru; 6) Universidad Nacional Mayor de San Marcos, Lima, Peru; 7) Instituto Nacional de Enfermedades Neoplasicas, Lima, Peru; 8) Instituto de Medicina Genetica, Lima, Peru; 9) Montreal Children's Hospital Research Institute - McGill University. Montreal, Canada.

Statement of purpose. By 2004, genetic consultations, molecular diagnosis and research in human genetics in Peru were performed in a few centers in Lima, the country's capital city. In the last decade, genetic services have significantly grown but are still unregulated. Our study aims to identify and characterize the institutions offering human genetic services in Peru. **Methods**. A cross-sectional study was carried out using web search and interviews that were conducted with geneticists and other field-related professionals from institutions offering clinical, molecular and/or research services in human genetics in Peru. The survey was realized using a standard collecting data form. The study obtained IRB approval. **Results**. From the sixty-two identified institutions, forty of them completed the survey. The majority (92.5%) are located in Lima. Public health institutions represent 25%; five percent are public forensic institutions; 10% belong to both private and public universities, from which only two offer postgraduate training in human genetics; 5% are non-profit organizations; 55% are private genetic centers dedicated to specific laboratory activities including karyotyping (52.5%), FISH (32.5%), paternity testing (17.5%) and other molecular diagnostic techniques, such as PCR (50%), qPCR (27.5%), gene sequencing (27.5%) and others (25%). More than half of the participating institutions (55%) claim to perform basic and clinical research; only ten institutions have published a total of thirty-eight MEDLINE indexed papers. Twenty-one institutions offer genetic counselling and four some kind of genetic treatment. **Conclusions**. Public genetic institutions in Peru generally combine clinical, laboratory genetic services and research; while private centers offer specific procedures. Molecular diagnosis for genetic diseases is still very limited. Karyotyping, FISH, PCR and qPCR are the most available techniques in the country. Despite scarce training opportunities, human genetic services are growing in Peru but are still mainly located in Lima. We envision this information to be helpful in generating human genetic networks and defining public health policies in Peru.

2522W

Diagnostic application of targeted resequencing for familial nonsyndromic hearing loss. *B. Choi[1], G. Park[2], J. Kim[2], A.R Kim[2], B.J Kim[2], T. Park[2], S. Oh[2], K. Han[2], W. Park[3, 4].* 1) Seoul National University, Bundang, Gyeong Giido, South Korea; 2) Seoul National University, Seoul, South Korea; 3) Sungkyunkwan University, Suwon, South Korea; 4) Samsung Genome Institute, Samsung Medical Center, Seoul, South Korea.

Identification of causative genes for hereditary nonsyndromic hearing loss (NSHL) is important to decide treatment modalities and to counsel the patients. Due to the genetic heterogeneity in sensorineural genetic disorders, the high-throughput method can be adapted for the efficient diagnosis. To this end, we designed a new diagnostic pipeline to screen all the reported candidate genes for NSHL. For validation of the diagnostic pipeline, we focused upon familial NSHL cases that are most likely to be genetic, rather than to be infectious or environmental. Among the 32 familial NSHL cases, we were able to make a molecular genetic diagnosis from 12 probands (37.5%) in the first stage by their clinical features, characteristic inheritance pattern and further candidate gene sequencing of GJB2, SLC26A4, POU3F4 or mitochondrial DNA. Next we applied targeted resequencing on 80 NSHL genes in the remaining 20 probands. Each proband carried 4.8 variants that were not synonymous and had the occurring frequency of less than three among the 20 probands. These variants were then filtered out with the inheritance pattern of the family, allele frequency in normal hearing 80 control subjects, clinical features. Finally NSHL-causing candidate mutations were identified in 13(65%) of the 20 probands of multiplex families, bringing the total solve rate (or detection rate) in our familial cases to be 78.1% (25/32) Damaging mutations discovered by the targeted resequencing were distributed in nine genes such as WFS1, COCH, EYA4, MYO6, GJB3, COL11A2, OTOF, STRC and MYO3A, most of which were private. Despite the advent of whole genome and whole exome sequencing, we propose targeted resequencing and filtering strategy as a screening and diagnostic tool at least for familial NSHL to find mutations based upon its efficacy and cost-effectiveness.

2523T

Automating Clinical Exome Analysis. *M.N. Bainbridge[1,2,3], E.B. Venner[3], C. Eng[2], Y. Yang[2], R.A. Gibbs[1,2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Whole Genome Sequencing Lab, Baylor College of Medicine, Houston, TX; 3) Codified Genomics, LLC, Houston, TX.

The clinical use of genome wide sequencing is now commonly available and is poised to be ubiquitous in the next year, however, several challenges remain in the timely and cost effect analysis of these data. Initial review and prioritization of the hundreds of potentially disease causing variants is critical for making accurate diagnoses. Unfortunately, this first step is expensive, laborious, time consuming and prone to error. We have developed an information pipeline and set of algorithms to fully automate this initial review process. Initially, we capture clinician expertise regarding the patient phenotype and use this information to prioritize genes based on phenotypic overlap with diseases. This enable those with less medical training to identify variants which likely contribute to the patients phenotype and promote these variants for further review. Additionally, using this data as well as a rich set of annotation information about each variant and gene, we can automate the categorization of the variants according to ACMG guidelines. We tested our algorithms on a series of clinical exome cases that were previously solved by standard methods and found that the causative variant was almost always within the top 15 of all rare, protein changing variants. Further, we were able to identify other, strong candidate variants that were not highly prioritized using manual methods including in cases where multiple genes were suspected of contributing to the phenotype. Further, these improvements reduce the time spent on the initial variant review to approximately half. These analysis methods are critical in a clinical setting. They aid in reducing costs, time, and errors in the initial review greatly improving the quality of the test.

2524F

Design and validation of Next Generation Sequencing reference standards for oncology. *J. Goodall, C. Lowe, C. Thorne, A. Mulligan, J. Frampton, B. Burke, K. Schmitt, P. Morrill.* Horizon Diagnostics, Cambridge, Cambridgeshire, United Kingdom.

Horizon Discovery has established a range of best in class, genetically defined, genomic reference standards, including FFPE blocks and purified gDNA. These standards offer a sustainable and highly defined source of reference material to laboratories, proficiency schemes, and manufacturers. Here we present the design and validation of our Next Generation Sequencing (NGS) quantitative multiplex reference standards. NGS offers significant advantages for mutation detection, enabling the simultaneous detection of multiple mutations in multiple genes, and provides a digital readout of the mutation frequency. Challenges remain however, not least the number of different platforms, each presenting different systematic bias, and the need for extensive validation of the analytical pipeline to ensure variant calling is correctly and consistently achieved. A method for standardisation and verification is therefore required that goes beyond the current widespread use of HapMap controls. In response to the need for a better NGS reference standard, Horizon has leveraged its proprietary genome editing technology to create a multiplex reference standard covering many commonly assayed cancer mutations. The power of this approach is that virtually every characteristic of our reference standard can be defined, from the molecular constitution of the genome, to the DNA output associated with each product batch. Each sample contains key oncogenic mutations including KRAS G13D and G12D, PIK3CA H1047R, BRAF V600E, EGFR T790M, NRAS Q61K and KIT D816V represented at defined allelic frequencies ranging from 1 to 25%. Furthermore, >20 defined mutations in other disease relevant genes including PTEN, CTNNB1, ALK, FGFR2, MET, IDH1, NOTCH and BRCA1 are present. Digital PCR is used to quantify each mutation, enabling the precise quantification of mutational frequency. We show that the data obtained by NGS on our multiplex standard closely matches our digital PCR data, thereby validating the use of the standard as a tool that can be used as routine reference material by any lab intending to use NGS for clinical diagnostics. The multiplex reference standard described here establishes the integrity of the NGS workflow from enrichment and sequencing through to bioinformatics and data interpretation.

2525W

Simultaneous detection of point mutations and exonic deletions by target gene capture and deep sequencing. *Y. Feng, G-L. Wang, H. Cui, J. Wang, V.W. Zhang, L-J. Wong.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Background: Next generation sequencing (NGS) has demonstrated its clinical utility in the identification of point mutations but not the detection of large intragenic deletions, which is extremely important in elucidating its presence as a compound heterozygous mutation to a point mutation or accounting for an apparently homozygous mutation in autosomal recessive cases. **Method:** An analytical algorithm to detect exonic deletions was developed by comparing the normalized coverage depth (CD) of each coding exon of the testing sample to CD of that particular exon from a group of controls. **Results:** This algorithm allows detection and confirmation of homozygous and heterozygous deletions of various captured genes. This includes homozygous deletions of *MPV17* exons 3-7 deletion of a patient with mtDNA depletion syndrome, *LPIN1* exon 18 of a patient with rhybdomyolysis, and a hemizygous *PHKA2* exons 27-30 deletion of a male patient with GSD. In addition, there are compound heterozygous exonic deletions with point mutations, including *CPS1* E9-11, *citrin* E3, *TK2* E1-2, partial *OTC* exon 2 deletions, and a *CACT* E5-9 deletion originally identified as an apparently homozygous point mutation. **Conclusion:** It is extremely valuable to be able to detect point mutations and large exonic deletions simultaneously using target gene capture/NGS approach.

2526T

Next Generation Sequencing Coupled with a Novel Multiplex PCR Protocol for Comprehensive Genetic Screening of Maturity Onset Diabetes of the Young in India. *A. Chapla[1], D.M. Mahesh[1], D. Varghese[1], S.V. Nadig[1], H.S. Asha[1], R.T. Varghese[1], M. Inbakumari[1], F. Christina[1], S. Mathai[2], T.V. Paul[1], N. Thomas[1].* 1) Department of Endocrinology, Diabetes and Metabolism, Christian Medical College, Vellore, Tamil Nadu, India; 2) Department of Pediatric Endocrinology, Christian Medical College, Vellore, Tamil Nadu, India.

Maturity Onset Diabetes of the Young (MODY) is a monogenic disorder with an autosomal dominant pattern of inheritance characterized by β-Cell dysfunction. It accounts for around 1-2% of patients with diabetes and typically presents before the age of 25 years. Till date mutations in at least 13 different genes have been reported to cause MODY. Due to an overlap of clinical features with polygenic Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D), identification of patients with MODY is a diagnostic challenge. MODY genetic screening is of immense clinical importance and a confirmed genetic diagnosis would help streamline therapy. However, due to limitations in the scalability of the current diagnostic platform, performing genetic screening of a comprehensive panel of the identified MODY genes has been hindered in the past. As a result, majority of MODY patients are stereotypically classified as T1D or T2D and may potentially receive inappropriate therapy. This study aimed to establish genetic screening of a comprehensive panel of 10 MODY genes consisting of HNF1A, HNF4A, GCK, HNF1β, IPF1, NEUROD1, KLF11, CEL, PAX4, and INS. A novel multiplex PCR was established to enrich the target genes and further sequencing was performed on Ion Torrent Personal Genome Machine. Using this approach 95% of the targeted DNA was covered at a depth of 50 or more reads per nucleotide and at a minimum average base coverage depth of 300X. MODY genetic testing was carried out in 50 subjects with young onset diabetes of which 35 met the clinical criteria of MODY. We Identified mutations in seven patients, which include four with NEUROD1 mutation (novel c.175 G>C p.E59Q, novel c.-162G>A 5'UTR, and c.723C>G p.H241Q), one with HNF4A mutation (c.505G>A p.V169I coseggregating with c.493-4G>A and c.493-20C>T), one with GCK mutation (c.1318G>T p.E440X) and one with HNF1B mutation (novel c.274 C>T p.L92F). These mutations and the other identified rare variants were confirmed by Sanger sequencing. Further, to validate this protocol we compared the Ion torrent sequencing data of HNF1A with conventional Sanger sequencing in 10 subjects. Novel Multiplex PCR coupled with Ion Torrent Next Generation sequencing has allowed us for the first time to perform comprehensive genetic screening in Asian Indians in a rapid and cost-effective way. With this parallelized sequencing approach we have identified a higher frequency of NeuroD1 mutations, a pattern of MODY different from the Western population.

2527F

Many types of DNA damage can be detected with Two-Dimensional Strandness-Dependent Electrophoresis (2D-SDE). *J.J. Jonsson[1, 2], B. Gudmundsson[1, 2], H.G. Thormar[2, 3], A.G. Sigurdsson[2], S. Thongthip[4], M. Steinarsdottir[1], A. Smogorzewska[4].* 1) Dept. of Genetics and Molecular Medicine, Landspitali, Reykjavik, Iceland; 2) Dept. of Biochemistry and Molecular Biology, University of Iceland, Reykjavik, Iceland; 3) Lifeind/BioCule Inc., Reykjavik, Iceland; 4) Laboratory of Genome Maintenance, Rockefeller University, New York, NY.

Two-Dimensional Strandness-Dependent Electrophoresis (2D-SDE) in manual minigels or premade microgels is a novel technique for nucleic acid analysis. In the first dimension nucleic acid fragments are separated based on length and strandness i.e. double-stranded DNA, single-stranded DNA and RNA•DNA hybrids. The nucleic acids are heat denatured before the second dimension electrophoresis and in the second dimension all fragments are single-stranded and separate only based on length. We tested if 2D-SDE could detect various types of DNA damage in vitro and in vivo. Each sample was run in duplicate both uncut and cut with Mbo I, an enzyme which cuts both single- and double-stranded DNA. Single-stranded breaks, either nicks or gaps, were detected as horizontal streaks on 2D-SDE extending from uncut DNA molecules too large to efficiently enter the gel. Double-stranded breaks generated an arc in the gel. DNA molecules with interstrand crosslinks (ICL) migrated as an arc behind normal dsDNA molecules. In contrast, DNA with intrastrand crosslinks and bulky adducts were bent and migrated in front of that arc. Single-stranded DNA molecules, too damaged for complementary strand binding, formed a diagonal line. 2D-SDE detected DNA damage at comparable level of sensitivity to the well known comet assay. However, 2D-SDE allowed direct detection of damaged molecules and subfractions could be isolated form the gel. 2D-SDE could also be used to analyze cfDNA. 2D-SDE can detect many common types of DNA damage and the effects of DNA repair. Applications include testing quality of biosamples and efficiency of various molecular procedures were damage to DNA is common. Examples include extensive damage detected in CHIP-Seq experiments and FFPE samples. Applications also include genotoxicity testing, chemosensitivity testing and diagnosis of genome instablity and DNA repair disorders.

2528W

Next generation molecular diagnosis of patients with retinal degeneration. *L. Lan, N. Li, J. Chiang.* Casey Molecular Diagnostics Laboratory, Portland, Oregon.

Molecular diagnosis of inherited retinal degeneration has been difficult due to the following reasons: (1) locus heterogeneity (more than 200 genes involved); (2) allele heterogeneity (many private mutations); (3) overlapping clinical presentations; and (4) progressive nature of some conditions. The arrival of mass parallel sequencing: Next Generation Sequencing (NGS) provides an opportunity to revolutionize the task. At the Casey Molecular Diagnostic Laboratory, we have developed a method combining the specificity and low cost of PCR enrichment with Illumina MiSeq NGS platform. Several hundred clinical samples have been sequenced using this combination of methods thus far. Based on our experience, false positive and false negative callings are concerns. The presence of low coverage regions (gaps) is another concern. Mutations can be missed below cutoff (false negative), while sequencing artifacts above cutoff generate false positive results. Homozygous deletions can be missed if gaps are not filled. Therefore, after data analysis, gaps are also filled in by PCR and Sanger sequencing, and mutations and novel variations are confirmed by Sanger sequencing regardless of the quality score. The biggest drawback of this individual disease panel approach is for patients with uncertain clinical diagnoses, as multiple panel testing may be required in order to find mutations. At this time, we are developing a PCR based approach of testing the entire non-syndromic retinal degeneration genes plus some common syndromic genes. With the buildup of our database by testing more genes and confirming mutations and novel variations by Sanger sequencing, we aim to improve molecular diagnoses of patients with retinal degeneration. Our unique approach and experience will be presented.

2529T

Using Targeted Next Generation Sequencing for Diagnosis and Screening in Newborns. *T.D. Sokolsky, E.W. Naylor, A. Bhattacharjee.* Parabase Genomics, Inc., 100 Morrissey Blvd, Boston, MA 02125.

Newborns are disproportionately affected by rare monogenic diseases including hearing loss, inborn errors of metabolism, and lysosomal storage disorders. A diagnosis is critical in order to begin treatment early and prevent mortality or lifelong debilitation. Unfortunately, the current tests for newborns are limited in scope and scalability. Diagnostic confirmation often includes time-consuming and expensive serial single gene testing, which can result in reducing the available treatment window and increasing the risk of long term effects. We believe that next generation sequencing technology has the power to help infants in a way that conventional diagnostic approaches cannot. Targeted next-generation sequencing (TNGS) has allowed expansion of genetic testing across a large number of diseases. Here, we report on a new multi-gene panel that targets hundreds of genetic diseases predominantly affecting newborns and which neonatologists have recommend for inclusion in population based screening. Our panel encompasses diseases and symptoms such as hearing loss, hypotonia, hepatomegaly and failure to thrive. Through our TNGS pipeline, sequencing capacity can be maximized for high coverage of targeted regions along with decreased turnaround time and increased throughput. This optimized pipeline with lower per sample cost and better accuracy in variant calling can translate in the clinic to an accelerated patient diagnosis. Our TGNS approach focuses on the exonic regions of approximately 300 genes and uniquely also includes the entirety of specific genes, such as those for hearing loss and cystic fibrosis, where variants of clinical relevance are often found in non-coding regions. This design expands detection beyond exclusively exonic panels while maintaining the advantages of a smaller targeted panel. In our lab we additionally demonstrate improvements to extraction methods for small sized samples from newborns (e.g. DBS, saliva) in order to yield DNA in quantities sufficient for target enrichment by hybrid-capture and sequencing. Thus mitigating the need for whole genome amplification and associated biases. Our pipeline provides a sample to answer solution by incorporating our advances in sample prep, reduced sequencing run times with the latest instruments, and implementing commercial grade bioinformatic analysis with automated variant calling tools and interfaces.

2530F

Targeted Sequencing of Genes Causing atypical Hemolytic Uremic Syndrome and Coagulation Disorders. *S. Theru Arumugam[1], K. Meganathan[2], D. Kissell[1], S. Jodele[3], R. Gruppo[2], K. Zhang[1].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Division of Hematology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Division of Bone Marrow Transplantation and Immune Deficiency, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Sanger sequencing of multigenic disorders can be technically challenging, time consuming and very expensive. Recent advancements in Next-generation sequencing (NGS) not only facilitated the discovery of new disease genes, but also transformed the routine clinical diagnosis of genetic diseases. In this study, we evaluated the performance of NGS in detecting the DNA sequence variations in 52 genes causing atypical hemolytic uremic syndrome and coagulation disorders. Targets covering the exons, flanking intronic and regulatory regions of these genes were enriched using RainDance microdroplet PCR and sequenced, as single end 50 bp reads, on the Illumina HiSeq 2500 instrument. The raw sequence reads from each sample were subjected to molecular genetics laboratory's (MGL) NGS data analyses procedure, which include: (1) quality filtering of raw reads (2) mapping high quality reads to reference sequence and simultaneously detecting sequence variants using NEXTGene software (3) identifying potential true sequence variants using MGL criteria and (4) confirmation of variants with Sanger sequencing. A total of eight samples with known mutation were included in the pilot study and an average of about 20 million raw sequence reads were generated from each sample. About 97% of reads in each sample met the quality criteria and of which, approximately 67.5% aligned to the reference sequences of these 52 genes. We observed that about 98.6% of 145 kb target regions, covering exon and 20 bp flanking intronic regions, had at least 40 sequence reads. The data is currently being analyzed and results of this study will be presented.

2531W

Application of targeted next-generation sequencing in clinical diagnostics. *B. Sikkema-Raddatz, L.F. Johansson, E.N. de Boer, K. van Dijk-Bos, P. van Norel, J. Dijkhuis, M. Viels, M. Meems, A. Schipper, Y. Vos, H. Westers, B. Leegte, J.G. ter Beest, L. van der Heijden, A.H. van der Hout, L.G. Boven, J.P. van Tintelen, R. Almomani, R.H. Sijmons, J.D.H. Jongbloed, R.J. Sinke.* Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, Netherlands.

Purpose: The challenge in genetic diagnostics is to apply one comprehensive test for heterogeneous diseases. The number of genes to examine in a particular clinical case can be relatively large because phenotypes of many hereditary syndromes are known to overlap and different genes may underlie a single syndrome. Our aim was to design and implement various targeted next generation sequencing (NGS) gene-panels, starting with those disorders that account for the majority of current diagnostic requests. Methods: As proof of principle we developed a gene-panel based on Agilent Sure Select Target Enrichment® for simultaneous mutation detection for 48 genes associated with hereditary cardiomyopathies. To assess test-sensitivity and specificity we performed a validation on 84 patients. For 24 of these Sanger Sequencing (SS) data for up to six genes were available. Pools of 12 samples were sequenced using 151 bp paired-end reads on an Illumina MiSeq® sequencer and analyzed using NextGene® and Cartagenia-NGS® software. Subsequently, an extended gene-panel targeting 55 cardiomyopathy-associated genes was implemented in routine diagnostics. To date, more than 200 patients have been analyzed. In parallel, a targeted NGS gene-panel for 70 cancer predisposition genes was designed and validated in a similar manner. Results: For the cardiomyopathy gene-panel, 99 percent of all bases had a coverage of ≥30 reads per nucleotide. Because of poor coverage a total of 11 regions were analyzed in parallel using SS. We identified ~21000 variants (~250 per patient). SS was performed for 168 variants (155 substitutions, 13 indels). Al were confirmed, including a deletion of 18 bps and an insertion of 6 bps. Compared to previous routine diagnostics based on SS, application of our NGS-method resulted in increase of diagnostic yield from 15 percent to about 50 percent. Results for validation of the cancer gene-panel were comparable; 99 percent of all bases with ≥30 reads per nucleotide, for 22 regions SS is performed in parallel, no false positive or negative results upon confirmation with SS of 180 variants. The first patients are now being tested in diagnostics. Conclusion: Targeted NGS of a disease-specific subset of genes can be reliably implemented in diagnostics to analyze large numbers of genes in parallel with significantly improved diagnostic yield. Additional gene-panels i.e. for epilepsy and neurodegenerative disorders are currently being designed and implemented.

2532T

Improved Accuracy and Precision in Clinical Next Generation Sequencing with the SmartChip TE™ Target Enrichment System. *J. Dunne, W. Dong, G. Hein, S. Silveria, S. Derveaux, A. Chang, S. Anandakrishnan, M. Leong, M. Sanchez, D. Batey, S. Husain.* WaferGen Biosystems, Fremont, CA.

The main challenge for clinical targeted next-generation sequencing methods is obtaining complete and uniform coverage of all target regions. Some popular methods for target enrichment rely on lengthy and inefficient hybrid capture or multiplexed PCR techniques, resulting in lower coverage and more off target reads. To address these challenges, WaferGen has developed the high-density SmartChip TE System. One SmartChip TE panel supports conducting hundreds to thousands of parallel, singleplex PCR reactions to efficiently enrich desired target regions in less than 3 hours. Results of a recent study used a SmartChip TE Panel (140 kb target region) targeting exonic regions in a 17 gene set in 16 cancer cell lines (NCI 60). High design rates and percent bases on target ensured specific amplification necessary for efficient enrichment. The sequencing results from SmartChip TE enriched targets showed coverage at 20x and 100x of 98.8%; and 98.2%; respectively with a uniformity of coverage of 98% at > 10% of mean coverage. Comparison of performance between AmpliSeq and HaloPlex designs and the SmartChip TE custom designs will be presented. The unprecedented enrichment quality is achieved with flexibility of running up to 2500 unique singleplex reactions on a single SmartChip TE chip. The results indicate that CLIA-certified or clinical research laboratories can utilize the SmartChip TE system to obtain the most complete and uniform coverage among target enrichment technologies.

2533F

Improved genetic testing for monogenic diabetes using targeted next generation sequencing. *H. Lango Allen, R. Caswell, E. De Franco, S. Flanagan, G. Hysenaj, K. Colclough, J. Houghton, M. Shepherd, A.T. Hattersley, M.N. Weedon, S. Ellard.* University of Exeter Medical School, Exeter, United Kingdom.

Current genetic tests for diagnosing monogenic diabetes rely on selection of the appropriate gene for analysis according to the patient's phenotype. Next generation sequencing enables the simultaneous analysis of multiple genes in a single test. Our aim was to develop a targeted next generation sequencing assay to detect mutations in all known MODY and neonatal diabetes genes.

We selected 29 genes in which mutations have been reported to cause neonatal diabetes, maturity-onset diabetes of the young (MODY), maternally inherited diabetes and deafness (MIDD) or partial lipodystrophy (FPLD). We designed an exon-capture assay to include the coding regions and conserved splice sites. A total of 114 patient DNA samples were tested: 32 with known mutations and 82 previously tested for MODY (n=33) or neonatal diabetes (n=49) but in whom a mutation had not been found. Sequence data were analysed for the presence of base substitutions, small indels and exonic deletions or duplications.

All known mutations and polymorphisms (n=70 different variants) were detected including 55 base substitutions, 10 small insertions or deletions and 5 partial/whole gene deletions/duplications. Previously unidentified mutations were found in 5 patients with MODY (15%) and 9 with neonatal diabetes (18%). Most of these patients (12/14) had mutations in genes that had not previously been tested. These included mitochondrial m.3243A>G mutation; previously missed intronic *HNF4A* (c.358+5G>A) mutation; mutations in *EIF2AK3* and *SLC19A2* not originally tested because patients were referred before the onset of syndromic features; *ABCC8* mutation that was missed because of allelic dropout caused by a polymorphism within a primer binding site; and novel mutations in *GCK*, *PDX1* and *GATA6*.

In conclusion, our novel targeted next generation sequencing assay provides a sensitive method for simultaneous testing of all monogenic diabetes genes. The increased number of genes tested led to an improved mutation detection rate.

2534W

NGS Data Analysis for a Primary Immunodeficiency Gene Panel using Haloplex Enrichment Method.
J. Durtschi[1], E.M. Coonrod[1], R.L. Margraf[1], H.R. Hill[1,2,3,4], K.V. Voelkerding[1,2], A. Kumáovics[1,2]. 1) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 2) Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT; 3) Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT; 4) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.

Haloplex (Agilent) gene target enrichment offers a cost effective and efficient work flow for library preparation followed by Illumina MiSeq sequencing and was used to develop a panel targeting 94 primary immunodeficiency associated genes. Haloplex capture uses restriction enzymes to digest genomic DNA and then probes selectively capture and amplify DNA in 100 to 500 base DNA segments tiled across target regions. Haloplex sequence data is characterized by non-randomly distributed sequence inserts defined by the limited set of restriction enzymes used and unwanted Illumina adapter sequence at read 3' ends when sequencing goes beyond the end of shorter Haloplex insert sizes. In this study we evaluate Haloplex data processing and analysis issues encountered in our 94 gene panel. A custom Haloplex panel was designed for 94 genes associated with primary immune deficiency disorders and used to prepare NGS libraries for 40 samples. Four samples were indexed and multiplexed per Illumina MiSeq, 2x150, paired end run. Read data were trimmed of MiSeq adaptor sequence present at some 3' read ends, aligned, and analyzed using Cutadapt, BWA, and GATK softwares, respectively. During Haloplex data processing, Illumina adapter sequence was identified and trimmed from an average 40% of reads amounting to 13% of bases trimmed per data set. After this trimming, over 10% of reads were shorter than 100bp suggesting that these reads came from unintentional Haloplex inserts below the 100bp minimum. Inferred insert size analysis of alignments also indicated insert sizes less than 100bp for over 10% of read pairs. Median coverage in coding regions of panel genes was much higher in Haloplex samples versus three typical Nimblegen v3 exomes (on average, 1062 versus 123) but the fraction of the same coding regions with coverage below 30 was higher in Haloplex versus exomes (11% versus 4%) indicating less uniform coverage of Haloplex data. Of the collection of 702 different coding region variants that passed basic variant filters in at least one of 40 samples, 23% were not seen in the 5,400 exome data set, indicative of a high false positive rate. However, many likely false positives appear at systematic, trackable locations. Our ongoing study indicates that, when properly managed, these Haloplex data characteristics lead to an effective enrichment method for our multi-gene panel assay for the clinical testing of primary immunodeficiencies.

2535T

Detection of copy number variants in whole exome sequencing data in routine genome diagnostics.
N. de Leeuw, J.Y. Hehir-Kwa, D. Lugtenberg, M. del Rosario, J. de Ligt, R. Pfundt. Dept Human Genetics, Radboud University Medical Centre, Nijmegen, Netherlands.

Genome wide high resolution SNP array analysis has been used in our laboratory for the detection of copy number variations (CNVs) as a first tier diagnostic tool since 2009. In the past two years, we have increasingly used whole exome sequencing (WES) for the detection of clinically relevant single nucleotide variations (SNVs) and small insertion-deletions in our diagnostic setting for a variety of genetic diseases. In addition to this SNV detection, we recently demonstrated the clinical utility of WES to detect CNVs in a representative dataset from patients with intellectual disability (ID) in whom diagnostic SNP array analysis previously detected de novo, pathogenic CNVs. Based on these results, we implemented CNV detection in WES data parallel to the existing SNV detection, thereby increasing the range of genetic variation and the diagnostic yield that can be obtained from WES data. We started with the retrospective analysis of a total of 205 patients with ID and their parents who had previously been sent in for WES diagnostic trio analysis. Parental data were used to determine the inheritance or de novo occurrence of the CNVs, and also served as an estimation for the likelihood of detecting an incidental finding. We will compare the results of CNV detection in exon-targeted WES data for a total of 615 samples with high resolution genome wide SNP array (Affymetrix CytoScan HD array) data from 3,500 patients and parents. Our discussion will focus on the differences in resolution, the detection rate of clinically relevant CNVs and their copy number state, as well as on the frequency of incidental findings. We show that WES can be successfully used to efficiently reach a diagnosis in a patient by analysing both SNVs as well as CNVs, making WES a suitable approach as a first tier diagnostic test for patients with ID and/or congenital anomalies.

2536F

Clinical interpretation and reporting of secondary findings from genome sequencing: lessons learned from the first 15 cases of The MedSeq Project.
H.M. McLaughlin[1,2], J. Krier[3,4], W.J. Lane[1,5], D. Metterville[2], I. Leshchiner[2,4,6], B.H. Funke[1,2,7], J.L. Vassy[4,6,8], M. Murray[9], I.S. Kohane[3,4,6,10], S.W. Kong[3,11], C. MacRae[4,6], M.S. Lebo[1,2,5], R.C. Green[4,6], H.L. Rehm[1,2,5] for The MedSeq Project. 1) Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA; 2) Laboratory for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, Massachusetts, USA; 3) Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA; 4) Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; 5) Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; 6) Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; 7) Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA; 8) VA Boston Healthcare System, Boston, Massachusetts, MA; 9) Geisinger Health System, Danville, Pennsylvania, USA; 10) Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts, USA; 11) Department of Medicine, Boston Children's Hospital, Boston, Massachusetts, USA.

The MedSeq Project is a randomized clinical trial that aims to develop tested approaches for the evaluation and reporting of genome sequencing data and seeks to assess the impact of integrating genome sequencing into clinical care. MedSeq participants in the sequencing arm of the study receive a summary of secondary findings returned in a General Genome Report (GGR). The GGR features a concise summary of genome-wide secondary findings with clearly delineated sections for (1) highly-penetrant monogenic disease risk; (2) carrier status for phenotypes associated with recessive inheritance; (3) pharmacogenomic associations; and (4) blood groups. Here, we describe bioinformatics, interpretation, and reporting approaches for The MedSeq Project. Population frequencies from 1000 Genomes and the Exome Variant Server, loss-of-function predictions, and variant databases such as the Human Gene Mutation Database are used to filter potential disease-causing variants from the 3-5 million variants produced from each genome sequence, resulting in 100-130 variants requiring manual assessment per case. Comprehensive evidence-based variant and gene assessments are performed for each variant to identify those with evidence for causing highly penetrant disease, typically 1-5 per case, which require Sanger confirmation. Finally, a team-based approach is utilized to determine variant inclusion and a concise, information-rich GGR is drafted using the GeneInsight® software system. All results are summarized on a single page with ~3 additional pages of supporting material providing the evidence for disease association, disease and phenotypic descriptions, details on pharmacogenomic associations, and serologically confirmed blood antigen results. Our experience from The MedSeq Project highlights important considerations in the reporting of secondary findings and provides a framework for interpretation and reporting practices in clinical genome sequencing.

2537W

Analytical Performance of a Next-Generation DNA Sequencing-based Clinical Workflow for Genetic Carrier Screening. *G.J. Porreca, M. Umbarger, C. Kennedy, P. Saunders, B. Breton, N. Chennagiri, D. Maganzini.* Good Start Genetics, Cambridge, MA.

Next-generation DNA sequencing (NGS) is poised to displace genotyping technology for clinical applications because it promises richer information at low cost. However, to date a number of considerations including sequence accuracy and completeness, as well as workflow scalability, have limited its adoption in the clinical laboratory. Here we describe a NGS-based platform designed for genetic carrier screening in a clinical setting. A set of 15 genes are isolated from genomic DNA by automated multiplex target capture, tagged with molecular barcodes, and pooled and sequenced on the Illumina Hiseq system. Reads from each sample are de-multiplexed, aligned to a reference, and integrated into accurate genotype calls which are then interrogated for pathogenic mutations. A total of 42,858 bp were targeted for capture by a set of molecular inversion probes designed to tile across the target such that each base was captured by at least three different probes. Across a set of 182 DNA samples derived from cell lines or blood, a median of 99.71% of bases were sequenced to sufficient depth and quality for inclusion in genotype calling. Replicate runs exhibited a high level of concordance, with 17 discordant single nucleotide variant (SNV) calls out of 5,177,206 across 126 samples. Concordance of genotype calls with bi-directional Sanger sequence of PCR amplicons derived from a set of 194 samples was 99.97% at SNV positions (1 out of 4,001 Sanger SNV calls discordant), and 99.9999% at non-variant positions (8 out of 6,992,754 Sanger non-variant calls discordant). We identified, in a set of 55 samples, a total of 92 mutations (out of 92) that have been previously reported to be causative of recessive Mendelian disease. We also determined that 9 previously uncharacterized samples contained mutations that were either known or expected to be pathogenic. Clinical carrier screening has traditionally been performed using genotyping technology, and demands high analytical accuracy. Our NGS platform achieves exceedingly high concordance with Sanger sequencing, long considered an accuracy gold standard. Furthermore, we made high-confidence genotype calls across the vast majority of basepairs within our target regions. Taken together, these results indicate that NGS, when paired with the appropriate sample preparation methodology, automation, and data analysis, can deliver the performance required for use in clinical carrier screening.

2538T

Targeted exome sequencing identifies two pathogenic *DYNC2H1* variants in a fetus with short-rib-polydactyly syndrome. *K.I. Varvagiannis[1,2], P. Makrythanasis[2], F. Santoni[2], J.-M. Pellegrinelli[3], P. Extermann[4], C. Brockmann[1], C. Gehrig[1], M. Guipponi[1,2], J.-L. Blouin[1,2], S.E. Antonarakis[1,2], S. Fokstuen[1].* 1) University Hospitals of Geneva, Service of Genetic Medicine, Genome Clinic, Geneva, Switzerland; 2) University of Geneva, Department of Genetic Medicine and Development, Geneva, Switzerland; 3) University Hospitals of Geneva, Department of Gynecology and Obstetrics, Geneva, Switzerland; 4) Dianecho, Geneva, Switzerland.

The short-rib-polydactyly syndromes (SRPS) represent a heterogeneous group of ciliary skeletal dysplasias characterized by a narrow thorax, short ribs and limbs and polydactyly. To date, more than 10 genes have been associated with this mainly autosomal recessively inherited phenotype.

We applied targeted exome sequencing on an affected fetus presenting with micromelia, postaxial polydactyly of both hands and feet, short ribs and a severe cardio-thoracic disproportion. The pregnancy was terminated at 18+3 gestational weeks based on the ultrasound findings.

Exome sequencing was performed on the fetus' DNA using an Illumina HiSeq2000. Bioinformatic analysis was carried out using BWA, samtools, pindel and ANNOVAR. Analysis was targeted to the 10 genes previously implicated in SRPS: *NEK1, DYNC2H1, IFT80, WDR35, EVC2, EVC, TTC21B, WDR19, IFT122, IFT43,* while masking the rest of the detected variants. An 8x and 30x coverage of 99.57% and 98.29% respectively, was achieved for the 10 genes. Variant calls were annotated and filtered for their presence in known genome databases (1000Genomes, dbSNP, Exome Variant Server) as well as in our in-house variant database. Two candidate variants were identified in *DYNC2H1* - a splicing variant (NM_001377.2:c.1953G>A) previously reported in the literature and a novel missense (NM_001377.2:p.(Ala1542Val)) - as well as a 16 bp insertion in *TTC21B* leading to a frameshift and a premature termination codon (NM_024753.4:p.(Thr8Valfs*4)). Sanger sequencing was performed for confirmation of the variants and verification of their phase. Both parents were found to carry one *DYNC2H1* variant, while the mother was carrier of the *TTC21B* insertion, establishing the genetic diagnosis of SRPS in the fetus.

While the analysis was ongoing the couple went through a second termination of pregnancy due to recurrence of an affected fetus. Sanger sequencing, revealed the presence of the two *DYNC2H1* variants but absence of the *TTC21B* variant.

Our results demonstrate that targeted exome sequencing is an efficient and cost-effective approach for the identification of pathogenic variants in genetically heterogeneous disorders.

2539F

Validation of an Accurate, High-Throughput Multiplex qPCR Assay to Confirm CMA Clinical Findings. *L.E. Northrop, V. Aggarwal, V. Jobanputra, M. Mansukhani, B. Levy.* Columbia University, Department of Pathology & Cell Biology, Division of Personalized Genomic Medicine, New York, NY.

Chromosomal microarray analysis (CMA) is the standard cytogenetic screening method in clinical constitutional laboratories, especially in patients with developmental delay and intellectual disability. The demand for CMA in prenatal patients is likely to increase following the findings of the recent NICHD multicenter prenatal study, recognizing the need for a more comprehensive screening technology. Confirmation of abnormal CMA results is easily achieved by Fluorescence *in situ* hybridization with BAC probes (BAC-FISH), especially when dealing with imbalances >500kb in size. However confirmation of duplications <500kb is often difficult due to interpretive challenges and confirmation of deletions <200kb there is often no available BAC probe. Furthermore, the cost of BAC probes can be high with a long turn-around-time (TAT) of several weeks. Given these limitations, we investigated an alternative molecular-based method for confirmation of abnormal CMA results. Several clinical labs use qPCR copy number (CN) testing for confirmation of their abnormal CMA results. However, there are no universal published guidelines on the design and validity of qPCR as a confirmation testing method in a clinical setting. A validated qPCR method for CN confirmation is merited in abnormal CMA cases with smaller imbalances. Especially in clinical labs that need to be in compliance with New York State Standards of Clinical Laboratory Practice. We have designed a cost-effective, high-throughput RT-qPCR method with a quick TAT for the confirmation of CMA aberrations that are reported in our clinical laboratory. A brief description of our design consists of taking direct genomic DNA to a 96 well plate with 28 references run in triplicate, the proband, parents &/or sibships (if available). We developed a multiplex assay which includes the 'test' gene and 'endogenous' gene in a combined SensiFast lo-rox master mix (Bioline) detected by fluorescence probe capture (IDT). This method is highly accurate & specific in confirming CN for CMA with a TAT of approximately one week from design to result. By this approach, we have accurately confirmed CMA abnormalities that could not be confirmed by BAC-FISH (either because of probe availability or due to BAC size limitations). Our validation cohort consists of imbalances as small as 5kb and as large as 483kb. We propose qPCR as the standard method when confirming CMA duplications <500kb and deletions <200kb in a clinical setting.

2540W

Detection and quantification of somatic mutations in Klippel-Trenaunay Syndrome using digitally counted nanodroplets. *N.M.K. Kamitaki[1], V. Luks[2], R. Murillo[2], S.A. McCarroll[1], M. Warman[1,2,3].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA; 3) Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA.

All human tissues contain cells with somatically acquired mutations. Those mutations that cause disease may be difficult to identify. When the fraction of cells having a somatic mutation is high in an available tissue, the disease causing mutation can be identified using commonly applied sequencing methods, such as massively parallel sequencing of exomes or genomes, or PCR-based amplifications and sequencing of candidate genes. When the frequency of mutant cells is low, mutation detection becomes much more challenging with these methods. Our molecular approach is based on using allele-specific digital PCR in nanodroplets. We designed primers and allele-specific probes, where each reference and mutation-specific probe for a given site have different fluorophores, to target five of the most common *PIK3CA* mutations found in overgrowth syndromes and cancer. We start off with a standard PCR mix and after adding the sample, emulsify the solution into tens of thousands of nanodroplets, each of which contains the materials requisite for PCR, but only some of which contain a template DNA molecule from the patient's genome. We count the number of droplets positive for each fluorophore using a Bio-Rad QX1000 Droplet Reader. This allows digital counting of the abundance of each allele, allowing us to not only detect if mutant alleles are present, but also to quantify their frequency. We previously identified somatic mosaic missense mutations in *PIK3CA* in three patients with Klippel-Trenaunay Syndrome (KTS). Using this molecular approach to screen affected tissue samples from 28 individuals with KTS we identified mutant allele frequencies ranging from 2.7% to 16% in 21 of these patients. This suggests that the majority of patients with KTS have recurring mutations in *PIK3CA*. Our results do not preclude all KTS being caused by somatic mutations in *PIK3CA*, since our assay was designed to query only five common mutations. Advantages of digital PCR using nanodroplets include low reagent costs, rapid turnaround, and ease in adding new mutations to the assay. Additionally, we have been able to consistently detect 0.1% mutation frequencies within dilution series, demonstrating the high sensitivity of this assay. Using this technique to initially screen affected tissue for *PIK3CA* mutations will reduce the number of samples that will need to be tested using more expensive and time consuming mutation detection methods.

2541T

Prenatal Testing of novel mutations of Maple syrup urine disease by next-generation sequencing. *S. Chen, X. Li, H. Ge, X. Pan, F. Chen, H. Jiang.* BGI-Shenzhen, Shenzhen, Guangdong, China.

Objectives: The applications of massively parallel sequencing technology to fetal cell-free DNA (cff-DNA) have brought new insight to noninvasive prenatal diagnosis. However, most previous studies based on maternal plasma sequencing have been restricted to fetal aneuploidies. Our study aim to report the first case to combine the target capture next-generation sequencing (NGS) for identifying new mutations in unexplained Mendelian disorders cases and NIPT of single-gene disorders by maternal plasma DNA sequencing together, to prove the feasibility and potential of clinical integration. **Methods:** We recruited a pregnant couple with a maple syrup urine disease (MSUD) proband child to develop our method. Target capture next-generation sequencing was performed to identify new mutations in the related gene. The maternal plasma was isolated for DNA extraction and target sequencing with a semi custom array. Then, a haplotype-assisted strategy was developed to detect whether the fetus is inherited the pathogenic mutations in target gene. **Results:** We identify a heterozygosis duplication in exons 2-4 of gene BCKDHA in proband and father, a potential variant, c.392A>G in BCKDHA, in proband and mother. The parental haplotype was constructed successfully in this trio family by target region sequencing. Then, a sensitive Hidden Markov Model (HMM) was used to identify the parental transmitted allele and recombination breakpoints in the maternal plasma. In the pathogenic gene, BCKDHA, the fetal inherited the same alleles with the proband, which indicated it was also a MSUD patient. Real-time PCR and Sanger sequencing performed on DNA samples of amniotic fluid and umbilical cord blood were consistent with our diagnostic results. **Conclusions:** Target capture and NGS have significant efficacy and scalability for identifying new genes or new mutations in unexplained Mendelian disorders cases and NIPT by maternal plasma DNA sequencing has been proved to be feasible for noninvasive detection of single-gene disorders. This is the first report to use the complicated pipeline in real clinical setting, to combine the target capture next-generation sequencing for identifying new mutation and noninvasive prenatal testing by maternal plasma DNA sequencing for prenatal testing together, which indicate the potential use in routine clinical practice.

2542F

New allele-specific real-time PCR system with automatic interpretative program for genotyping of SNPs related to pharmacogenetics. *D. Kim, S. Byun, J. Seo, C. Lee.* Molecular Diagnostics Division, Bioneer Corporation, Daejeon, Korea.

Various human SNP (single nucleotide polymorphism) markers have been found to be involved in drug metabolism. The accurate analysis of genetic variations of particular genes encoding drug metabolic enzymes is critical for pharmacogenetics. Direct sequencing is used as a gold standard for genotyping of candidate genes, however, a faster and higher-throughput method is often required for the tests in clinical laboratories. Recently, we developed a new real-time PCR system, ExiGenotyper, with the automatic interpretation program for SNP genotyping, and genotyping kits of VKORC1 and CYP2C9, CYP2C19, and TPMT (thiopurine S-methyltransferase). SNP genotyping of VKORC1 and CYP2C9 genes is useful for determining the optimal warfarin dose. The CYP2C19 gene is a subtype of cytochrome p450 which is related with clopidogrel dosing. The TPMT gene encodes a protein that catalyses S-methylation of aromatic and heterocyclic sulfhydryl compounds, and TPMT activity can be estimated through the SNP genotyping of the TPMT gene. The allele-specific primer method was used for real-time PCR in this system for the determination of SNP genotype. For the evaluation of the system and the genotyping kits, SNP verified clinical samples by direct-sequencing were used. ExiGenotyper warfarin genotyping results of 82 human genomic DNA samples agreed with direct sequencing results. ExiGenotyper CYP2C19 genotyping results of 59 human genomic DNA samples were in concordance with direct-sequencing and ExiGenotyper CYP2C19 genotyping results of 26 samples agreed with results from a commercially available POC device. ExiGenotyper TPMT genotyping results of 246 human genomic DNA samples were in agreement with direct sequencing. The results suggested that the new system we developed allows accurate and rapid genotyping of SNP and is applicable in clinical laboratories.

2543W

Discrepant Tay-Sachs disease enzyme and DNA carrier screening results in the African American population. *D. Neitzel, H. Travassos, N. Faulkner, S. Hallam, V. Greger.* Good Start Genetics, Cambridge, MA.

Tay-Sachs disease (TSD) is a common autosomal recessive condition in the Ashkenazi Jewish population. It has also been observed at higher frequencies in other populations. TSD carrier screening is currently recommended by ACOG and other societies for high-risk populations, as well as for couples where at least one member is high-risk. Screening consists of both DNA analysis of the HEXA gene as well as HexA enzyme analysis. Screening has led to a significant reduction in the incidence of TSD in the high-risk populations. Today, most new cases of TSD are born to couples where at least one member of the couple does not belong to a high-risk population. According to ACOG's position statement, 'biochemical analysis should be used for individuals in low-risk populations'. At the time of result reporting, we noted a high percentage of indeterminate and positive TSD enzyme results without mutations present in the African American population (AA), leading us to retrospectively evaluate the data on all samples tested. Good Start Genetics' TSD DNA sequencing test detects 67 known disease-causing mutations (including the 7.6kb deletion) plus novel truncating mutations. HexA enzyme activity was assessed on leukocytes. Out of 2656 individuals tested, 144 self-reported as AA. 16/144 (11%) were enzyme positive and 46/144 (32%) were indeterminate. In comparison, 15/1038 (1.5%) were enzyme positive and 82/1038 (7.9%) were indeterminate in self-reported Caucasians. No mutations or pseudodeficiency alleles were identified in any of the TSD enzyme positive or indeterminate AA individuals. Based on enzyme data alone, the observed carrier frequency in our AA population is 1/9 versus an expected frequency of 1/300. Given the low incidence of TSD in the AA population and the fact that no mutations could be identified in these samples, this result raises questions about the reliability of enzyme analysis in the AA population. Another commercial laboratory has also reported a higher than expected percentage of TSD enzyme indeterminate and positive results in the AA population without identifying any DNA mutations. However, only the most common mutations were tested. Our data is based on more extensive DNA analysis, and we corroborate their findings. The reference ranges for HexA enzyme activity need to be re-evaluated. Despite guidelines, enzyme analysis is NOT currently an accurate method to assess TSD carrier status in the AA and possibly other populations.

2544T

Diagnostic Sequencing - implemention into routine processes. *K. Stangier, T. Paprotka.* GATC Biotech, Konstanz, Germany.

Next generation sequencing is on its way to play a more and more important role for analyzing genetic and genomic diseases and finding its way into clinical diagnostics. We will present the workflow of the implementation of next generation sequencing into routine processes for diagnostic and clinical purpose. Starting from the approach to design and set up pilot projects through validation studies the whole cycle till the product launch will be illuminated. Data on sensitivity and sensibility from a clinical study as well as from the final product will be shown. The data sets include analysis of genetic disorders, cancer and prenatal diagnostic of more than 7,500 patients. Also, the implementation of the final test in a ISO accreditation will be described. Utilizing Next Generation Sequencing in clinical and genetic diagnosis and personalized disease risk profiling is very important in the future. Further optimization of samples isolation and preparation as well as sensitivity and data analysis pipelines will help to integrate Next Generation Sequencing into a common clinical setup.

2545F

Molecular Genetic Diagnosis of Fanconi Anemia in Chinese patients.
X. Chen[1], H. Liu[2], W. Teng[2], F. Wang[2], Y. Wang[2], Q. Yin[2], M. Wang[2], L. Guo[2], P. Zhu[1]. 1) Department of Hematology, Peking University First Hospital, Beijing, China; 2) Molecular Medicine Lab, Hebei Yanda Hospital, Hebei, China.

Background: Fanconi anemia (FA) is a group of hereditary diseases, the typical clinical manifestations including anemia and multiple congenital malformations. However, FA patients didn't always with malformation, thus gene mutation inspection is useful in diagnosis and differential diagnosis of FA and autoimmune aplastic anemia (AA), aplastic anemia - paroxysmal hemoglobinuria syndrome (AA-PNH). Method and Cases: FANCA, FANCC, FANCG gene mutation analysis by PCR and Sanger sequencing, PNH phenotype detection by flow cytometry and chromosome breakage test was performed in 28 cases which clinically suspected FA. Results: 1) 12 cases were identified carrying FA gene mutation, 4 males and 8 females. Bone marrow failure symptoms began at 0.5 to 15 years old, with a median age of 5 years old. 2) 10 cases carrying FANCA mutations, 1 with FANCC mutations, 1 with FANCG mutations. Biallelic mutations were identified in 3 cases, respectively were FANCG W184X homozygous, FANCA L1249RfsX16 and 3067-2A>C, FANCA H330LfsX2 and P1324L. Only one heterozygous mutation were identified in each other 8 cases, including FANCA Q123X, R413C, Q563R, A840P, L871V, S858R, E878RfsX8, A958V and FANCC R334W. Pedigree analyses were performed in 5 families and all confirmed being hereditary. 3) 6 out of the 12 gene mutation-positive cases clinically with café au lait spots and/or organ malformations. Chromosome breakage test were positive in 4 out of 6 gene mutation-positive cases. 4) A significant proportion of PNH cells were detected by flow cytometry in one FANCA mutated patient, with CD55 (-) 66.68% and CD59 (-) 68.81% in granulocyte. However, the patients showed multiple sporadic mutations rather than clonal mutations in PIGA gene. Conclusions: Gene mutation inspection is useful in diagnosis and differential diagnosis of FA, AA and AA-PNH.

2546W

Estimating the contribution of unidentified mutations in autosomal recessive disorders.
L.P. ten Kate[1], M.E. Teeuw[1], A. Sefiani[2,3], F-Z. Laarabi[2,3], I. Hama[3], L. Henneman[1], M.C. Cornel[1]. 1) Dept Clinical Genetics, VU Univ Med Ctr, Amsterdam, The Netherlands, Netherlands; 2) Dept Medical Genetics, Nat. Inst. of Health, Rabat, Morocco; 3) Center of Human Genomics, Fac. of Medicine and Pharmacy, Univ. Mohammed V Souissi, Rabat, Morocco.

It has been shown before that the total pathogenic gene frequency (q) of an autosomal recessive disorder can be estimated from existing mutational data, provided that (a) the overall inbreeding coefficient (F) of patients in the sample is known and above zero; and (b) for all patients in the sample bi-allelic mutations are found (1,2). Not infrequently however, some patients reveal only one mutation or none at all even though they meet the clinical criteria. For instance, in a sample of 175 Moroccan Familial Mediterranean Fever (FMF) patients 43 were found with bi-allelic MEFV mutations (27 homozygotes and 16 compound heterozygotes), 23 with only one identified mutation, and 109 without any identified mutation. Average F in this sample was 0.0122. Suppose we may assume that all 23 patients with only one identified mutation are in fact compound heterozygotes, with the second mutation unidentified, how many of the 109 patients without an identified mutation would then be homozygotes of the same unidentified mutation? This question can be solved because there is a fixed ratio between homozygotes (HO) and compound heterozygotes(CH) for any given allele ai, depending on F and q only, as shown in the following equation: HO/CH = $[F + (1\text{-}F)\, a_i.q]/[2q(1\text{-}F)(1\text{-}\, a_i)]$, in which a_i represents the relative frequency of allele a_i. Taking now the observed HO/CH ratios of identified alleles as a reference, one will be able to make a proximal estimate of the number of homozygotes with the unidentified allel (in this case 17), and subsequently the total pathogenic gene frequency including the unidentified mutation. As the HO/CH ratio will be smaller when there are two or more different unidentified alleles, and as some of the heterozygotes may in fact be real heterozygotes, the above estimate is a maximum one. So, at least 92 of the FMF cases in this sample were not caused by recessive MEFV mutations. Other explanations for their existence are needed. (1) Ten Kate et al. J Community Genet 2010; 1: 37-40 (2) Giallusi et al. Eur J Hum Genet 2013; Mar 13, doi: 10.1038/ejhg.2013.43 [Epub ahead of print].

2547T

The effect of Long-Term Frozen Storage of Urine Samples on the detection of Chlamydia trachomatis in comparison to Preserved - Room temperature Urine Samples.
M.AK Abdalla[1], M. El-Mogy[2], C. Moreira[3], R. DiPietro[4], T.A Haj-Ahmad[5], Y. Haj-Ahmad[3,5]. 1) Department of Biochemistry, Faculty of Science, Alexandria University, Egypt; 2) Molecular Biology Department, National Research Center, Dokki, Cairo, Egypt; 3) Brock University, 500 Glenridge Avenue, St. Catharines, ON, L2S 3A1; 4) Department of Biochemistry, Faculty of Science, McMaster University, ON, Canada; 5) Norgen Biotek Corp., 3430 Schmon Parkway, Thorold, Ontario, Canada, L2V 4Y6.

Chlamydia trachomatis is a widespread sexually transmitted disease that can be avoided by the early detection of the pathogen. Urine is a biological sample that can be considered for the detection of the pathogen. However, some studies have shown that long terms frozen urine is not suitable for molecular based assays, likely due to degradations. The main goal of this work was to compare the effect of urine chemical preservation and storage at room temperature and urine stored at 4ºC and -70ºC on the quality of the isolated DNA and hence the detection of pathogens in urine. Briefly, urine sample was spiked with known amount of Chlamydia trachomatis, and then subdivided into three sets of samples: One set of samples were stored at 4ºC, the second set at -70ºC and to the third set was chemically preservative and stored at room temperature. Total genomic DNA was isolated both at time zero and after three months. Total Urinary DNA as well as Chlamydia DNA were purified and were analyzed qualitatively using conventional and Real-time PCR. The data shows the effect of storing urine samples at different storage conditions, with or without preservative, on the quality of the isolated DNA and hence on the detection of Chlamydia from urine.

2548F

De-novo occurrence of SMN1 deletions: parental origin and frequencies.
K. Eggermann, T. Eggermann, K. Zerres, S. Rudnik-Schöneborn. RWTH Aachen, Aachen, Germany.

Spinal muscular atrophy (SMA) is a severe neuromuscular disorder characterized by degeneration of anterior motor neurons, resulting in a progressive muscle weakness and paralysis. SMA is caused by mutations in the survival motor neuron 1 (SMN1) gene in 5q13. The homozygous deletion of the SMN1 gene accounts for nearly 95 %; of SMA patients and the vast majority of parents are heterozygous carriers, as de-novo mutations occur in only 1-2 % of patients. However, in a number of parents two SMN1 copies can be detected. This can either be explained by two SMN1 copies on the same chromosome 5q13 and the deletion on the other (genotype 2 +0), or by a de-novo deletion and a regular SMN1 distribution in the parent (genotype 1+1). Several tests for SMN copy number analysis (e.g. MLPA, qPCR) have been developed, aiming at the identification of heterozygous carriers of the SMN1 deletion. However, they do not allow the differentiation between carriers of the 2+0 and the 1+1 genotype. Interestingly, in the patients with a de-novo deletion a significant preponderance of paternal chromosomes affected by the mutation has been observed (12 paternal de-novo deletions versus 2 maternal ones). Here we report on our molecular findings of more than 200 parents of SMA patients with a homozygous deletion. All probands were investigated by MLPA; in parents with two SMN1 copies microsatellite typing was performed. In 7 parents two SMN1 copies were identified, among them four with a 2+0 genotype. Three parents had a regular 1+1 genotype (non-carriers). The de-novo deletion twice affected the maternal chromosome and once the paternal SMN1 copy. Our results and those from the literature (n>830) show that (a) about 5 % of SMA parents carry two SMN1 copies, (b) about one third of those can be explained by de-novo deletions in the patients. We also conclude that the previously observed imbalance between paternal and maternal chromosomes affected by the de-novo deletion may be biased by the small number of cases. Finally we want to emphasize that the differentiation between the different genotypes in SMA parents (2 +0 versus 1+1) does not only provide information that will assist the SMA families in further family planning but it also contributes to a more accurate risk assessment for further relatives.

2549W

A Connective Tissue Disorders NGS Panel: Development, Validation, and Novel Findings. *J. Lee, M. Basehore, S. McGee, K. Kubiak, K. King, K. Champion, J. Jones, M. Friez.* Greenwood Genetic Center, Greenwood, SC.

Connective tissue disorders represent a heterogeneous group of more than 200 recognized conditions for which the connective tissues are the primary pathologic target. Connective tissues are the structurally supportive components of the body that form a framework or structural matrix, connect body tissues, and provide protection of organs and storage of energy. While certain connective tissue disorders are associated with definitive phenotypes, some patients may have non-specific or atypical presentations. Many patients might also present with similar features due to the clinical variability and variable expressivity observed among these disorders. For these cases, we have developed and validated a targeted sequencing panel for routine clinical diagnostic testing, using RainDance™ microdroplet enrichment and SOLiD™ Next Generation Sequencing (NGS), to analyze the coding and flanking intronic regions of 31 genes associated with various connective tissue disorders simultaneously. This method enables the detection of single nucleotide changes and small duplications and deletions in genes that, collectively, are associated with at least 80 distinct connective tissue disorder phenotypes. Molecular testing was performed using our diagnostic NGS panel on an initial cohort of 22 patients with suspected connective tissue disorders. We identified novel likely-pathogenic changes in four (18%) of these patients, including a heterozygous frameshift alteration in *COL5A1* in a patient with a phenotype consistent with Ehlers-Danlos syndrome, an apparently homozygous frameshift alteration in *ZNF469* in a patient with features of Brittle Cornea syndrome, a heterozygous nonsense alteration in *TNXB* in a patient with joint hypermobility, and a heterozygous frameshift alteration in *COL1A1* in a patient with features of Osteogenesis Imperfecta. Of the remaining patients, only two (9%) were found to have normal results, with no molecular alterations, and 16 (73%) had at least one variant of uncertain clinical significance. These preliminary findings suggest that our targeted NGS panel may have a potentially high diagnostic value for patients with suspected connective tissue disorders. Furthermore, the simultaneous analysis of a panel of 31 connective tissue disorder genes enables a more efficient diagnostic strategy compared to a single-gene approach using traditional Sanger sequencing.

2550T

Validation of clinical NGS vs. sanger sequencing: Measuring the value of orthogonal testing. *S. Lincoln, S. Kash, Y. Kobayashi, G. Nilsen, J. Sorenson, M. Cargill, R. Scott.* InVitae, San Francisco, CA.

Next Generation Sequencing (NGS) technology can help expand clinical use of constitutional genetic testing by allowing large numbers of genes to be rapidly tested in patients at low cost. However, broad adoption of NGS requires demonstration of clinical-grade accuracy, in part to address concerns about false positive findings in increasingly large gene panels. Because of these concerns many clinical NGS labs confirm pathogenic variants in patients using Sanger sequencing, adding to cost and slowing turn-around times.

We have embarked on a set of studies to measure the performance of clinical NGS by comparison with Sanger, and these studies may help determine the need for orthogonal confirmation. We apply NGS to individuals who have undergone traditional genetic tests, including both patients with various indications as well as reference DNAs from the Get-RM Program. We enrich for individuals with pathogenic variants (~50%) to measure sensitivity as well as specificity, and among those individuals we enrich for certain variation classes known to be challenging for NGS. At the same time, one study cohort is being selected prospectively to include unbiased representation of all patients indicated for genetic testing under current clinical guidelines.

To date over 200 individuals have been tested by Illumina NGS and also by Sanger, with our NGS lab blinded to the other results. In these we see 100%; concordance for both variant and reference bases called. We plan to expand these studies to include about 1000 samples by the time of the ASHG 2013 meeting and to 2000 thereafter.

We are also augmenting the study with 9 HapMap samples, for which we have collected multiple data sets into a 'gold standard' for over 200 clinically relevant genes. These samples allow many variant and invariant positions to be studied per sample, with the caveat that known pathogenic variants are found less frequently than they are in clinical samples.

NGS data quality depends on careful choice of platform, assay, algorithms, parameters and QC criteria. We will describe our methods in addition to the validation study details. We intend to quickly make as much of the above data public as possible.

2551F

Success and flaws when using NGS for molecular diagnosis. *K. Rocha, M. Lazar, G. Yamamoto, M. Aguena, V. Takahashi, R. Pavanello, M. Passos-Bueno.* Gen e Biol Evolutiva, Inst Biocien Univ Sao Paulo, São Paulo, SP, Brazil.

The Human Genome Research Center (HGRC) offers molecular diagnosis to several diseases, mainly neuromuscular disorders and some cases of skeletal dysplasias using traditional molecular biology techniques such as MLPA, for detecting targeted deletions or duplications, or Sanger sequencing, in cases caused by small nucleotide changes. Unfortunately, a considerable number of cases of these disorders remain unconcluded, mostly because several of them are characterized by clinical and genetic heterogeneity, while in some others, as Duchenne muscular dystrophy, where around 30% are caused by point mutations, the high cost of Sanger sequencing method is unaffordable for most of the Brazilian families. In order to expand the number of conclusive molecular diagnosis, we designed a custom next-generation sequencing panel containing approximately 200 genes for MiSeq (Illumina) sequencing. These genes are associated with two main disease groups: neuromuscular disorders and skeletal dysplasias. We analyzed 26 patients: in six of them, the disease-causing mutation was known (control group), 15 had clinical symptoms compatible with neuromuscular disorders and 5 with skeletal disorders. We were able to detect the pathogenic mutations in five out of six control patients (83.3%). Among the neuromuscular patients, we identified disease causing mutations in 11(73.3%), but we were only able to confirm the clinical hypothesis for 2 skeletal dysplasia patients (40%). The low detection rate in the later group could be explained by two main reasons: the absence of the gene associated with their clinical features in our panel and lack or insufficient coverage at the targeted region due to the high CG content of some genes (e.g. *TWIST1* and *COL18A1*). The low coverage of some exons of *COL18A1* also explains why we could not detect, in the control group, all the previously known mutations. In spite of that, our experience has been positive: we identified mutations in at least 8 different genes or regions that we currently do not offer testing by Sanger sequencing, including *DMD*. In conclusion the results have shown that NGS is a powerful tool to improve the detection of mutations for clinical diagnosis. FAPESP/CEPID, CEGH.

2552W

Clinical utility of next generation sequencing for the molecular diagnosis of genetically heterogeneous retinitis pigmentosa. *J. Wang[1], V.W. Zhang[1], FY. Li[1], C. Truong[1], G. Wang[1], P.W. Chiang[2], R.A. Lewis[1,3], L.J. Wong[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Casey Eye Institute, Oregon Health and Science University, Portland, OR; 3) Department of Ophthalmology, Baylor College of Medicine, Houston, TX.

Background: Retinitis pigmentosa (RP) is one of the most genetically heterogeneous disorders. Mutations in over 100 genes have been associated with both non-syndromic and syndromic RP. Identification of disease-causing mutations is essential for genetic counseling, carrier testing, and future gene-specific therapies. Simultaneous sequencing of multiple genes for RP by a high throughput next generation sequencing (NGS) is a cost-effective approach to that goal. . Methods: We developed a target gene capture sequencing (TCS) approach with the NimbleGen solution-based capture design of 66 genes currently known to cause RP, followed by NGS analysis on Illumina HiSeq2000. Results: A total of 202,800 bp of target sequences including 939 coding exons and 20 bp of flanking intron regions were sequenced to an average depth of 700X per base. Six consistently poorly covered exons were completed by PCR/Sanger sequencing to ensure 100% coverage for the entire coding regions. The phase I validation sample has a total 135 variants being identified, which are 100% concordant with Sanger sequencing results. Phase II validation was performed on 12 samples with known mutations in 4 different genes: ABCA4, RP1, RPE65, and USH2A. All previously detected disease-causing mutations were identified correctly. We also have analyzed 25 samples from unrelated individuals with pigmentary retinal dystrophies, LCA, or RP-related disorders. Deleterious mutations were detected and confirmed in 19 patients with autosomal recessive, dominant and X-linked inheritance. Our positive detection rate is 76%. In addition to single nucleotide substitution and small indels, homozygous insertion of a 355 bp Alu sequence in exon 10 of the MAK gene in two unrelated patients, and a homozygous exonic deletion in the EYS gene in one patient were detected in this cohort. Conclusion: Our data underscore the important clinical utility of NGS-based analysis in the molecular diagnosis of RP. A strategic data analysis has proven accurate for mutation- and variant-identification and annotation. The deep exonic base-to-base coverage of all coding regions plus Sanger confirmation allow the accurate identification of all point mutations, small indels, and even large Alu insertions, which is difficult to detect from NGS data. The TCS approach greatly improves the diagnosis of RP in a cost and time efficient manner.

2553T

Evaluation of fragile X screening methods for early detection of affected infants. *P. Mueller, J. Lyons, G. Kerr.* Newborn Screening and Molecular Biology Branch, CDC, Atlanta, GA.

Purpose: Expanded CGG repeats (≥ 200 repeats) in the FMR1 promoter are known to cause Fragile X in males; premutations (55 to 199 repeats) cause fragile X-associated primary ovarian insufficiency in females and tremor/ataxia syndrome in both males and females. Since the GC content in these expanded repeats is near 100%, the larger CGG repeats are very difficult to PCR amplify while normal repeats amplify preferentially. Therefore, it is difficult to distinguish between normal homozygous females (20% to 40% of females), and those with one normal allele and a CGG expansion. The goal of this study was to evaluate and modify, as needed, published methods for detecting CGG repeat expansions in the FMR1 gene promoter region for potential use as a Fragile X screening assay. Methods Used: We have evaluated and tested modifications of PCR-based screening methods including those using CGG targeted primers, melting point curves, and heat pulses during PCR extension to facilitate the detection of large CGG repeats. Results: Our analysis of 80 normal, 3 gray-zone, and 38 pre- and full mutation samples using the protocol by Tassone that includes a CGG targeted primer (J Mol Diagn, 10:43-49, 2008) called all correctly except one female heterozygous normal and expanded sample. The Orpana (Anal Chem, 84:2081-2087, 2012) heat pulse method using agarose gel electrophoresis detects pre- and full mutation males since there is no visible product for the normal allele even if the amplicon from an expansion is not evident. Female full expansions with a normal allele are less evident. The Teo method (Clin Chem 58:568-579, 2012) using a melting curve analysis of triplet-primed PCRs in both the 5' and 3' directions resulted in optimal temperature cutoffs that were different from the published method necessitating optimization in each laboratory. The 3' assay gave the more robust optimal combination of sensitivity and specificity for pre- and full mutations. Conclusions: The Tassone method gave the best performance for both males and females, while the Orpana method worked well for males. The 3' Teo assay was more robust than the 5' assay. A cost analysis based on reagent and supply cost with amortization of major equipment over 5 years, assuming 55,602 samples/year (median of annual state births, 2010), indicated that the Orpana method was the least expensive at $2.25/sample, followed by the Tassone method at $4.25/sample and the Teo method at $5.71/sample.

2554F

Development and implementation of AJPNxt, a 51 mutation Ashkenazi carrier screening panel built on the Illumina BeadXPress™ platform. *J.M. Buis, S. Birkeland, J. Sugalski, C. Holland, J. Stoerker.* Research & Development, aMDx Laboratory Sciences, Ann Arbor, MI.

Members of the Ashkenazim are at elevated risk for a number of inherited genetic syndromes. Widespread and successful adoption of carrier screening amongst the Ashkenazi Jewish community has greatly reduced the incidence of these inherited diseases. We have developed an expanded screening panel for 51 mutations across 18 diseases prevalent in this community using the Illumina, Inc. BeadXpress Platform. This panel comprises 8 multiplex PCR amplification reactions and subsequent allele specific primer extensions pooled for the detection of 103 independent wild type or mutant signals at once via VeraCode™ technology. Due to widespread adoption of carrier screening in the Ashkenazi community and population dispersion, a significant percentage of recently born disease affected individuals have either uncommon or de novo mutations. As such, we have included the addition of several less common variants across multiple diseases and will track incidence of these for justification of additional rare variants of pathological importance in future panels. This laboratory developed test was validated using a mutation enriched third party blinded set of 479 samples containing 232 known positive samples, from mixture of controls or samples split with an authority laboratory. The validation proved the test is 100% sensitive for mutations tested and greater than 99% specific. Here, we report on the first cohort of patients screened across a geographically diverse section of the US over the first 8 months of testing. This includes our initial individual mutation disease rates, overall detection rates across our panel and the serendipitous finding of homozygotes.

2555W

Development of a Target-Capture Gene Panel for Clinical Genomic Sequencing in Bronx, NY. *M. Delio, K. Patel, A. Maslov, J. Cai, J. Shan, S. Maqbool, B. Calder, A. Golden, J. Greally, B. Morrow, J. Vijg, C. Montagna.* Albert Einstein College of Medicine, Bronx, NY.

Next-generation sequencing (NGS) has the great potential to improve clinical care for children to diagnose rare Mendelian diseases as well as for adults to provide specific treatments such as for various cancers. Based upon medical need of individuals receiving clinical care at Montefiore Medical Center, we designed a custom gene panel, spanning 5 Mb and consisting of 650 genes targeting known Mendelian loci, some pediatric diseases and several hotspot genes in various cancer types. To determine the sensitivity and specificity of our custom panel we also used several predesigned panels (Ion Torrent) and Whole Exome Sequencing (WES) on cancer samples. The gene panel was created using the Roche-NimbleGen SeqCap V3 capture system and samples were sequenced using the Illumina HiSeq 2500 Rapid Run technology. A total of 24 DNA samples were analyzed, including 7 control DNA samples with known variants. Sequence data was annotated and variants were identified through the Wiki-based Automated Sequence Processor (WASP) analytical pipeline (PMID: 22942009) and customized analysis using the Broad Institute Genome Analysis Tool Kit. Of the 17 subjects (15 unrelated, 2 related), eight had known Mendelian disorders, eight had pediatric diseases and one had breast cancer (tumor/normal). We identified likely causative variant(s) in two unrelated subjects that included mutations in CASQ2, p.Gln67*, c.532+1G>A (Ventricular Tachycardia, Catecholaminergic Polymorphic, 2; OMIM#611938) and TP53, p. Arg213* (Breast Cancer; OMIM#114480), two related subjects (siblings) that included mutations in LDB3, p.V118M (Cardiomyopathy, Dilated, 1C; OMIM#601493) and possible causative variants in eight additional subjects. To further evaluate two breast cancer samples, two additional targeted cancer panels were used (AmpliSeq Cancer Hotspot and the Comprehensive Cancer Panel using the Ion Torrent Technology). In addition, we performed (WES) on the breast cancer sample and matched control tissue. We identified a pathogenic nonsense variant in the TP53 tumor suppressor gene. This variant was detected on all platforms. We found targeted capture high-throughput sequencing to be a cost effective, time sensitive and an efficient approach in detecting pathogenic variants to aid the diagnosis of individuals in our Bronx patient population.

2556T

Clinical whole exome sequencing (WES) coupled with whole mitochondrial genome sequencing at Baylor Whole Genome Laboratory (WGL). *Y. Ding[1], D.M. Muzny[1], J.G. Reid[1], A.C. Hawes[1], M. Wang[1], M.N. Bainbridge[1], N. Veeraraghavan[1], Y. Han[1], H. Dinh[1], D.P.C. Ng[2], C.J. Buhay[1], J.V. Korchina[1], M.L. Landsverk[2], V. Zhang[2], M. Scheel[1], W. Liu[1], N. Saada[1], J. Ma[1], J. Chandarana[2], L.K. Dolores-Freiberg[2], S. Matakis[1], R. Najjar[1], R.A. Gibbs[1], A.L. Beaudet[2], C.M. Eng[2], Y. Yang[2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX77030; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030.

Baylor WGL is a CLIA and CAP certified clinical laboratory which has provided clinical WES since October 2011. To date, we have received approximately 1700 cases including about 1600 germline WES cases and 100 cancer WES cases. Clinical reports have been issued from over 1000 cases, with a positive rate of 26%. Our data indicate the efficiency and cost effectiveness of WES for patients with complicated and unspecific clinical presentations suggestive of genetic etiology. However, while WES can replace nuclear gene sequencing as the first tier test, mitochondrial genome cannot be accurately sequenced from standard WES assays and has to be ordered separately if mitochondrial related disorders are indicated. In order to streamline test ordering process for referral physicians, we have been offering clinical WES coupled with mitochondrial genome sequencing (MGS) for every WES sample at no additional charge. Our clinical WES utilizes solution whole exome capture probe set VCRome 2.1, which covers approximately 200K coding exons while the entire mitochondrial genome is captured by one long range PCR amplification. Currently, manual Illumina Barcoded Paired-End (BC PE) capture library and mitochondrial whole genome shotgun library are sequenced on Illumina HiSeq2000 or 2500 platforms. Samples are multiplexed in loading (3 samples per HiSeq lane) and the corresponding mitochondrial barcoded libraries are spiked into the same lane at a ratio of 1:20 without additional sequencing cost. Quality control (QC) check points are set up for each major processing step, success rates of 99% and 96% for WES capture library and mitochondrial shotgun library respectively, and success rate of 99% for sequencing on HiSeq2000 or 2500 have been achieved. The median coverage for whole exome is 160X with approximately 13 Gb per WES sample, while median coverage for mitochondrial genome is 28,000X with 0.9 Gb per MGS sample. Greater than 95% of the targeted exome sequences and 100% mitochondrial sequences are covered at 20X or more. Currently, about 200 samples are processed at WGL for whole exome and whole mitochondrial genome analyses every month. Validation of Beckman Biomek NXp/Span-8 robots platforms has been initiated to accommodate the increasing sample volume with preliminary results demonstrating high quality data comparable to those from the manual processing.

2557F

BRCA1/2 Genomic Rearrangements and Characterization of Familial Breast Cancer.
Y.J. Hyun[1], S.H. Seo[1], M.W. Seong[1], S.I. Cho[1], S.S. Park[1], S.W. Kim[2]. 1) Seoul National University Hospital, Seoul, South Korea; 2) Department of Surgery, Seoul National University Bundang Hospital, Seongnam, Korea.

Introduction: Germline mutations in BRCA1 and BRCA2 are known to increase the risk for developing familial breast and ovarian cancer. The large genomic rearrangements (LGRs) involving the BRCA genes have been reported to account for variable proportion of familial breast cancer patients according to ethnic populations. Here, we studied the contribution of LGRs in BRCA1 and BRCA2 to high-risk breast cancer patients in Korea. Methods: A total of 243 patients with two or more familial history of breast cancer (BC) or one or more familial history of ovarian cancer (OC) were enrolled in the study. All patients were screened by direct sequencing (n= 177), or mutation scanning like fluorescence-based conformation-sensitive gel electrophoresis (F-CSGE) or denaturing high performance liquid chromatography (DHPLC) and/or direct sequencing (n=66). Multiplex ligation-dependent probe amplification (MLPA) was done in 144 patients who were mutation negative for the BRCA genes. Results: BRCA1 mutations were detected in 48 patients and BRCA2 mutations in 53 patients. One patient was harboring both BRCA1 and BRCA2 mutations. Among 144 BRCA mutation negative patients, BRCA1 LGRs were identified in three one and no BRCA2 LGR was identified. In addition, the BRCA1 LGRs account for 6.25% (3/48) of all BRCA1 mutations in this population. Conclusion: This study suggests that MLPA might be considered in BRCA1/2 mutation negative familial BC or OC patients in Korea, although contribution of LGRs is low in this population.

2558W

Utility and Limitations of Exome Sequencing for the Molecular Diagnosis of Bilateral Sensorineural Hearing Loss.
V. Jayaraman[1], J.S. Brunton[1], A. Sasson[3], M. Sarmady[3], J.L. Abrudan[1,2], M.C. Dulik[2], E.T. DeChene[1,2], S.E. Noon[1], A. Wilkens[1], A. Dickinson[1], M. Kaur[1], L.K. Conlin[2], N.B. Spinner[2], P.S. White[3], I.D. Krantz[1]. 1) Dept of Pediatrics, Children's Hospital of Philadelphia, PA; 2) Dept of Pathology & Laboratory Medicine, Children's Hospital of Philadelphia, PA; 3) Center for Biomedical Informatics, Children's Hospital of Philadelphia, PA.

Bilateral sensorineural hearing loss (BLSNHL) has significant genetic heterogeneity with more than 200 associated genes. Current testing for BLSNHL is either done by Sanger sequencing or next generation sequencing (NGS) of gene panels. These tests can be cumbersome and are not comprehensive. We report the results of applying exome sequencing for the diagnosis of BLSNHL. This work was carried out through the CHOP/UPenn Pediatric Genetic Sequencing (PediSeq) project, an NHGRI funded grant, to study the application of exome and genome sequencing to clinical care. Samples from 30 affected individuals with BLSNHL (10 positive controls and 20 unknown samples) were analyzed following exome capture (Agilent SureSelect v4), and sequencing (Illumina HiSeq 2000) at an average coverage depth of 100X. Reads were mapped to human genome GRCh37 using Novoalign; post processing and analysis was performed using a combination of Picard and GATK. A list of 267 BLSNHL genes was used as a filter for analysis. Our goals were to determine 1) the coverage for all 4,208 exons in the 267 target genes; 2) the number of BLSNHL variants not captured and 3) the number of known variants in our 10 controls that were correctly identified and the ability to identify causative mutations in the unknown cohort. The percentage of exons completely covered at 20X varied across individuals from 72.3% to 78.7% with the average being 75.3%. The average percentages of partially covered and not covered exons were 20.5% and 4.3% respectively. Overall, 65.3% of exons had consistent complete coverage in at least 90% of our sample population. Based on the number of uncovered exons, we calculate that 192 out of a total of 3,119 known variants (6.2%) reported in HGMD would be missed. Of the 10 positive controls, all 16 pathogenic mutations were covered. However, 3 of those mutations were partial exon deletions and therefore not identified through the variant calling pipeline. When assessing a targeted gene panel, such as the 267 BLSNHL panel studied here, dropping read depth criteria to 5X did not significantly alter the number of variants needing to be evaluated and increased the ability to identify causative mutations. Although WES may be effective for diagnosis of heterogenous disorders, there are significant limitations to the current platforms which greatly impact the significance of a ‘negative’ result and appropriate single gene or NGS panels should be considered for targeted diagnostic testing.

2559T

The Mayo Clinic experience with short stature and search for genetic causes of idiopathic cases.
A. Kochhar[1], J. Rustin[2], M. Goodenberger[2], Q. Stein[3], A. Mroch[3], L. Borovik[3], L. Davis-Keppen[3], E. Thorland[2], G. Velagaleti[4], S. Kirmani[1], A. Lteif[5], P.L. Crotwell[3], J.C. Hodge[2]. 1) Department of Medical Genetics, Mayo Clinic, Rochester, MN; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Sanford Clinic USD Genetics Laboratory, Sanford Children's Hospital, Sioux Falls, SD; 4) Pathology Department, University of Texas Health Sciences Center, San Antonio, TX; 5) Pediatric Endocrinology, Mayo Clinic, Rochester MN.

Short stature is a multifactorial trait that is influenced by numerous genes and environmental factors. The subcategory of idiopathic short stature (ISS) is defined as height two standard deviations below the mean for age and sex in the corresponding population, without evidence of an endocrine, nutritional or genetic syndrome cause. While the etiology of most ISS cases is unknown, ~2-5% of such patients have defects in the transcription factor-encoding gene *SHOX* or its enhancers, which is located in the Xp/Yp pseudo-autosomal region 1 (PAR1). Our study aims to describe the diagnostic spectrum of short stature observed at Mayo Clinic and to identify copy number changes in *SHOX* and its regulatory elements to further define the mutation heterogeneity at this locus as well as in 121 other genes potentially associated with stature in ISS patients. A custom high-density Agilent oligonucleotide 8x60k chromosomal microarray was developed that targets genes associated with idiopathic and syndromic short stature, genes suggestive of a potential role in stature, and the PAR1 area including *SHOX*. An IRB-approved retrospective effort to identify ISS cases involved chart review of demographics, clinical presentation, lab data and imaging results from patients referred to Mayo Clinic Rochester for short stature between 2009 and 2012. We classified 200 patients with short stature: 43 (21%) ISS, 45 (22%) constitutional delay of growth and puberty, 23 (11%) nutritional, 27 (13%) identified genetic syndrome, 19 (9%) syndromic short stature with no identifiable cause, 12 (6%) endocrinopathy, 12 (6%) chronic systemic disease, 5 (2%) intrauterine growth retardation and 14 (7%) short stature due to miscellaneous causes. The ISS patients were approached for consent to perform custom chromosomal microarray testing. A prospective effort also ascertained ISS patients from pediatric endocrinology and medical genetics at Mayo Clinic and Sanford Children's Hospital. Finally, cases referred for short stature that were normal by clinical testing using a chromosomal microarray not specifically targeting stature genes were also included. To date, these combined recruitment methods have resulted in analysis of 66 samples and the identification of two PAR1 deletions (1.3 Mb and 1.9 Mb) encompassing *SHOX* and two duplications (448 kb) that overlay an evolutionarily conserved region near *SHOX* of possible significance. This study is ongoing with a recruitment target of 200 patients with ISS.

2560F

Two *HEXB* genotypes affect Tay-Sachs disease carrier identification by enzymatic activity assay. *J. Liao, M. Luo, L. Shi, J. Goldman, L. Edelmann, C. Yu, R. Kornreich.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Tay-Sachs disease (TSD) is an autosomal recessive, neurodegenerative disorder caused by mutations in the *HEXA* gene, which lead to deficient activity of the Hex A enzyme and the accumulation of its substrate GM2 ganglioside in neural cells. Hex A is comprised of two subunits, α and β encoded by *HEXA* and *HEXB* respectively. A related isozyme, Hex B consists of two β subunits. To prevent TSD in high-risk populations, measurement of Hex A enzymatic activity has been routinely used as the primary method of carrier screening. This assay utilizes a synthetic substrate, 4-MUG, which can be hydrolyzed by both Hex A and Hex B, in combination with a heat-inactivation step, to differentiate the activities of the heat-labile Hex A from heat-stable Hex B in fluids or blood cells. TSD carriers, noncarriers, and affected individuals can be identified by their different ranges of Hex A%. Here we describe four TSD carriers identified by mutation analysis whose Hex A% results from enzymatic activity assay in leukocytes showed that one was a noncarrier (Hex A%: 68.4) and the other three were within the inconclusive range (avg. Hex A%: 53.0±2.7). *HEXB* gene sequencing revealed a c.[1627G>A; *81_*82delTG] genotype in the first case (noncarrier by Hex A%) and a c.[619A>G(+)*81_*82delTG] genotype in the other three (inconclusive by Hex A%). An additional 41 TSD carriers confirmed by both mutation analysis and Hex A% (avg. Hex A%: 43.1±4.7) were screened for these two genotypes. None of them had c.[1627G>A; *81_*82delTG] genotype and the allele frequency of c.[619A>G(+)*81_*82delTG] genotype in these carriers was 9.8% (8/82), no different than the general population. These results, therefore, suggest that these two genotypes are strongly associated with atypical Hex A% findings as all four mutation positive individuals who also had these polymorphisms did not have enzyme results in the carrier range. The 1627G>A and the *81_*82delTG polymorphisms have been reported to reduce the thermostability and enzymatic activity of Hex B, respectively. The c.[619A>G(+)*81_*82delTG] genotype has also been previously found in five TSD carriers with normal serum Hex A%. Therefore, these two genotypes may cause ambiguous TSD carrier identity by raising Hex A% as a result of decreasing Hex B thermostability and/or activity. They have potential to mask the true TSD carrier status in population screening and should be considered when performing TSD carrier screening by enzymatic activity assay.

2561W

Classification and interpretation of *PRSS1*, *SPINK1* and *CFTR* sequence variants found in idiopathic and hereditary pancreatitis. *A. Millson[1], C. Miller[2], E. Lyon[1,2,3].* 1) ARUP Inst Clin & Exp Path, ARUP Labs, Salt Lake City, UT; 2) ARUP Laboratories, Salt Lake City, UT; 3) Pathology Department, University of Utah, Salt Lake City, UT.

Hereditary pancreatitis is an autosomal dominant disease with variable expression caused by mutations in the *PRSS1* gene. Pathogenic mutations in two other genes, *SPINK1* and *CFTR*, may also contribute to idiopathic pancreatitis and are usually inherited in an autosomal recessive or multifactorial fashion. Thus, sequencing of the *CFTR*, *PRSS1* and *SPINK1* genes helps determine if there are genetic factors contributing to hereditary and idiopathic pancreatitis. Progeny, a clinical data management software program, is used at ARUP Laboratories to track patient symptoms and molecular sequencing results. This enables us to determine if there is a genotype/phenotype correlation, allows for consistent classification of identified variants and provides for quality assurance in laboratory reporting. Data entered in Progeny includes patient demographics, symptoms, risk factors, family history, and *CFTR*, *SPINK1* and *PRSS1* sequencing results (nucleotide and amino acid changes, mutation type, mutation effect and classification). The ARUP Progeny database was queried to find the pathogenic mutation detection rates in 913 individuals undergoing full gene sequencing of *PRSS1*, *SPINK1* and *CFTR*. Twenty-two percent of patients carried one *CFTR* mutation, 12.9% carried one *CFTR* and one *SPINK1* mutation, 4.9% carried one *SPINK1* mutation, 2.4% carried one *PRSS1* mutation and 2.5% carried two *CFTR* mutations. One case out of the 913 had two *CFTR* and one *SPINK1* mutation, and two cases had one *SPINK1* and one *PRSS1* mutation. Three cases carried two *SPINK1* mutations. No mutations were found in 64.2% of the cases.

2562T

Challenges of reporting incidental findings: follow-up on a FBN1 mutation identified by clinical WES testing. *Z. Niu[1], M.R. Bekheirnia[1], P. Ward[1], A. Braxton[1], F.J. Probst[1], G.S. Patel[2], L. Immken[2], Y. Yang[1], C.M. Eng[1].* 1) Molecular Human Genetics, Baylor College of Medicine, Houston, TX; 2) Specially for Children, Austin, Texas.

The recent advances in next generation sequencing and the success of whole exome sequencing have brought many changes to laboratory diagnosis. Beyond identifying causative mutations, the comprehensive nature of the genomic approach necessitates proper reporting of incidental findings, recommended by the '13 ACMG guideline. Here, we discuss our experience and follow-up on an *FBN1* mutation, as an example of the challenges associated with reporting incidental findings from clinical WES testing.

Marfan syndrome (MFS, MIM#154700) is a connective tissue disorder with variable clinical manifestations in the skeletal, ocular and cardiovascular systems. Aortic dilation, dissection and mitral valve prolapse are common in MFS patients. However, aortic root dilation may not be noted without aortic insufficiency or rupture, as one of leading causes of premature death in MFS patients. Clinical WES performed in the Medical Genetics Laboratories at Baylor College of Medicine identified a heterozygous c.3509G>A (p.R1170H) mutation in the *FBN1* gene in a 2 year old patient without known cardiovascular issues, who was diagnosed with mental retardation AD 5 (MIM:612621) based on clinical presentation and a *de novo* mutation in *SYNGAP1* (PMID:23161826). Considerable literature discussed segregation of the p.R1170H mutation with Marfanoid habitus, myopia and mitral valve prolapse, but aortic dilation was not present in most cases (PMID:7870075,9837823, '11ICHG-1132W). We report it as a medically actionable finding based on potential benefit, noting variable phenotypic expression. Further evaluation of the child revealed an enlarged aortic root (22 mm, z=3.99), hindfoot deformity and retrognathia. The parent of this patient, who also carries the p.R1170H mutation, is asymptomatic for MFS with the exception of malar hypoplasia. To our knowledge, this is the 1st patient with the p.R1170H mutation in *FBN1* exhibiting aortic root dilation at a young age. Based on above results, close monitoring is needed for cardiovascular complications and intervention.

In summary, the case reported here demonstrates the benefit and challenges of returning incidental findings with proper consent, thorough communication and deliberate follow-up evaluation. Collective information and additional studies could broaden the phenotypic spectrum of the p.R1170H mutation in the *FBN1* gene.

2563F

Clinical experience with "TaGSCAN," a targeted next-generation based sequencing test for 514 genes using symptom-driven analysis. *C.J. Saunders[1,2], E. Farrow[1], S.E. Soden[1,3], N.A. Miller[1], D. Dinwiddie[4], L. Willig[1,3], L. Zellmer[1,2], L. Smith[1,3], S.F. Kingsmore[1,2,3].* 1) Center for Pediatric Genomic Medicine; 2) Department of Pathology; 3) Department of Pediatrics, Children's Mercy Hosp, Kansas City, MO; 4) Department of Pediatrics, University Of New Mexico, Albuquerque, NM.

We designed and clinically validated a targeted next generation sequencing-based test for 514 genes, representing 795 childhood genetic diseases. "TaGSCAN" (for Targeted Gene Sequencing and Custom Analysis) uses Illumina hybrid enrichment with a custom probe set, 2 × 100 cycle sequencing (500X coverage), and a trio of novel software applications, developed in-house. The latter include Sign and Symptom Assisted Genome Analysis (SSAGA), Rapid Understanding of Nucleotide variant Effect Software (RUNES) and Variant Integration and Knowledge in Genomes (VIKING). These tools tailor variant results by customizing the list of genes analyzed for each patient based on phenotype. This limits incidental findings, carrier status, and the number of variants requiring interpretation. This low-cost assay may be used as a broad screen for unknown genetic conditions as well as a more specific test to sequence discrete genes or small panels of genes with high clinical suspicion. TaGSCAN includes a diverse range of genes for which no clinical testing is available in the United States, including IL12RB1, LRP2, IFNGR1, IFNGR2, PLEKHG5, and SNAP29, ANTXR2. Since clinical validation of this test, we have run a total of 121 samples, including both research and clinical patients with 34 (28%) of these returning a positive or likely positive result. Positive genes in this sample set include common genes like CFTR, ACADM, PAH, F8 and G6PD, as well as genes associated with very rare conditions, such as IGHMBP2, SLC16A2, IL12RB1, NR5A1, ERCC6, and PLA2G6. Of the 56 tests run clinically, 43 were ordered as single gene or small (<5 genes) panels, for which the diagnostic yield was 35%. 13 clinical tests were run as symptom-driven screens, for which 3 (23%) were positive. A prospective cost analysis performed on the first 51 patients comparing charges for sending the test out versus billing for TaGSCAN indicates an estimated savings of over $185,000.

2564W

Establishment of Molecular Diagnostic Platform for Leber Congenital Amaurosis using Extensive Multi-Gene Panel Sequencing. *S.H. Seo[1], Y.S. Yu[2], JM. Hwang[3], H. Park[1], S.I. Cho[1], S.S. Park[1], MW. Seong[1].* 1) Departments of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 2) Departments of Ophthalmology, Seoul National University Hospital, Seoul, Korea; 3) Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Korea.

Introduction: Leber congenital amaurosis (LCA) leading to blindness is genetically heterogeneous, and thus confirming the molecular diagnosis can be challenging. So far, more than 10 genes have been discovered to be associated with this disease, but these genes account for a small proportion of LCA in this country. Herein, we developed an extensive diagnostic multi-gene panel including more than 200 retinal dystrophy-associated genes. **Methods:** Twenty LCA patients were participated in the study, including 2 patients with previously defined pathogenic mutations, in order to verify the performance of our diagnostic platform. A total of 203 genes were analyzed by enrichment of all coding regions and flanking intronic regions with a customized NimbleGen SeqCap EZ Choice library (Roche) and sequencing with a HiSeq (Illumina). All pathogenic variants discovered and low-coverage regions under 10X of coverage depth were resequenced by Sanger method. **Results:** On average, the coveage depth was 800X per base pair (bp), and only less than 1% of the regions showed coverage under 10X. Two patients for assay validation were successfully confirmed using multi-gene panel sequencing. Among remaining 18 patients, two pathogenic mutations were detected in 5 patients including two novel LCA genes as well as known LCA genes like CEP290 and RPGRIP1. In additional 8 patients, only single possibly pathogenic mutations were found. **Conclusion:** Our extensive diagnostic platform has identified novel genetic causes in LCA patients. We demonstrated the successful use of massive parallel sequencing in confirming the molecular diagnosis of LCA patients, and that multi-gene panels can be used with more advantages over exome sequencing in the clinical fields.

2565T

Detection of low-level mosaic microdeletion in Neurofibromatosis type 1. *J. Xie, A. Poplawski, C. Fu, T. Callens, J. Williams, H. Zhan, L. Messiaen.* Dept of Genetics, Medical Genomics Laboratory, University of Alabama at Birmingham, Birmingham, AL.

Neurofibromatosis type 1 (NF1; MIM#162200), an autosomal dominant neurocutaneous disorder affecting ~1 in 3000 individuals worldwide, is caused by mutations in *NF1* gene. The *NF1* gene has a high mutational rate compared to other disease-related genomic loci. As many as 30-50% of NF1 patients present as sporadic cases, i.e. do not have a parent affected by the disorder. A fraction of these 'founder' cases presents with mosaicism for a *NF1* mutation, as a consequence of *de novo* mutations arising postzygotically during embryonic development. Mosaicism is an important consideration in NF1, given its impact on the clinical phenotype and transmission risks. Estimation of the frequency of mosaicism in sporadic NF1 is however challenging due to the notorious variability of the phenotype, the complexity of the *NF1* mutational spectrum and the limitations of the current techniques to accurately detect low-level mosaicism for all of the potential mutational targets in the *NF1* gene. A first estimation on the frequency of *NF1* mosaicism was provided through the study of patients carrying an *NF1* microdeletion as identified using Multiplex Ligation-dependent Probe Amplification (MLPA) and confirmed using FISH (Messiaen et al. Hum Mutat. 32, 213, 2011). The frequency of 10% in this cohort necessarily is an underestimate, as low-level mosaic microdeletions would have been missed in the initial assessment. We recently identified a mosaic NF1 patient carrying an *NF1* microdeletion in only 8% blood cells (4% of the alleles). Such low-level mosaicism will escape detection using MLPA or aCGH. Therefore, in order to facilitate more sensitive and accurate molecular diagnosis for *NF1* microdeletions, we developed a fast, sensitive and cost-effective method utilizing droplet digital PCR (ddPCR). ddPCR allows measurements of smaller fold changes compared to MLPA or aCGH. TaqMan® copy number assay probes were used for the target *NF1* and reference *RNase P* genes and absolute measurements were obtained using the droplet reader. Our data showed that ddPCR can reliably detect mosaic *NF1* microdeletion in as low as 5-10% blood cells ($p<0.05$). As upon transmission, the *NF1* microdeletion tends to result in a more severe phenotype with earlier age of onset in the constitutionally affected patient. Sensitive detection of this specific type of mutation in 'founder' patients is warranted as a part of a comprehensive assessment.

2566F

Incorporating new disease genes into clinical whole exome sequencing (WES): annotation update, interpretation challenges and customized familial studies. *Y. Yang[1], F. Xia[1], J. Beuten[1], Z. Niu[1], M.S. Leduc[1], R.E. Person[1], M.T. Hardison[1], J. Zhang[1], M. Bainbridge[2], J.G. Reid[2], A.C. Hawes[2], Y. Ding[2], A.A. Braxton[1], P.A. Ward[1], M.L. Landsverk[1], A. Willis[1], D.M. Muzny[2], S.E. Plon[1,3], J.R. Lupski[1,2], A.L. Beaudet[1], R.A. Gibbs[2], C.M. Eng[1], Baylor Whole Genome Laboratory (WGL).* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX.

The pace of disease gene discovery is accelerating through utilization of NGS technologies, availability of large-scale biology data and accumulated knowledge of gene function. Hundreds of new disease genes are identified each year. Translating new research discoveries into clinical laboratory practice in an accurate and timely manner can lead to improved diagnostic sensitivity and better understanding of genotype-phenotype correlations. For example, of the 265 positive cases diagnosed by clinical WES at Baylor WGL during the period of January 2012 through May 2013, 13% harbor causative mutations in disease genes discovered since January 2012. WES testing is an ideal platform for timely incorporation of new disease genes since it interrogates the entire coding regions of the genome, making it possible for updating disease gene annotation at any time during or even after the analysis process. However, extensive literature review to verify new Mendelian disorders and disease-causing mechanisms must be performed. The clinical laboratory should also make every effort to facilitate familial mutation studies for new disease genes, which usually do not yet have single-gene clinical tests available.

Clinical WES offered by Baylor WGL has a turnaround time of 15 weeks. During this period the disease gene list is updated twice, one time prior to the initial data review and the second time during laboratory director review, to ensure the most inclusive consideration possible for candidate disease genes. We also re-analyze cases after completing WES reports as new disease genes are published. In addition, we re-investigate the strength of evidence for new genes by reviewing published data on inheritance, functional studies and disease causing mechanism. This close follow-up of clinical cases leads to additional definitive and possible diagnoses. For example, we reported a patient with mutations in *MEGF8* (gene published November 2012, WES diagnosis reported the same month), another patient with a mutation in *HDAC8* (case signed out July 2012 as negative, gene published September 2012, diagnosis updated the same month), and recurrent diagnoses from disease genes (*MAGEL2,ASXL3*) discovered by our laboratory. Follow up communications with referring physicians are conducted and customized targeted mutation analyses, including prenatal testing when applicable, are made available on a clinical basis for assessment of family members.

2567W

Implementation of a Quality Assurance Program for Next Generation Sequencing Based Tests. *Z. Yu[1], M. Sarmady[2], T. Tischler[1], B.A. Brown-Kipphut[1], M. Shafiq[1], N.N. Fernandes[1], Z. Fan[1], M.S. Boltz[1], J.B. McKnight[1], B.N. McLarney[1], L.C. Davidson[2], M.J. Italia[2], J.W. Pennington[2], P.S. White[2], P.V. Warren[1], C.A. Stolle[1], A. Santani[1].* 1) Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Center for Biomedical Informatics, Children's Hospital of Philadelphia, Philadelphia, PA.

Clinical laboratories are governed by regulations from Clinical Laboratory Improvement Amendments (CLIA) and select laboratories also follow guidelines from the College of American Pathologists (CAP). These regulations are in place to ensure high standards and reliable test results in the clinical environment. Next generation sequencing (NGS) has revolutionized genomic medicine and is now being increasingly applied to clinical molecular diagnosis. Despite this increasing popularity, implementation of NGS in a clinical environment is challenging due to the inherent complexities of the rapidly evolving NGS technologies and the accompanying requirements for sophisticated data analysis. Moreover until recently, few guidelines or resources were available for institutions interested in implementing a clinical NGS program. Using the NGS accreditation criteria from the CAP checklist as a guideline, we developed a framework to address four major areas of a NGS based quality management program: i) validation testing, ii) quality assurance, iii) proficiency testing, iv) bioinformatics. We share our experience in this regard with the development and implementation of four targeted NGS based panels that were created to identify germline mutations in four common disorders. Validation testing was performed on 33 specimens to establish the analytical performance characteristics of these four laboratory developed tests. Quality control procedures were established to monitor all components of these tests including reagents, instrumentation and bioinformatics. Proficiency testing was performed to assess laboratory performance. A workflow management quality control system (Pegasus) was used to support a customized data management, processing, and analysis workflow. In summary, we have developed a comprehensive quality management program in the laboratory for NGS based tests. This program, which has been implemented in our clinical practice, will help meet regulatory standards as well as assure the integrity and quality of NGS results.

2568T

Non-invasive detection of genomic mutations by targeted sequencing of plasma cell free DNA. *J. Namkung, K. Seo, Y. Lee.* Bioinformatics Tech Lab, SK Telecom, Sungnam, South Korea.

Cell free DNA (cfDNA) offers a non-invasive diagnostic approach to a wide range of clinical disorders. Next-generation sequencing (NGS) has enabled the detection of unknown fetal genomic variations from cfDNA of maternal peripheral blood. However, due to the lengthy sequencing time, some platforms are still not suitable for clinical application. The relatively low concentration of cfDNA in the blood has presented many challenges as well. By using targeted sequencing, only a subset of genes or a defined region in a genome are sequenced, which allows high sequence coverage critical for identifying rare genetic variations. The purpose of our study was twofold, first, to sequence and analyze the genomic coverage of cfDNA using NGS technologies and second, to test the sensitivity and accuracy of detecting the genomic mutation from cfDNA by using Ion AmpliSeq™ Inherited Disease Panel (IDP), which enables multiplexed target selection of exons of 328 genes implicated in genetic disorders. Whole blood (6 mL) was collected from healthy donors, and plasma was separated by centrifugation. The cfDNA and genomic DNA (gDNA) were extracted using column-based methods. The concentration of cfDNA was 8.9 ± 0.5 ng/µl (30 ng/ml of blood), which was in agreement with reported values in healthy controls (5-30 ng/ml). Libraries were prepared accordingly: 1) Ion Xpress ™ Plus gDNA Fragment library kit was used for sequencing with the Ion Torrent PGM™ platform with 316 chips; 2) ThruPLEX™-FD Prep Kit (Rubicon Genomics) was used for Illumina® NGS platform. Result showed that the genomic coverage of cfDNA sequencing was 97% of that of the genomic DNA, given the same number of reads. The results were similar in both Illumina® and Ion Torrent PGM™ platforms. The study also showed that the cfDNA present in the blood stream is representative of the global content of the gDNA. To study the feasibility of detecting the genomic variation of cfDNA, semiconductor sequencing will be performed by using IDP. To mimic the condition of maternal blood, cfDNA samples will be mixed in the range of 5 to 20%, which is the ratio of cell free fetal DNA in the maternal blood. Through this study, we will determine the possibility of using cfDNA from blood for detection of genomic mutations. This suggests that the use of cfDNA in conjunction with accurate and rapid semiconductor sequencing technologies could be the next step forward in developing non-invasive diagnostic testing.

2569F

A systematic approach to assessing the clinical significance of genetic variants. *H. Duzkale[1,2], J. Shen[1,2], H. McLaughlin[1,2], A. Alfares[1,2], M.A Kelly[2], T.J Pugh[1,2], B.H Funke[1,2], H.L Rehm[1,2], M.S Lebo[1,2].* 1) Harvard Medical School, Boston, MA; 2) Partners HealthCare Center for Personalized Genetic Medicine, Cambridge, MA.

Molecular genetic testing can improve the accuracy of diagnosis, prognosis, and risk assessment for patients and their family members. Recent advances in low-cost, high-throughput DNA sequencing technologies have enabled the rapid expansion of genetic tests, as well as an expansion of variant- and gene-level data and their associations with human disease. The number of variants assessed in our laboratory has tripled in the past two years through launching next generation sequencing (NGS) gene panels. While NGS panels have increased the analytical and clinical sensitivity of our assays, the increased content has also added a significant burden of interpretation and higher rate of variants of uncertain significance (VUS). Our laboratory has developed a systematic approach to accurate, efficient and timely assessment of variants for pathogenicity over the past decade. Using our semi-automated tool, the average time for a clinical variant assessment has been reduced to 22 min. We first gather and validate variant information from publications and clinical notes, from internal, collaborative and public variant databases, and from various bioinformatics resources; we then perform statistical analyses, evaluation of research and functional data, and computational predictions on the data to determine the likelihood of pathogenicity; finally, we weigh all the evidence to reach an overall interpretation on the potential for each variant to be disease-causing. In this report, we highlight the principles of variant assessment, address the caveats and pitfalls, and provide customizable tools and examples to illustrate the process. We also demonstrate the importance of variant reassessment, as we recently downgraded 11 of 106 (10%) pathogenic or likely pathogenic cardiomyopathy variants to VUS, mostly due to their presence in large population studies. Our laboratory has evaluated 245 genes in 53 diseases in a total of 22,490 cases since its inception. 17,245 variants have approved clinical classifications and 7,955 were validated and reported in patients. We have recently submitted 7,129 variants to ClinVar database, an initiative to publicly share genetic test results as well as phenotype information of patients. Of these variants, 40% have associated rsIDs in dbSNP, 20% have been reported in the literature, and 40% were novel. By sharing our experience, we hope to bring the clinical and research communities together to build a framework for variant assessment.

2570W

Assessment Of Detection Of Proviral DNA and RT (MET 184 VAL) Gene Resistance Mutation in HIV-1 Identified by Multiplex PCR and Restriction Fragment Digestion Assay. *R. SHRESTHA[1,2], S. KHADKA[1], S.R. WAGLE[1], A. SAPKOTA[1].* 1) Center for Molecular Dynamics Nepal (CMDN), Intrepid Nepal, Pvt. Ltd. Thapathali-11, Kathmandu, Nepal; 2) Department of Laboratory Medicine, Nobel Hospital and research Institute, Kathmandu, Nepal.

ABSTRACT Proviral DNA and ART resistance forms the corner stone of a short chemotherapy course for Post exposure prophylaxis. Potent antiretroviral therapies that suppress cell-free plasma viral RNA levels below the limit of current assay detection necessitate other complementary approaches for assessing viral burden, such as quantification of cell associated proviral DNA.The quantification of human immunodeficiency virus type-1 (HIV 1) RNA is the cornerstone for monitoring the effectiveness of antiretroviral therapy. For the majority of therapy naive patients, three to six months of potent antiretroviral therapy usually suppresses cell-free HIV 1 RNA levels to less than 50 copies/mL of plasma. Objective: To rapidly identify the proviral DNA and To study mutation in specific part (meth184val; ATG - GTA substitution) of M184V gene of HAART resistance in HIV-1 Reverse Transcriptase. Method: An analytical study was design to explore the role of proviral DNA and RT (M184V) gene mutation identified by PCR and RFDA. Results and Discussion: The male patients outnumbered female. Among 13 male subject 2(15.4%) were proviral DNA positive and 11(84.6%) were proviral negative. Among 2 female subject 1(50.0%) were proviral DNA positive and 1(50.0%) were proviral negative. All isolates were only digested by CViAII restriction enzyme and confirmed as wild type. No mutation was detected in the analyzed sample. Conclusion: In summary, we demonstrated here that in contrast to Proviral detection drug response was good as proviral DNA load was below the limit of detection in patients. Since, we can say that the incidence of developing of resistance virus is in increasing order. In addition, Potent antiretroviral therapies suppress cell-free plasma viral RNA levels below the limit of current assay detection necessitate other complementary approaches for assessing viral burden, such as quantification of cell associated proviral DNA. Therefore, it is recommended to use this technique for the diagnosis of Drug Resistance Mutation in HIV-1 patients and the response of ART for their Better Health Care. Keywords: HIV/AIDS, HAART, Drug resistance, mutation, env gene, RT gene,.

2571T

International external quality assessment for diagnostic Next Generation Sequencing. *S. Abbs[1], J. Coxhead[2], P. Westwood[3], K. Thomson[4], H. Scheffer[5], S. Bhaskar[6], G. Taylor[7], Z. Deans[8], S. Patton[9].* 1) Genetics Laboratories, Cambridge University Hospitals, Cambridge, United Kingdom; 2) Newcastle University, Newcastle, UK; 3) Western General Hospital, Edinburgh, UK; 4) Oxford University Hospital, Oxford, UK; 5) University Medical Center Nijmegen, Nijmegen, the Netherlands; 6) Central Manchester Foundation Trust, Manchester, UK; 7) University of Melbourne, Victoria, Australia; 8) UK National External Quality Assessment Scheme for Molecular Genetics, Edinburgh, UK; 9) European Molecular Genetics Quality Network, Manchester, UK.

Next Generation Sequencing (NGS) is increasingly being introduced into clinical genetics laboratories worldwide. The huge amount of data generated by NGS cannot be duplicated by alternative methods for laboratories to internally validate all results, therefore external assessment of data is required. The UK National External Quality Assessment Scheme (UKNEQAS) for Molecular Genetics and the European Molecular Genetics Quality Network (EMQN) have developed a joint EQA scheme for NGS, with the aims to: (a) assess and improve quality; (b) enable laboratories to benchmark their NGS service against others and against best practice; (c) work towards consistency of reporting clinical results generated by NGS; and (d) contribute towards best practice. EMQN and UKNEQAS offer numerous disease specific EQA schemes, and the challenge for developing NGS EQA was to ensure it does not duplicate what is already available, making it generic (independent of genes, diseases, and platforms) and applicable to as many users as possible. A survey of 1020 labs worldwide in Dec 2011 generated replies from 52% about methods, platforms, and loci being tested. Diagnostic experience of NGS varied from less than a year (76% labs) to over 3 years (1.9%). All respondents expressed an interest and need for NGS EQA. The survey results were used to develop a follow-up survey to a subset of labs, and then establish a pilot EQA. 30 labs were selected on the basis of diagnostic experience with NGS and covering a variety of technology platforms. These labs were sent a genomic DNA sample and asked to sequence either the smallest gene panel or largest single gene which the lab tested, submit technical details, and genotypes at known SNPs. The DNA was validated in 3 diagnostic labs and by 3 NGS platform manufacturers. The number of genes sequenced per lab varied from 1-625. 1011 variants were reported against reference sequences, in a total of 145 genes. 30 genes were sequenced by more than one lab, allowing comparison of results between participating labs and the validation data. The minimum acceptable depth of coverage required by labs for reporting results varied from 16x to 100x. Disease specific EQA has drastically improved the quality of results and consistency in diagnostic reports. This NGS EQA will play an important role in enabling labs to benchmark this new technology, assess the accuracy of data and facilitate high quality reporting for patient benefit.

2572F

Development of a rapid and comprehensive genetic testing service for nephrotic syndrome using next generation sequencing. *E.J. Ashton[1], D. Bockenhauer[2], N.J. Lench[1].* 1) NE Thames Regional Genetics Service, Great Ormond Street Hospital, London,WC1N 3BH United Kingdom; 2) Department of Nephrology, Great Ormond Street Hospital for Children, London, WC1N 3JH.

Nephrotic syndrome is a renal disorder presenting with proteinuria, oedema and hyperalbuminemia. Mutations in at least 18 different genes are known to be responsible for nephrotic syndrome and genetic testing has been a long and time consuming process as each gene would be sequenced independently until a mutation could be identified. Identification of a genetic basis provides an explanation for the disease, enables genetic and prognostic counselling and increasingly affects management of the disease. The advent of next generation sequencing has meant that it is possible to screen a large number of genes simultaneously, which would previously have been impractical and unaffordable. We have developed a targeted re-sequencing screen comparing two different library preparation methods (the Haloplex™ target enrichment system from Agilent and TruSeq Custom Amplicon from Illumina) and used these to screen for mutations in up to 18 genes known to be involved in nephrotic syndrome. Sequencing was carried out using the Illumina MiSeq system using both 2x150 bp and 2x250 bp sequencing chemistries and data analysis was performed using NextGENe® software from SoftGenetics. All mutations detected by Next Generation Sequencing were confirmed using Sanger sequencing. We have screened over 60 patients so far who were referred with a clinical diagnosis of steroid-resistant nephrotic syndrome aged from new-born to adulthood, using a combination of the two different library preparation methods, with the majority of mutations so far being identified in the NPHS1 gene. In cases where we were unable to identify a definite pathogenic mutation (or mutations) we considered that this may be due to mutations in regions of a gene not covered by our design, an uncertain clinical diagnosis or mutations in genes yet to be identified as being involved in nephrotic syndrome and therefore not included in our designs. The flexibility of the library designs means that it is simple to add in genes newly associated with nephrotic syndrome. We have developed a rapid and comprehensive genetic testing service for patients and families with nephrotic syndrome. This well lead to improved detection rates combined with a much reduced cost and turn-around time. The low rate of identified mutations may indicate that the majority of cases of steroid-resistant nephrotic syndrome are not due to monogenic mutations in currently known disease genes.

2573W

Diagnostic exome sequencing can alter a primary clinical diagnosis. *F. Taylan[1,2], M. Kvarnung[1], A. Lindstrand[1], T. Bui[1], A. Nordgren[1], E. Blennow[1], M. Nordenskjöld[1], D. Nilsson[1,2].* 1) Department of Molecular Medicine and Surgery and Center of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Science For Life Laboratory (SciLifeLab), Solna, Sweden.

Whole exome sequencing (WES), which is now feasible and clinically available, allows us to identify disease causing variants in rare diseases. Advancements in bioinformatic tools also make the data produced by WES more interpretable and easy to handle. As part of our effort to use WES as a diagnostic tool, we have collected samples from distinct families with consanguineous marriages and at least two affected children with complex phenotypes. Whole exome sequencing was performed for all available family members, such as parents, affected and healthy children. The family case presented here has two affected children with a clinical diagnosis of Bardet Biedl Syndrome (BBS; MIM209900), a multisystemic ciliopathy characterized by retinal degeneration, obesity, intellectual disability and renal dysfunction. The major findings observed in the affected children were retinal dystrophy, obesity, hypogonadism, eczema, small teeth and flat feet. Whole exome sequencing revealed no mutation in any of the seventeen known BBS genes and genome wide array comparative genomic hybridization did not detect any deletion or duplication at the chromosomal level. We then expanded our analysis to other ciliopathy genes associated with phenotypes similar to BBS. A novel homozygous stopgain mutation was detected at position 2437 of the ALMS1 (ALMS1:NM_015120:exon8:c.C7310A:p.S2437X) in both of the affected children, and the mother, the only parent available, was found to be a carrier of this mutation. The ALMS1 has been known to cause Alström Syndrome, an autosomal recessive ciliopathy with a clinical picture partly overlapping with BBS. This finding demonstrates how exome sequencing can alter a clinical diagnosis and reveal differential diagnoses of phenotypically similar syndromes. As the sequencing technologies become cheaper and bioinformatic tools improve further, whole exome sequencing for clinical diagnostic purposes finds its place in routine medical practices.

2574T

The importance of considering autosomal genes for the diagnosis of Non-Syndromic Intellectual Disability. *C. Tan[1], S. Topper[1], V. Nelakuditi[1], K. Arndt[1], F. Kobiernicki[1], D. del Gaudio[1], N. Meeks[2], J. Saari[2], V. Misra[3], S. Sastry[3], S. Levesque[4], L. Russell[4], G. Sillon[4], S. Das[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Pediatrics, Section of Genetics, University of Colorado, Aurora, CO; 3) Division of Genetics and Metabolic Disorders, Children's Hospital of Michigan, Detroit, MI; 4) Department of Medical Genetics, Montreal General Hospital, Montreal, QC.

Intellectual Disability (ID), with a prevalence estimated to be between 1-3% of the general population, is a lifelong disability typically presenting in infancy or early childhood. It is estimated that mutations in the genes on the X chromosome may account for about 10% of all cases of ID, thus molecular investigations into the X chromosome in elucidating the etiology of an individual with ID have been routinely suggested. Due to the genetic and phenotypic heterogeneity of patients with non-syndromic ID, multi-gene testing, the concurrent analysis of multiple genetic loci, can be beneficial. The utilization of high throughput, massively parallel sequencing, or next generation sequencing (NGS), has greatly improved the ability to simultaneously analyze multiple genetic loci, and X-linked ID (XLID) panels including analysis of a large panel of X-linked genes can be used for diagnosis. We performed next generation sequencing analysis of a total of 95 genes implicated in non-syndromic ID. The 95 genes included 61 X-linked genes, 20 Autosomal Recessive genes and 14 Autosomal Dominant genes. Genes were selected based on a criterion of 'non-syndromic ID' although some syndromic genes were also included, as a range of mutations in a single gene can sometimes confer both syndromic and non-syndromic phenotypes. As more genes are identified and implicated in non-syndromic ID, it seems appropriate to consider the inclusion of autosomal genes in multi-gene panel testing. To date, we have studied 10 patients and have identified pathogenic mutations in 3 out of the 10 patients. All three mutations are novel and include two frameshift mutations in *SYNGAP1* (c.1783del and c.2602del) and one splicing mutation in *TCF4* (c.991-2A>G). All genes in which mutations were identified are autosomal, typically de-novo in inheritance. Phenotypic information was collected on all three patients. Both patients identified with the *SYNGAP1* mutations were non-dysmorphic and were characterized as having global development delays. The patient identified with the *TCF4* mutation had global developmental delays and minor facial dysmorphisms that were possibly suggestive of a spectrum of monogenic conditions. Our three mutation positive cases provide further evidence supporting the utility of including autosomal genes in multi-gene panel analysis of patients with intellectual disability.

2575F

Clinical usefulness of copy number variants detected by affymetrix high-resolution genome-wide array. *E. Cho[1], E. Lee[1], J. Jang[1], H. Kim[2].* 1) Green Cross Laboratories, Yongin , South Korea; 2) Gachon University, Gil Medical center, Incheon, South Korea.

Background: Recently, chromosomal microarray has proven to be an effective tool for detection of submicroscopic chromosome abnormalities causing congenital disorders and has been adopted for clinical applications. Here, we investigated the usefulness of chromosomal microarray, as a first-tier tool in detecting the etiology of developmental delay, intellectual disability, autism spectrum disorders and multiple congenital anomalies in a large number of Korean pediatric patients. Methods: We applied Affymetrix Cytogenetics Whole Genome 2.7M array and Affymetrix Cytoscan 750K array in 333 patients (2.7M array: 200 cases and 750K: 133cases) to assess CNV detection and evaluated the clinical significance of detected CNVs. Results: We found 103 cases (30.9%) with known pathologic CNVs (64 cases: 19.2%) and CNVs of uncertain clinical significance (39 cases: 11.7%). 3 cases(1.0%) of uniparental disomy were detected. 70% (45/64) of known pathologic CNVs were <5Mb and would likely not be detected by G-banded chromosome analysis. 50 cases with copy number losses, 45 cases with copy number gains, 6 cases with both copy number gain and loss and 2 cases with mosaic copy number gain (12p mosaic gain) were detected. The smallest pathologic CNV detected was 23Kb. 2.3 Mb gain at 8p23.2 found in 7 cases was revealed as a probably benign CNV by parental study. Conclusions: We concluded microarray analysis significantly improves the diagnostic yield than G-banded chromosome analysis.

2576W

Performance analysis of saliva generated genomic DNA used for genotyping on the Affymetrix DMET Plus array as part of the Coriell Personalized Medicine Collaborative. *N.P. Gerry, N.C. Weiner, D.E. Lynch, L.A. Swanson.* Coriell Institute for Medical Research, Camden, NJ.

The use of saliva as a source of genomic DNA for research and clinical studies has grown in popularity due to the ease of collection and participant compliance. The Coriell Personalized Medicine Collaborative (CPMC) has been using Oragene collection kits for the past six years as the source of genomic DNA, initially for genotyping on the Affymetrix Genome-Wide Human SNP Nsp/Sty 6.0 array, and for the past four years, also for genotyping on the Affymetrix DMET Plus array. Although multiple studies on the SNP 6.0 array have utilized saliva as the source of DNA, the CPMC is one of the first large scale studies to also use that DNA on the DMET Plus array. To date, the study has successfully processed more than 4000 samples on the DMET Plus array with an average call rate of 99.5%. This was similar to the 99.4% call rate achieved on the SNP 6.0 array. The genotyping results from saliva generated genomic DNA have also proven to be highly reproducible on DMET Plus. Independent extractions from individual Oragene kits as well as across multiple kits for 4 samples used as processing controls have resulted in average call rates between 99.5% and 99.8% across thirty or more replicates of each control. Furthermore, in the control replicates, the concordance rates in a set of 166 variants of interest to the CPMC study ranged between 99.7% and 99.9 %. Finally, because the genomic DNA samples were run on both Affymetrix arrays, it was also possible to examine the performance of 212 SNPs that are present on both platforms. For a set of 1920 samples, the average concordance was 98.2%. When excluding SNPs where either or both arrays had a No Call, the average concordance increased to 99.5 %. The use of saliva generated genomic DNA in the Affymetrix DMET Plus assay has proven to be very successful and has allowed the CPMC to expand its genotyping options while maintaining a single DNA source.

2577T

Improved diagnosis of mitochondrial disorders by next generation sequencing approach. *V.W. Zhang, J. Wang, Y. Feng, X. Tian, L-J. Wong.* Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Background: Step-wise analyses of the genetically and clinical heterogeneous mitochondrial disorders are time consuming and cost ineffective. Next generation sequencing (NGS) technology allows simultaneous sequence analysis of any number of target genes. **Method:** The intact circular mitochondrial genome is enriched by long range PCR as a single amplicon and sequenced with captured target nuclear genes on HiSeq2000. **Results:** At an average coverage of 20,000X and 1,000X for the mtDNA and nuclear genes respectively, heteroplasmic mtDNA point mutations and large deletions can be detected, quantified, with breakpoints determined. Poorly covered coding regions of nuclear target genes are filled in by specific PCR/Sanger sequencing to ensure 100% coverage. Our results demonstrated that NGS-based analysis improved the diagnostic yield of mtDNA disorders to 30% using muscle specimen. The detection rate in nuclear genes varies from 10 to about 25%. About 25% of complex deficient samples had mutations in mtDNA, and 16/19 (84%) samples with mtDNA depletion in liver had mutations identified in one of the genes causing mtDNA depletion syndrome. **Conclusion:** The NGS approach greatly improves the diagnosis of mitochondrial disorders. Due to the extreme clinical heterogeneity, it remains challenging to diagnose atypical cases despite the technical advance.

2578F

Detection of Disease-Causing Mutations in the Ashkenazi Jewish Population Using a BioFilm Microarray. *M. Procter[1], C. Smith[2], R. Mao[3].* 1) Research and Development, ARUP Laboratories, SLC, UT; 2) AutoGenomics, Vista, CA; 3) Department of Pathology, University of Utah School of Medicine. SLC, UT.

Background: Jews of eastern European descent, referred to as Ashkenazi Jews, are at a higher risk of carrying certain genetic diseases at greater frequencies than the general population. The American College of Obstetricians and Gynecologists (ACOG) recommends genetic testing for people of Ashkenazi Jewish descent for mutations associated with the following conditions: Tay-Sachs disease, Canavan disease, and Familial Dysautonomia. Further ACOG recommendations include testing be offered to these same individuals for mutations associated with the following diseases: Gaucher disease, Bloom syndrome, Fanconi Anemia group C, Neimann-Pick disease type A, and Mucolipidosis type IV. These autosomal recessive diseases occur 20-100 times more frequently in the Ashkenazi Jewish population, and are associated with life-threatening conditions. DNA-based carrier screens for the above conditions in this ethnic group are sensitive due to a number of common shared mutations. Materials and Methods: The assay involves multiplex PCR of DNA at 30 ng/μL, followed by analyte specific primer extension (ASPE) prior to hybridization of the ASPE primers to a BioFilm microarray. The microarray chip with bound ASPE product is washed repeatedly before optical scanning with signal detection and analysis. Steps subsequent to PCR were performed in the Infiniti™ bioanalyzer. For our evaluation, 31 samples representing all 30 mutations and 1 polymorphism were run in the assay to determine accuracy. We utilized DNA purchased from Coriell Cell Repositories and known-positive and known-negative samples from ARUP's clinical Molecular Genetics laboratory as well as samples acquired from collaborators. Results: Twenty-six mutations were detected with 100% concordance with expected results using clinical samples, Coriell samples, and samples from collaborators. The remaining 5 variants were detected with 100% concordance using a commercially available 'supercontrol' containing all AJ variants included in this assay. Conclusions: We found this assay to be reliable in the detection of a large number of mutations associated with Ashkenazi Jewish diseases. The microarray format of this test allows simultaneous detection of many mutations in a single reaction. Automation of all steps subsequent to the PCR step in the Infiniti™ instrument allows for high throughput with minimal hands-on time.

2579W

Comprehensive mutation analysis by next generation sequencing in patients with neonatal intrahepatic cholestasis. *T. Togawa, T. Sugiura, K. Ito, T. Endo, S. Saitoh.* Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

Next generation sequencing (NGS) technology has revolutionized genomic and genetic research. Neonatal intrahepatic cholestasis is a heterogenetic disorder and caused by mutations in a number of genes, making genetic diagnosis challenging. We developed a diagnostic panel of causative genes for neonatal intrahepatic cholestasis using AmpliSeq (LifeTechnologies) and the Ion Torrent PGM technology. DNA samples were obtained from 12 individuals with neonatal intrahepatic cholestasis; 4 patients with Alagille syndrome (AS), 2 patients with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD), 4 patients with progressive familial intrahepatic cholestasis (PFIC), and 2 patients with unknown etiologies. One patient with AS, 2 patients with NICCD and 2 patients with PFIC had known mutations detected by conventional Sanger sequencing, and were invited as positive controls. A panel of genes included *JAG1*, *NOTCH2*, *SLC25A13*, *ATP8B1*, *ABCB11*, *ABCB4*, *AKR1D1*, *HSD3B7*, *CYP7B1*, *TJP2*, *BAAT*, *EPHX1*, *ABCC2*, and *VPS33B*. NGS was performed on Ion Torrent PGM and variant call and annotation were performed with CLC Genomics Workbench version 6.0 (CLC bio). We detected disease-causing mutations in 6 patients; one JAG1 mutation in one patient with AS, two SLC25A13 mutations in one patient with NICCD, and 4 ABCB11 mutations in 3 patients with PFIC as well as one patient with AS. For positive controls, five single nucleotide variants (SNVs) detected by Sanger sequencing were all detected, while only one out of three small insertions/deletions was identified. This study showed clinical usefulness of comprehensive mutation analysis by NGS for neonatal intrahepatic cholestasis. However, improvement on detection of small insertion/deletion remains to be fulfilled.

2580T

CIGMA (Clinical Impact of Genetic Mutational Analysis): a new approach to mutational classification in large-scale clinical genetic testing. *C. Turnbull[1, 2], R. Sultana[1], S. Mahamdallie[1], E. Ruark[1], H. Hanson[2], N. Rahman[1,2].* 1) Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) Royal Marsden NHS Foundation Trust, London, UK.

The evolution of Next-Generation sequencing (NGS) technologies has rendered dramatic expansion in the delivery of germline genetic testing for cancer predisposition both affordable and technically viable. To deliver such large-scale expansion, germline genetic testing for cancer predisposition will likely increasingly be undertaken by oncologists non-specialist in genetics. In the current model of low volume clinical genetic testing, classification of variants represents a sizeable challenge to both those who report the tests and clinicians who manage the patients. Currently, the majority of rare variants detected are classified as 'variants of uncertain significance' and these are detected in 5-10% of many individual cancer gene tests. This classification can confer substantial ambiguity with regard to communication of risk and management implications. Large-scale gene panel testing will identify significant numbers of rare variants per test, potentially scaling up this ambiguity to a level rendering large-scale non-specialist delivery of these tests unfeasible. Hence, a new era of expanded genetic testing will require commensurate new approaches to classification and reporting of mutational data. Most current approaches to variant classification are largely informed by mutational data from disease cases, which enables clear classification as pathogenic or non-pathogenic for only a small minority of variants. We have undertaken large-scale analyses of population mutational data from cancer predisposition genes and integrated these with case mutational data and gene-specific parameters relating to mechanisms of pathogenicity and the clinical relationship between gene and phenotype. Through large-scale bioinformatic integration of these data sources superimposed with clinical interpretation and decision-making, we have developed a novel system of variant classification to deliver clear, robust, consistent classification of mutational results into categories defined by clinical impact and management. We have developed an NGS panel comprising 97 cancer predisposition genes (Illumina TruSight Cancer panel) and are piloting a new model of expanded delivery of genetic testing by oncologists. The application of the CIGMA system for variant classification and reporting of mutational data is central to successful delivery of this programme (www.mcgprogramme.com). The CIGMA design principles and pilot data from our first 100 tests will be presented.

2581F

A comprehensive genetic analysis of Japanese patients with Charcot-Marie-Tooth disease using a Next-Generation Sequencing System. *A. Hashiguchi, A. Yoshimura, S. Nozuma, Y. Higuchi, Y. Sakiyama, T. Nakamura, Y. Okamoto, E. Matsuura, H. Takashima.* Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Science, Kagoshima, Japan.

[Background and Objective] Our objective was to perform genetic analysis in Japanese patients with Charcot-Marie-Tooth disease (CMT) and identify the molecular epidemiology. A DNA microarray consisting of 28 known CMT-causative genes has been applied for mutation screening since 2004. However, the positive rate of mutation detection was as low as 10%-20%. Therefore, in May 2012, we increased the number of target genes to 60 using a next-generation sequencing (NGS) system (Illumina MiSeq). [Methods] We conducted a genetic test of 182 Japanese patients with CMT from May 2012 to April 2013. Based on the median nerve motor conduction velocity (MCV), patients with CMT were broadly classified into demyelinating type (MCV < 38 m/s) and axonal type (MCV > 38 m/s). For demyelinating type CMT, the presence of the PMP22 duplication mutation was excluded using FISH before this study. When novel suspected mutations were detected, a segregation study was performed to identify the pathogenicity. The negative cases, filtered by the NGS system, proceeded to exome sequencing. [Results and Discussion] Using the NGS system, we identified pathogenic mutations in 42 (23%) of 182 patients. Thirteen mutations were identified in MFN2 (CMTX1) and GJB1 (CMT2A2). We also discovered mutations in MPZ (6/182), NEFL (4/182), SH3TC2 (2/182), GDAP1 (2/182), PRPS1 (1/182), and TRPV4 (1/182). However, segregation study data were not available for 20 variants, although they were suspected as causative mutations. If we considered these variants, the positive rate would be improved to 34% with the NGS system.

2582W

Multiplex-PCR coupled to Next-Generation Sequencing (NGS) and SNP array technologies greatly improve molecular diagnosis of Usher syndrome. *C. Bonnet[1,2], S. Chantot-Bastereaud[3], I. Sliesoraityte[1,4,9], A. Fakin[5], F. Testa[6], L. Martorell Sampol[7], S. Gherbi[2,3], S. Dad[8], S. Marlin[2,3], S. Kohl[4], D. Zobor[4], S. Mohand-Saïd[1,2,9], F. Simonelli[6], S. Banfi[6], J. Rodriguez Jorge J[7], L. Birk Moller[8], A. Kurtenbach[4], M. Hawlina[5], A. Aurricchio[6], JA. Sahel[1,2,9], I. Audo[1,2,9], E. Zrenner[4], C. Petit[1,2,10].* 1) Institut de la Vision, Paris, France; 2) UPMC, INSERM, Paris, France; 3) Hôpital Trousseau, Paris, France; 4) Institute For Ophtalmic Research University of Tübingen, Tübingen, Germany; 5) Eye Hospital University Centre Ljubljana, Ljubljana, Slovenia; 6) Department of Ophthalmology, Second University of Naples, Naples, Italy; 7) Genetica Molecular Hospital Sant Joan de Deu, Barcelona, Spain; 8) National Research Center for Genetics, Glostrup, Denmark; 9) Hôpital des XV-XX, Paris, France; 10) Institut Pasteur, Collége de France, Paris, France.

Usher syndrome (USH) is the most prevalent cause of hereditary deafness-blindness in humans. Three types of USH (USH1, USH2, and USH3) have been distinguished clinically. These are defined according to the severity of the sensorineural hearing impairment, the presence or absence of vestibular defects, and the precocity of retinitis pigmentosa onset. To date, 12 USH loci have been characterized and ten causative genes identified. In order to improve molecular diagnosis and to minimize its cost, we here developed multiplex PCR coupled to Next-Generation Sequencing (NGS) technologies. By using this unique approach in a large cohort of 400 diagnosed Usher patients, biallelic mutations were detected in 83% of patients, monoallelic mutations were detected in 13% of them and no mutation was found in 4% of patients. SNP array was performed for patients with no mutation or with only one pathogenic mutation. Remarkably, large rearrangements (deletion and/or duplication) were observed in 25% of the Usher patients with only one pathogenic mutation. The patients with no mutation found, were examined by whole exome sequencing. Of note, one patient carried a homozygous nonsense mutation in USH2A and one pathogenic nonsense mutation in USH1G. Another patient carried two pathogenic mutations (nonsense and deletion) in USH2A and two predicted pathogenic mutation (splice site variation and predicted pathogenic missense mutation) in CDH23. These observations are crucial in the perspective of gene therapy. An early and reliable diagnosis is one critical step to develop specific and adapted therapies to cure this disorder.

2583T

Panel Diagnostics for Deafness Disorders using Next-Generation Sequencing. *I. Vogl, S.H. Eck, S. Datter, S. Küçük, D. Wahl, I. Rost, S. Chahrokh-Zadeh.* Molecular Genetics, Center of Human Genetics and Laboratory Diagnostics, Martinsried, Germany.

Hearing impairments can be classified in many different ways and have an incidence rate of approximately one in 1000 births and additionally affect 50% of the population at age 80 or higher. There are multiple genetic or non-genetic causes for hearing loss. The non-genetic factors include ototoxic drugs, perinatal infections or traumas. In most cases both exogenous factors and mutations in one or more genes contribute to the phenotype. 80% of all familial, non-syndromic deafness cases are inherited in an autosomal recessive manner. In half of them mutations in the GJB2 (gap junction protein connexin 26) are responsible for the phenotype. Sometimes heterozygous mutations in the GJB2 can occur in combination with a heterozygous deletion del(GJB6-D13S1830) in GJB6 (gap junction protein connexin 30). Furthermore there are over 70 genes known which can cause different types of deafness (autosomal dominant, autosomal recessive, X-linked recessive or mitochondrial). In a pilot study we designed a deafness gene panel comprising 75 nuclear genes and 6 mitochondrial genes associated with hearing impairment. In addition to 3 positive and 7 negative controls we sequenced 2 related patients with non-syndromic deafness. Conventional diagnostics by GJB2 sequencing and GJB6 deletion analysis was exhausted and yielded no results. Sequencing was performed on the Illumina MiSeq Next-Generation Sequencing platform. Data analysis was performed using CLCbio workbench (v6.0) and custom developed Perl scripts. The target regions, in total encompassing 413,454bp and 16,569bp, were enriched via in-solution oligonucleotide hybridization and capture (Illumina TSCE). On average, 40-70% of the reads could be mapped to the human genome (build hg19) and 30-60% to the mitochondrial genome, of which between 55-60% and 96-97%, respectively were on target. On the basis of the above mentioned controls a variant calling pipeline was established and validated. Using that pipeline the 2 patients (mother and daughter) were analyzed and the A7445G mutation of the mitochondrial genome was detected. This mutation is known in the literature to be associated with deafness induced by ototoxic drugs or independent of them. Any intake of ototoxic drug was not investigated. The mutation was confirmed independently by Sanger sequencing.

2584F

Next-Generation Sequencing in the Molecular Diagnostics of Rare Diseases using a Gene Panel Approach. *S.H. Eck, I. Vogl, S. Datter, S. Kuecuek, W. Rupprecht, B. Busse, J. Hoefele, S. Chahrokh-Zadeh, C. Marschall, K. Mayer, I. Rost, HG. Klein.* Molecular Genetics, Center for Human Genetics and Laboratory Diagnostics, Dr. Klein, Dr. Rost , Martinsried, Germany.

The implementation of Next-Generation Sequencing (NGS) in clinical diagnostics opens vast opportunities through the ability to simultaneously sequence all genes contributing to a certain disease at lower cost and higher speed compared to traditional sequencing approaches. In rare and heterogeneous disorders NGS may lead to a significant improvement of the diagnostic yield. On the other hand, the practical implementation of NGS in a clinical diagnostic setting involves a variety of new challenges which need to be overcome. Among these are the generation, analysis and storage of unprecedented amounts of data, strict control of sequencing performance, validation of results, interpretation of detected variants and reporting. Here we present a panel approach for the molecular diagnostics of rare disorders. Exonic regions of more than 250 custom selected genes are enriched in parallel by oligonucleotide hybridization and capture (Illumina TSCE) and sequenced on the Illumina MiSeq instrument. During analysis, only genes from the requested indication (grouped in subpanels) are selected to limit analysis to relevant genes, while simultaneously minimizing the possibility of incidental findings. Data analysis is performed using the CLC Genomics Workbench (v.6.0.3, CLCbio) and custom developed Perl scripts. Target regions which fail to reach the designated coverage threshold of 20X are reanalyzed by Sanger sequencing and candidate mutations are independently confirmed in a seperate reaction. All detected variants are imported into an in-house relational database scheme which can be queried via a web interface for dynamic data analysis and filtering. Information from all 250 genes is used in an anonymized way for internal variant frequency calculation, quality control and the detection of potential sequencing artifacts. We have applied this approach to more than 150 samples from a variety of different disorders. In particular we use the outlined approach for the diagnostics of arrythmogenic cardiac disorders (LQTS, HCM, DCM), connective tissue disorders (EDS, TAAD), rare kidney disorders (Nephrotic Syndrome, CAKUT), neurological disorders (Noonan syndrome, Microcephalies), metabolic disorders (MODY diabetes) and pharmacogenetics.

2585W

Sequencing *CFTR* in a clinical diagnostic setting using the ion torrent personal genome machine. *L.Y. Lau[1], M. Eliou[2], J. Orr[2], C.R. Marshall[1,3], T.L. Stockley[2], S.W. Scherer[1,3], P.N. Ray[1,2,3].* 1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Ontario, Canada.

The advent of Next Generation Sequencing (NGS) technologies offers a rapid and low cost approach to replace traditional gene sequencing in the diagnostic setting. Until whole genome sequencing (WGS) becomes affordable as a diagnostic approach, the current strategy involves targeted sequencing of known disease genes. The diagnostic lab at The Hospital for Sick Children (Toronto, Canada) currently uses Sanger sequencing of *CFTR* to screen patients with cystic fibrosis (CF). In collaboration with Life Technologies™, we examined the feasibility of using targeted *CFTR* amplification followed by sequencing with the Ion Torrent Personal Genome Machine (PGM) as a Cystic Fibrosis diagnostic test. Following PCR amplification with Ion Ampliseq™ we sequenced 40 individuals to an average depth of 3500X using the PGM. On average, 25.1Mb sequenced data were obtained per sample with average read lengths of 120 bases and 354,454 reads on target. With the exception of exon 10, all 27 exons and flanking splice sites were sequenced to a read depth of over 500X. For exon 10, we were only able to obtain coverage above 15X for 58% of the exon with uniquely mapping reads because there is a high sequence similarity to a region on Chromosome 20. Using Ion Torrent Suite 3.4 for alignment and variant calling, we were able to detect 39 out of 42 (93%) known mutations of *CFTR*. We had difficulties in reliably identifying the remaining 3 mutations which included changes of the homopolymer T track on intron 8, a duplication of a G on exon 22 c.4028dupG (p.Gly1343fs), and a SNP, c.-8G>C located in the promoter region. The SNP was seen in the alignment file; however, it was not called by the software. When using the updated Ion Torrent Suite 3.6, again on default parameters, we lost the ability to detect the deletion on intron 6a c.744-33GATT[6]+[7]. The updated software also missed a SNP upstream of exon 9 c.1210-13G>T but it is seen in the alignment file. We believe we can improve the results by optimizing the parameters of the updated software. Our data suggest that the PGM can be used as a reliable diagnostic tool for clinical screening of *CFTR*.

2586T

Reproducibility of Fetal Fraction Estimates in Maternal Plasma using the Harmony™ Prenatal Test. *E. Wang, C. Struble, T. Musci, A. Batey, J. Schmidt, K. Song, A. Oliphant.* Ariosa Diagnostics, Inc, San Jose, CA.

Introduction: Fetal cell-free DNA (cffDNA) in maternal plasma enables techniques for screening for fetal aneuploidies using next generation sequencing technologies. It is known that the power to detect fetal aneuploidy is directly related to the amount of cffDNA, regardless of whether the approach employs sequencing random DNA fragments using a shotgun method or a directed analysis of targeted DNA fragments. Here, we demonstrate that Digital ANalysis of Selected Regions (DANSR™) assays for single-nucleotide polymorphisms (SNPs) on chromosomes 1-12 and chromosome Y accurately and reproducibly measure the fraction of cffDNA in maternal plasma (fetal fraction). We present data showing fetal fraction measurements using SNPs or chromosome Y alone are highly correlated, and these measurements are reproducible, as evidenced using two blood tubes from a single blood draw. **Objective**: To report clinical reproducibility results on Harmony Prenatal Test's ability to measure fetal fraction. **Methods**: This study included 11,024 maternal plasma samples containing observable counts on the Y chromosome. Samples were processed between 15 March 2013 and 17 May 2013 from pregnant women carrying a fetus at least 10 weeks in gestational age. A separate set of 1,266 research samples, in which both blood tubes were drawn at the same time but processed on different dates, were used to study the reproducibility of the fetal fraction estimates. Samples were processed as previously described (Norton et al., 2012). Fetal fraction was assessed simultaneously by assays against a set of 192 SNP loci on chromosomes 1-12 (FF-SNP) and a set of Y assays (FF-Y). **Results**: In the first study of 11,024 samples, the FF-SNP and FF-Y were correlated with an r-squared of 0.99. Among the duplicated blood tubes, the median processed time difference between a pair of blood tubes was 80 calendar days. Despite this large time difference and various possible reagent, robotic, and operator differences, the fetal fraction calculated between tubes were strongly correlated with an r-squared of 0.99 and a slope of 1. The median absolute relative difference between fetal fractions measured by the tube pairs was 4.5% (S.E. 0.2%). This implies a fetal fraction of 0.04, on average, may result in either 0.038 or 0.042 on a repeated measure from a different blood tube. **Conclusions**: Fetal fraction computed from the Harmony Prenatal Test is accurate, precise, and reproducible.

2587F

Fast STR-PCR protocol enabling rapid and high quality chimerism analysis after allo-HSCT. *W. Teng[1], H. Liu[1], F. Wang[1], Y. Wang[1], X. Chen[2], J. Fan[1], P. Zhu[2].* 1) Molecular Medicine Lab., Hebei Yanda Hospital, Sanhe, Hebei, China; 2) Department of Hematology, Peking University First Hospital, Beijing, China.

Background: Chimerism analysis based on STR-PCR is the standard method for chimerism analysis after allogeneic hematopoietic stem cell transplant (allo-HSCT). The cycling time needs about 3.5 hours, and incomplete adenylation of PCR products often become a problem. Here we aimed to develop a fast PCR protocol enables rapid and high quality chimerism analysis. Methods: Peripheral blood and fingernail samples were collected from healthy donors and patients. Simulated mixed chimerism samples were prepared by fold dilution. The AB Identifiler Kit with 15 STR loci was used. AmpliTaq Gold polymerase and standard PCR cycling program were used as contrast. Newly bio-engineer modified fast polymerase MyTaq, Phusion and Q5 were tested with rapid cycling programs. The PCR products were capillary electrophoresed and chimerism was analyzed using AB3500XL and GeneMapper IDX software. Results: All the four kinds of polymerase can be successfully used for STR-PCR and genotyping. The STR-PCR cycling time for MyTaq, Phusion and Q5 were about 30 minutes, considerably less than 3.5 hours when using AmpliTaq Gold. For the 3 kinds of fast polymerases, the Q5 polymerase showed the best balanced amplification efficiency among the 15 STR loci. The PCR product of Phusion and Q5 don't have non-template addition 'A' tail due to lack of adenylate activity, yet the allelic ladders without 'A' tail must be specially made. The MyTaq has extremely efficient of adenylate activity and showed complete adenylation even using crude DNA samples extracted from fingernail, thus actually fully resolved the problem of incomplete adenylation. Conclusion: The STR-PCR cycling time got a drastic reduction and the problem of incomplete adenylation was resolved by using the novel MyTaq and rapid PCR program, thus enabling rapid and high quality chimerism analysis after allo-HSCT.

2588W

New method for detecting mutation with high sensitivity using capillary electrophoresis DNA sequencer. *T. Yokoi, M. Yoshida, T. Anazawa.* Central Research Laboratory, Hitachi.Ltd, Tokyo, Japan.

By recent progress in NGS technologies, a genomic era for cancer studies is growing rapidly. In this era, automated capillary electrophoresis (CE) DNA sequencer by sanger dideoxy terminator is still a widely used for validating gene sequences, and is considered gold standard for target gene Analysis. However, CE DNA sequencers cannot detect low-frequency mutation, because sequencing workflow and subsequent signal processing of CE DNA sequencers are optimized for detecting germline mutation (50% existence). To address this issue, we developed a new method for detecting mutation with high sensitivity using CE DNA sequencers. In the method, we first modified sequencing workflow, namely, measuring four kinds of bases in separate capillaries, to reduce to crosstalk-fluorescence signal. Second, we implemented a procedure of extracting relative intensity of each peak to a neighbor peak of the same kind of base in a raw fluorescent signal data. Dispersion of relative intensity of each peak was found to be small enough to detect mutation with high sensitivity by comparing with relative intensity of reference data. This developed method successfully detected 5% existence mutation in EGFR and KRAS without visual inspection. The proposed method shows a possibility to detect mutation of much lower existence, and is promising for practical use to detect genetic mutations.

2589T

Whole-genome prenatal sequencing and integrative genomics: Detection of structural variation from invasive and non-invasive approaches. *M.E. Talkowski[1,2,3], V. Pillalamarri[1], H. Brand[1], I. Blumenthal[1], M.Z. Ordulu[5], S. Pereira[5], J. Kitzman[6], J. Shendure[6], J.F. Gusella[1,2,3,4], C.C. Morton[4,5].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute, Boston, MA; 5) Departments of Obstetrics, Gynecology and Reproductive Biology and Pathology, Brigham and Women's Hospital and Harvard Medical School; 6) Department of Genome Sciences, University of Washington, Seattle, WA.

Technical innovation has had a dramatic impact in prenatal genetic diagnostics, particularly for congenital anomalies where de novo structural variations (SV) represent a major source of risk but are not considered by non-invasive cell free fetal DNA ($_{cff}$DNA) studies that are limited to detection of aneuploidy. We recently performed the first example of prenatal clinical diagnosis from whole-genome sequencing (WGS) using large-insert 'jumping libraries' at a cost and timeline comparable to karyotyping. Here, we present real-time clinical sequencing and integrative genomic interpretation of fetal DNA from four individuals (one twin pair) referred for advanced maternal age harboring *de novo* balanced SVs. In two subjects, DGAP247 and DGAP248, protein-coding genes were disrupted (*KHDRBS3* and *RCF3*, respectively) but interpretation from large independent datasets suggested a variant of unknown significance (VOUS) for each. The parents of DGAP247 chose to receive these findings and carried DGAP247 to term resulting in a well newborn examination. The parents of DGAP248 chose not to receive the findings. We also performed deep WGS on cffDNA isolated from maternal plasma of DGAP247 during the third trimester. We generated 53X genome coverage from 1.72 billion reads, finding a fetal DNA composition of ~25% of reads. We used a split read algorithm to detect the inversion from 14% of reads (~28% fetal coverage at the locus), demonstrating feasibility of SV detection from very deep WGS of $_{cff}$DNA. Transcriptome studies are ongoing. In the twins (DGAP258) we identified a pericentric inversion that rearranged the known copy number morbid 6p25.3 locus, as well as an independent cryptic inversion, but neither disrupted genic sequence. The consequences of inverting a recurrent copy number morbid region but not disrupting genic sequence are unknown, but the twins were recently born with a well newborn examination. These data highlight the power of prenatal WGS by jumping libraries to detect pathogenic SVs of all classes, however they are reliant on invasive methods that carry substantial risk to the fetus. They also show the potential feasibility of SV detection by non-invasive methods, though costs and analytical effort were exorbitant. Collectively, our studies emphasize the unique challenges facing development of a comprehensive prenatal genetic screening strategy that accesses the entire pathogenic mutational spectrum but minimizes risks to the fetus.

2590F

A novel and rapid digital PCR-based method for the identification of 22q11.2 Deletion Syndrome in large population screening. *V. Hwang[1], D. Maar[2], J. Regan[2], T. Simon[3], F. Tassone[1, 3].* 1) Department of Biochemistry and Molecular Medicine, UC Davis, Davis, CA; 2) Digital Biology Center, Bio-Rad Laboratories, Pleasanton, CA; 3) MIND Institute, UC Davis Medical Center, Sacramento, CA.

22q11.2 Deletion Syndrome (22q11DS), the most common microdeletion syndrome in humans, is characterized by a wide range of clinical manifestations including craniofacial defects, developmental delay, neurological and psychiatric problems, autism spectrum disorders and congenital heart defects. 70-80% of individuals have a 3 Mb deletion, 15-30% have a nested 1.5 Mb deletion, and the remainder have atypical deletions within the 22q region. Because the syndrome is associated with over 100 different diagnostic findings that range from mild to severe and life-threatening conditions, diagnosis is often delayed or missed. Therefore, due to this high variability, propositions to include 22q11DS in newborn screening panels have been suggested.

The currently accepted method for diagnosis is fluorescent in-situ hybridization (FISH). This approach, however, is expensive, labor intensive, and requires special technique and equipment. We have developed an inexpensive, rapid, sensitive and specific test that can identify newborns with 22q11DS. The approach uses a droplet digital PCR (ddPCR) that processes samples collected on blood spot cards and provides accurate and rapid diagnosis while keeping costs at a minimum-the ideal premise for large population screenings, such as newborn screening. We performed a validation study where we blindly screened over 500 anonymous blood spots collected from the general population that included a small number of 22q11DS samples. We demonstrate that ddPCR definitively assesses DNA copy number variation from DNA isolated from 903 blood spot cards and provide evidence that this methodology is highly specific and reliable. A large population screening is needed to establish the prevalence of 22q11DS in the general population, which has not been well assessed to date. In addition, the need for newborn screening for 22q11DS is highly sought as diagnosis at birth allows for immediate mediation of symptoms that leads to a better disease prognosis. To our knowledge, this study is the first to demonstrate the efficacy of ddPCR in large population screening studies, an advent that holds great promise for families and clinicians alike.

2591W

Detection of 22q11.2 deletion syndrome in Colombian patients with isolated congenital cardiopathy by MLPA. *T. PINEDA[1], O. MORENO[2], I. ZARANTE[3].* 1) Medical Genetics Resident, Genetics Institute, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Biol, Msc. Genetics Institute, Pontificia Universidad Javeriana. Bogotá, Colombia; 3) Md, Phd. Genetics Institute, Pontificia Universidad Javeriana. Bogotá, Colombia.

22q11.2 deletion syndrome is one of the most common genetic syndromes and it has a wide phenotypic spectrum, abnormalities can include cardiac defects, immunodeficiency, hypocalcaemia, speech and language impairment, cleft palate, development delay and learning difficulties. The aim of this study was to detect the 22q11.2 deletion by MLPA (kit SALSA MLPA Probemix P250 B1 DiGeorge) in Colombian patients presenting with isolated congenital cardiopathy in order to establish the prevalence of this syndrome in this group of patients. The study included 26 patients with congenital cardiopathy that were enrolled from different hospitals in Bogotá (23 patients with contruncal cardiopathy, 1 patient with VSD and Coarctation of the aorta, 1 patient with VSD and overriding aorta and 1 patient with ASD and PDA), the group of patients range from 1 day old to 13 years old. Exclusion criteria included cleft lip or cleft palate, mental retardation and any other major anomaly. The results showed that the classical 22q11.2 deletion is present in the 15,3% of the patients (4/26), we did not find any atypical deletion in this group. With this study we can conclude that MLPA is a very useful molecular method that provide an accurate diagnosis and that it should be implemented in all patients presenting with isolated congenital cardiopathy related to 22q11.2 deletion syndrome.

2592T

A comparison of CNV endpoint accuracy between CytoScan® Dx assay and Next Generation Sequencing. *A. Roter[1], B. Eynon[1], S. Close[2], K. Kwiatkowski[1], D. Ballinger[3], S. Yang[3], R. Duttagupta[1], C. Chen[1], K. Suyenga[1], A. Singh[1], T. Chen[1], M. Chadha[1], E. Fung[1].* 1) Affymetrix, Inc., Santa Clara, CA; 2) Indiana University School of Medicine, Division of Clinical Pharmacology, Indianapolis, IN; 3) Complete Genomics, Mountain View, CA.

Background: The genome-wide distribution of ~2.7 million markers enables CytoScan® Dx assay to determine copy number state with very high resolution. To assess the resolution of CytoScan Dx assay, the accuracy in determining the genomic location of the endpoints of copy number variation (CNV) regions was evaluated. Endpoint location was compared between next generation sequencing and CytoScan Dx assay for 1,387 concordant CNV endpoints in post-natal samples.

Materials & Methods: 137 samples (blood and cell-line derived DNA samples) were selected to maximize variation across the genome. The CytoScan Dx assay was run at Affymetrix and next generation whole genome sequencing was independently performed and analyzed in a blinded fashion at Complete Genomics. Endpoint agreement was pre-specified as the sequence-defined endpoint of ± 12 markers for loss CNVs and ± 25 markers for gain CNVs, translating into an effective resolution of 25kb for losses and 50kb for gains. In addition, copy number accuracy was assessed by comparing results with those determined by routine patient care methods such as FISH or karyotype.

Results: The overall endpoint agreement was 93.4% and was similar for copy number loss regions (92.5%) and for copy number gain regions (94.7%). The median absolute endpoint difference was 1 marker across all comparable endpoints. Additionally, CytoScan Dx assay and routine patient care methods agreed 100% of the time on copy number variation.

Conclusions: The CytoScan Dx assay is highly accurate in detecting copy number variation. For those copy number variation regions of 25kb and 25 markers for loss segments and 50 markers and 50kb for gain segments, CytoScan Dx assay was able to determine the genomic location within ± 1 marker of the sequence determined endpoint demonstrating very high resolution and a high degree of accuracy in determining the copy number variation endpoint locations.

2593F

Accurate detection of small and large copy number events from targeted next generation sequence data. *K.B. Jacobs, J.P. Paul, G. Nilsen, M. Mikhaelian, R. Hart, M. Johnson, S.E. Lincoln, J.M. Sorenson.* InVitae Corporation, San Francisco, CA.

Germline copy number variants (CNVs) can be detected from next-generation sequencing (NGS) data generated using targeted DNA capture technologies (e.g. exomes and other panels), however methods for doing so must overcome many technical challenges. Several algorithms have been published to detect CNVs in such data, though they may not yet be adequate for use in diagnostic testing laboratories, particularly for detection of small single-exon CNVs. Thus, diagnostic testing laboratories often resort to expensive and low-throughput methods such as MLPA to discover and confirm small CNVs. As a result, clinicians must carefully decide whether to order both a sequencing test and a deletion/duplication test for their patients. A single test that can accurately assay both types of alterations would improve patient access to comprehensive genetic testing.

We present a new method, CNVitae, which is designed to detect single-exon CNVs as well as larger regions sequenced using NGS. CNVitae is based on a statistical model for read counts and employs model-based segmentation algorithms optimized for use with sparsely distributed and highly variable targets across the genome. This framework estimates the most likely copy number for all segments, and, critically for clinical use, each called segment is assigned a robust quality score indicating confidence in the copy number determination.

We evaluated CNVitae on high-depth targeted NGS data generating using Agilent SureSelect capture and Illumina MiSeq 2x150 paired-end sequencing. Under a simulation model, altering the observed laboratory data *in silico*, we achieved a sensitivity and specificity of >99% to detect single exon hemizygous deletions at a strict confidence threshold of Q25 (probability of error < 0.5%). We saw 97% sensitivity and >99% specificity to detect single exon duplications (CN=3) while four exon duplications were detected with sensitivity of >99%. In 40 patients independently known to carry clinically relevant CNVs, we detected all 37 of the single-exon or larger events with high confidence, although three sub-exon scale CNVs were not detected by the current algorithm. Additional validation results will be represented, including results on additional clinical samples, the Get-RM Copy Number Variation Reference Panel, and a re-analysis of a subset of exome data from the 1000 Genomes Project.

2594W

Assessing Common Maternal Copy Number Variation during cfDNA Analysis for Non-Invasive Prenatal Testing using Digital Analysis of Selected Regions (DANSR™) Assays. *C. Struble, E. Wang, J. Schmidt, A. Batey, T. Musci, K. Song, A. Oliphant.* Ariosa Diagnostics, Inc. San Jose, CA.

Objective: To report on Harmony™ Prenatal Test's ability to provide accurate results in the presence of a maternal CNV. **Method**: As previously described, DANSR assays were used to analyze common trisomies in maternal samples. The DANSR assays were selected to target nucleic acid regions unique to the chromosomes of interest in a maternal sample, and to avoid common copy number variations (CNVs) (Sparks et al, 2012). Fetal fraction was assessed in the maternal samples simultaneously with the trisomy analysis using DANSR assays against a set of 192 single nucleotide polymorphism (SNP)-containing loci on chromosomes 1-12. Using the estimated fetal fraction, a sample specific expectation is determined to evaluate if increased chromosomal counts using the DANSR assay are consistent with fetal aneuploidy and/or other clinical factors such as maternal CNV. **Results**: In approximately 64,000 samples, 35 (0.05%) putative maternal CNVs were observed. Sixteen of these 35 (47%) belong to a 1.5 mega base maternal duplication on chromosome 21q21.3. Of these 16 cases, there was one case of a high-risk for trisomy 21 result, and 15 cases of low-risk results in the presence of this maternal CNV. The case of high-risk for trisomy 21 was subsequently confirmed by amniocentesis. Genes in this duplicated region includes two secreted metelloproteases, ADAM metallopeptidase with thrombospondin type 1 motif (ADAMTS1) and ADAM metallopeptidase with thrombospondin type 5 motif (ADAMTS5). ADAMTS1 has been reported to play a critical role in follicular rupture. ADAMTS5 has been reported to degrade aggrecan, which is a proteoglycan of the cartilage and is associated with inflammation and arthritis. An additional gene in this region include the Cysteine and typrosine-rich 1 (CYYR1) with high conservation among vertebrates and high expression in cells belonging to the diffuse neuroendocrine system. **Conclusion**: DANSR's design allows for efficient and uniform sequencing across the chromosome loci with enough depth to evaluate the risk for fetal aneuploidy in the presence of previously unreported maternal CNV.

2595T

Generating High Confidence Next-Generation Sequencing Variant Calls for Clinical Diagnostic Use. *P.J.B. Sabatini[1], L. Lau[2], M. Eliou[1], J. Orr[1], P. Ray[1], T. Stockley[1].* 1) The Hospital for Sick Children, Toronto, Canada; 2) The Centre for Applied Genomics, Toronto, Canada.

Using next-generation sequencing for clinical diagnostic testing poses significant challenges associated with analytical sensitivity and specificity. Discovery of false positive variants lead to misdiagnoses and false negatives neglect clinically relevant ones. Understanding the properties of false variant calls from next-generation sequencing platforms will assist with tailoring the variant calling programs to generate high confidence sequence variants important for clinical diagnostics. The molecular genetics laboratory at the Hospital for Sick Children in Toronto, Canada is developing next-generation sequencing panels using the Ion Torrent platform. To observe the properties of false variant calls, three separate runs were performed using a known genotyped sample - J. Craig Venter - with the Inherited Disease Panel from Life Technologies. The Inherited Disease Panel targets the coding regions of over 300 genes associated with human genetic conditions and covers a wide variety of sequencing contexts that is important in understanding the properties of variant calling and their genotypes. After sequencing on the Ion Torrent the reads were mapped to the human reference genome sequence (hg19) using TMAP 3.4.3 and variant calls were made using the Ion Torrent Variant Caller's (3.4.2) default settings with an average read length of 135 base pairs. Of the known 1,005 SNPs present in the area targeted from the Venter sample an average of 74 variants were called incorrectly (93% sensitivity) and of the 150399 known references bases sequenced 74 were missed (99.9% specificity). The average read depth for correct homozygous calls was 347 (s.d=312) versus 157 (s.d=173) for false positives and for true heterozygous calls the read depth was 374 (s.d=325) versus 312 (s.d= 230) for false positive heterozygous calls. The ratio of the alternative allele was 1.00 (s.d.=0.02) for correct homozygous calls and 0.81 (s.d.=0.15) for false homozygous calls. Average heterozygous alternative allele ratio was 0.49 (s.d.=0.07) for correct calls and 0.30 (s.d.=0.27) for false positives. There were no significant differences in strand bias. Limiting the analysis to read depths over 350 improved the positive predictive value from 93% to 96%. By designing testing strategies that generate read depths above 350 and narrowing the filtering range of the allelic ratios we can improve the confidence of the variant calls important for clinical diagnostics.

2596F

Validation and implementation of a 19-gene 2nd tier Rett/Angelman syndrome next generation sequencing panel. *J.R. Jones, M.J. Basehore, S. McGee, K. Kubiak, J. Butler, K. King, J.A. Lee, M.J. Friez.* Greenwood Genetic Center, Greenwood, South Carolina, USA.

Angelman and Rett syndromes are common genetic disorders that share many similar features, including developmental delay, intellectual disability and seizures. The genetic heterogeneity among these two syndromes and many other similar disorders make it challenging to determine the appropriate second tier of diagnostic testing once the most common causes have been ruled out. In an effort to provide a more rapid, comprehensive and cost-effective analysis for patients with features overlapping those of Rett and Angelman syndromes, the Molecular Diagnostic Laboratory at the Greenwood Genetic Center (GGC) has designed and validated a 19-gene panel which utilizes RainDance™ Technologies microdroplet enrichment and SOLiD™ Next Generation Sequencing (NGS). The panel includes *MECP2* and *UBE3A* along with 17 additional genes known to cause conditions that often present with clinical features similar to Rett and Angelman syndromes. The coverage of the panel includes all exons for each gene as well as flanking intronic regions. Thus far, 38 samples have been submitted for clinical testing. Of the 17 completed analyses, eight patients had normal results, and six patients had at least one variant of unknown clinical significance. We identified pathogenic changes in three patients (~18%). Interestingly, one of these patients was found to have a one basepair deletion in *MECP2* (c.271delC). *MECP2* sequencing had not been previously requested for this patient since she does not have a classic Rett phenotype. Unlike most patients with Rett syndrome who have microcephaly, short stature and cachexia, at seven years of age this patient has a normal head circumference (47th percentile), height (90th percentile) and weight (75th - 80th percentile). She is ambulatory and non-verbal with severe intellectual disability. Therefore, this panel may also be useful in detecting *MECP2* and possibly *UBE3A* mutations in individuals with atypical presentations of Rett and Angelman syndromes, in addition to being a second tier test for *MECP2* and *UBE3A* mutation-negative patients. Our initial findings indicate that this expanded Rett/Angelman panel will provide a cost-effective method for testing the genes believed to be of the highest diagnostic priority for individuals who fall into this clinical spectrum.

2597W

High-throughput screening for *SMN1* copy number loss by next-generation sequencing. *E. Boyden, G. Porreca, M. Umbarger.* Good Start Genetics, Cambridge, MA.

Spinal muscular atrophy (SMA) is a lethal autosomal recessive neuromuscular disorder caused by functional loss of the *SMN1* gene. The high carrier rate of ~1 in 40 is attributable primarily to *SMN1* copy number loss produced by either deletion of part or all of *SMN1*, or conversion of *SMN1* to *SMN2*, a linked paralog that encodes an identical protein but is poorly expressed due to a silent non-coding variant that disrupts proper splicing. Clinical SMA carrier screening is currently performed via methods such as multiplex ligation-dependent probe amplification (MLPA) that enable the assessment of the copy number status of *SMN1* in a manner that distinguishes between *SMN1* and *SMN2*, typically by interrogating the exon 7 variant. Such methods are sensitive and specific, but relatively low-throughput.

We have developed a method by which the copy number status of *SMN1* and *SMN2* can be measured directly by next-generation sequencing. Our approach uses molecular inversion probes to capture loci specific to *SMN1* and *SMN2*, and subsequently employs ratios of normalized sequence read depths to infer copy number status. Our method shows sensitivity and specificity for detection of *SMN1* copy number loss similar to that of MLPA, but has the advantage of being compatible with automated high-throughput screening. Furthermore, because our method uses sequencing as the readout, it can detect deleterious *SMN1* point mutations and indels that would be missed by MLPA and related approaches. Thus, our new SMA carrier screening assay represents a significant improvement over standard SMA carrier screening methods.

2598T

Establishing Performance Specifications for Clinical Whole Exome Sequencing. *M.O. Dorschner[1,2], S.J. Anover-Sombke[2], J. Gasper[2], S. McGee[2], T. Shaffer[2], K. Patterson[2], J.D. Smith[2], G.P. Jarvik[2,3], D.A. Nickerson[2].* 1) Psychiatry & Behavioral Sciences; 2) Genome Sciences; 3) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle WA.

Whole exome sequencing (WES) is rapidly becoming an effective tool for the molecular diagnosis of rare genetic disorders. WES is particularly well suited for disorders in which the genes are known but diagnostic assays are unavailable or those conditions with closely, difficult to distinguish phenotypes. Prior to launching WES as a clinical test, laboratories must establish performance specifications in accordance with regulations set forth by the Clinical Laboratory Improvement Amendments (CLIA) or College of American Pathologists (CAP). Sanger-based sequencing tests have typically been validated by the examination of a small set of known, mutation-positive and negative samples. The ability to detect these variants accurately, serve as the basis for analytical sensitivity, specificity and reproducibility of the test. Testing only a small number of variants, from a limited set of samples is not adequate to represent the performance of large-scale assays such as WES and whole genome sequencing. To validate our WES protocol, we chose to examine 12 well-characterized HapMap samples, in duplicate. All samples were previously genotyped using the Illumina 1M array, and >60,000 markers overlapping the exome were examined for concordance with exome sequence data. With a mean coverage of 120X (90% of the exome covered at >20X), we calculated the sensitivity and accuracy of our WES protocol at >99% for single nucleotide substitutions. Poorly covered exons were largely the result of high GC content and/or low read mapability. We believe that comparing a much larger set of variants, more accurately reflects the true performance characteristics of our WES assay. With such high accuracy, we have developed an algorithm based on coverage and sequence quality to eliminate the need for Sanger confirmation. Only those variants exhibiting sub-optimal coverage, quality or strand bias are routinely Sanger confirmed. Sequencing technology continues to evolve at a rapid pace necessitating an organized procedure for re-validation of WES protocols as changes are incorporated into clinical workflows. Scalable quality control measures will be required for wider implementation of genomic medicine.

2599F

Development and validation of a synthetic, single-reagent, positive control for comprehensive high-throughput carrier screening. *A.M. Fedick[1,2], C. Jalas[3], N.R. Treff[1,2].* 1) Department of Microbiology and Molecular Genetics, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ; 2) Reproductive Medicine Associates of New Jersey, Basking Ridge, NJ; 3) Bonei Olam, Center for Rare Jewish Genetic Disorders, Brooklyn, NY.

Carrier screening tests that offer parallel analysis of multiple mutations, genes, and diseases are in high demand. Contemporary platforms include testing for hundreds of pan-ethnic mutations at once using methods such as TaqMan allelic discrimination, DNA microarrays, or next generation sequencing. One of the greatest challenges in developing new carrier screening assays is obtaining positive controls. Furthermore, there is no single reagent that can be used as a positive control for all mutations being evaluated, particularly given that there are no individuals who naturally carry the mutant allele for all mutations tested. As a proof-of-principle, we set out to develop a novel synthetic control to provide a single reagent suitable for routine quality control during the development, validation, and performance of TaqMan-based comprehensive carrier screening. A set of 29 ACOG recommended Ashkenazi mutations, for which TaqMan assays had been previously developed and validated using known carriers, were included in designing a synthetic control solution. The synthetic component of the reagent involved the creation of gBlocks (Integrated DNA Technologies). Each gBlock consisted of sequence covering the entire TaqMan amplicon and incorporated sequence for the mutant allele. Multiple amplicons were included in a single gBlock with the maximum size reaching 500 basepairs to enhance cost efficiency. A total of 8 gBlocks were designed, ordered, and pooled together. Initial validation was performed on a 7900 HT real-time PCR instrument using 384 well plates. Two synthetic controls failed to amplify and were excluded from further analysis. The remaining controls were added at a 1:1 ratio with DNA from an individual known to possess normal alleles for all mutation sites. The final mixture was used as template in conventional TaqMan allelic discrimination analysis using first a 7900 and then a QuantStudio 12K Flex Real-Time PCR System for high-throughput genotyping. Successful genotypes were made for 96.3% (26/27) of the synthetic controls on the 7900 and for 92.6% (25/27) on the QuantStudio. The two failed controls both involved mutations in the GBA gene, which has a known pseudogene, and are now being redesigned to achieve 100% coverage. In conclusion, this proof-of-principle study demonstrates a novel strategy to provide a single reagent suitable for routine quality control for high throughput carrier screening.

2600W

Limited clinical utility of Whole Exome Sequencing in the diagnosis of hereditary neuropathies. *O. Jarinova[1], J. Warman[1,2], J. Schwartzentruber[3], C. Goldsmith[1], N. Carson[1], E. McCready[1], G. Yoon[4], S. Baker[5], A.M. Innes[6], C. Beaulieu[1], A. Smith[1], T. Hartley[1], K. Boycott[1].* 1) Children's Hospital of Eastern Ontario, Ottawa; 2) Division of Neurology, Department of Medicine, The Ottawa Hospital; 3) McGill University and Genome Quebec Innovation Centre, Montréal, Québec; 4) Division of Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto; 5) Department of Medicine, Divisions of Physical Medicine and Rehabilitation, McMaster University, Hamilton; 6) Department of Medical Genetics, Faculty of Medicine, University of Calgary, Alberta.

Hereditary neuropathies comprise a distinct group of inherited disorders in which peripheral nerves undergo progressive degeneration, leading to significant physical disability of the hands and lower legs. While autosomal recessive and X-linked forms have been described, hereditary neuropathies predominantly show autosomal dominant inheritance. Causative mutations can reside in any of a number of genes responsible for various aspects of maintenance of peripheral nerves and result in an overlapping clinical spectrum for these disorders. Currently available molecular diagnostic options only allow the sequential analysis of a limited number of these genes. Therefore determining the genetic cause for a specific patient/family is often time-consuming and expensive.

We questioned whether Whole Exome Sequencing (WES) could facilitate the identification of causative mutations in patients with hereditary neuropathies. We sequenced the exomes of twelve unrelated patients without molecular diagnoses after mutations in the four common neuropathy genes (*PMP22*, *GJB1*, *MPZ*, *MFN2*), which account for over 70% of molecular diagnoses, had been ruled out. We screened the sequence data for the presence of likely damaging coding and splice-site variants using an in-house analysis pipeline and manual review of variants. While this approach has proven to be successful for diagnosis of other genetic disorders, it failed to identify obvious causative mutations in our patients. Our findings suggest that at this time using WES for the diagnosis of patients with hereditary neuropathies is challenging. This study emphasizes the need for refinement of analysis methods and further assessment of clinical utility of WES for the diagnosis of autosomal dominant heterogeneous conditions.

2601T

Comprehensive massive parallel DNA sequencing strategy for the genetic diagnosis of the Neuro-cardio-facio-cutaneous syndromes. *A. Justino[1], P. Dias[2], M.J. Pina[1], C. Ribeiro[1], S. Sousa[1], L. Cirnes[1], A.B. Sousa[2], J.L. Costa[1], J.C. Machado[1,3].* 1) Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; 2) Department of Genetics, Hospital de Santa Maria de Lisboa, Portugal; 3) Medical Faculty of the University of Porto, Porto, Portugal.

Background: The Noonan, Cardio-facio-cutaneous, Costello and LEOPARD syndromes are members of the Neuro-cardio-facio-cutaneous syndromes group (NCFCS). Mutations in 11 genes have been causally linked to these disorders (*PTPN11, SOS1, BRAF, RAF1, MEK1, MEK2, NRAS, KRAS, HRAS, SHOC2 and CBL*). Recently, an exome sequencing study also associated the gene *A2ML1* with this syndrome. Due to the genetic and clinical heterogeneity of these disorders it is challenging to define straightforward strategies of sequential gene analysis for their molecular diagnosis. Therefore, the aim of this study was to develop and validate a massive parallel sequencing (MPS) based strategy for the molecular diagnosis of NCFCS. Methods: A multiplex PCR-based strategy for the enrichment of the 12 genes and a dedicated variant prioritization pipeline (VPP) was established. Two sets of genomic DNA samples of clinically defined cases of NCFCS were studied using the Ion PGM: a training set (15 cases) used to optimize the strategy and a validation set (20 cases) used to validate and evaluate the power of the new methodology. Sanger sequencing was performed to confirm all variants and fill in regions with insufficient read coverage. Results: All variants identified by Sanger sequencing were detected with our MPS approach. The most frequent mutated gene was *PTPN11* (n=8). Other known disease causing mutations were found in *BRAF* (n=1), *SHOC2* (n=1) and *RAF1* (n=2). Additional genetic alterations of unknown significance were identified in *SOS1* (n=2), *CBL* (n=3) and the novel *A2ML1* (n=10). The methodology resulted in an experimental approach with a specificity of 99.4% and a maximum analytical sensitivity ≥98.1% with a confidence of 99%. The analysis of the 12 genes incremented in 15% the diagnostic yield of the strategy currently used by our laboratory for the molecular diagnosis of NCFCS. Conclusions: Here we present a workflow that provides a comprehensive genetic screening strategy for patients with NCFCS in a fast and cost-efficient manner. This approach demonstrates the potential of a combined MPS-Sanger sequencing based strategy as an effective diagnostic tool for heterogeneous diseases.

2602F

Blood Group Genotyping on the TaqMan® OpenArray®. *M. Laig[1], C. Colvin[2], M.A. Keller[2], E. Grigorenko[3], T. Horn[2], J. Crowley[2], D. Fantin[1].* 1) Life Technologies, South San Francisco, CA; 2) American Red Cross, Philadelphia, PA; 3) Diatherix Laboratories, Inc, Huntsville, AL.

Transfusion patients depend on well characterized blood donations. Incompatible blood types can cause severe hemolytic reactions. Blood group antigens have a genetic basis, in most instances they are based on a single nucleotide polymorphism (SNP). Genotyping is more accurate than serology based tests since polyclonal antibodies may miss detection of low expression of antigens. For some blood group antigens, no antibodies are available at all. We developed a high-throughput screening panel on the TaqMan® OpenArray® covering 28 markers across 14 blood groups: Colton Co(a), Cromer Cr(a), Diego Di(a/b), Dombrock Do(a/b), Hy, and Joa, Duffy Fy(a/b), Fy(x) and GATA Silencing, Kell Js(a/b), K/k, and Kp(a/b), Kidd Jk(a/b), Knops Kn(a/b), Landsteiner-Wiener LW(a/b), Lutheran Lu(a/b), MNS MN (GYPA nt59, GYPA nt71/72) and MNS Ss and U (Intron5, nt143, nt230), Rh RHCE C/c, E/e, L245V, M238V, G336C, Scianna Sc(a/b), Yt(a/b). Challenges overcome with this panel are extensive homologies in the MNS group between Glycophorin A, B, and E genes and in the Rh group between the RHCE and RHD genes that pose challenges to assay development. The OpenArray plates are run on the QuantStudio™ 12K Flex with a throughput of roughly 400 samples in 3.5 hours. Genotypes were successfully validated on control samples. Blood Group Genotyping on the OpenArray platform provides an accurate tool for high-throughput screening of blood units which can be a complementary approach to existing serology tests.

2603W

Clinical Validation of Noninvasive Prenatal Screening for Fetal Sex Chromosome Aneuploidies in Maternal Plasma using Direct ANalysis of Selected Regions (DANSR™) Assays. *T. Musci, C. Struble, E. Wang, J. Hooks, J. Schmidt, K. Song, A. Oliphant.* Ariosa Diagnostics Inc., San Jose, CA.

Introduction: Fetal cell-free DNA(cffDNA) in maternal plasma enables techniques for screening for fetal aneuploidy using next generation sequencing technologies. We have previously described a method using DANSR assays for biochemical analysis of chromosomes 13, 18, and 21, combined with the Fetal fraction Optimized Risk for Trisomy Evaluation (FORTE™) algorithm to compute the risk of trisomy based on the DANSR assay results with high sensitivity and specificity. Both components are integral parts of the commercially-available Ariosa Harmony Prenatal™ Test. In this study, we have developed additional DANSR assays for the X and Y chromosomes and have applied the FORTE algorithm on a blinded set of samples. The set contained samples with and without sex chromosome aneuploidies (SCA). **Objective**:Report clinical validation results on Harmony Prenatal Test's ability to detect fetal SCA. **Methods**: 432 subjects were selected. Participants provided informed consent under an IRB approved protocol. All subjects underwent invasive testing. Resulting karyotypes were used to confirm the Harmony Prenatal Test results. Samples were processed as previously described with lab and analysis personnel blinded to the fetal karyotype. FORTE models were built against monosomy X, XXX, XXY, XYY, and XXYY genotypes. Results were compared against the fetal karyotypes. **Results**:Of the 414 plasma samples that passed standard QC metrics for the Harmony Prenatal Test, all generated a sex chromosome result (100%; 95% CI: 99.1-100%). All samples were concordant with the fetal karyotyping result for fetal sex (100%; 95% CI: 99.1-100%). 26 of 27 monosomy X samples classified as high risk agreed with karyotype (sensitivity 96.3%; 95% CI: 81.7-99.8%), with two discordant high risk results (specificity 99.5%; 95% CI:98.1-99.9%). One XXX sample classified as high risk was concordant with karyotype with two discordant high risk results (specificity 99.5%; 95% CI:98.1-99.9%). All of six XXY calls were concordant (sensitivity 100%; 95% CI:61-100%; specificity 100%; 95% CI:99.0-100%). **Conclusions**:Directed analysis of cffDNA with DANSR assays allows for risk assessment of non-mosaic fetal SCA. While this is the largest fetal SCA validation study done to-date, due to the complex nature of SCAs, and possible maternal mosaicism, future larger studies are warranted. This study also demonstrates the ability to expand the Harmony Prenatal Test to genetic conditions other than trisomies 13,18 and 21.

2604T

Comparative Study For The Evaluation Of A New Technology For Cystic Fibrosis Screening. *M.CA. Rongioletti[1], F. Papa[1], C. Vaccarella[1], MB. Majolini[1], A. Luciano[1], C. Centrone[2], B. Minuti[2], V. Mazzucchi[1], M. Belli[1], I. Giotti[2], C. Giuliani[2], F. Torricelli[2], G.M. Liumbruno[1].* 1) Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy; 2) SOD Diagnostica Genetica, AOU Careggi, Florence, Italy.

Introduction Cystic fibrosis (CF) is one of the most frequently diagnosed autosomal-recessive diseases in the Caucasian population. Screening for Cystic Fibrosis Transmembrane conductance Regulator (CFTR) gene mutations, including poly T, is strongly recommended in infertile couples planning a pregnancy by assisted reproductive technology (ART). This study evaluated the performance of the new Nanochip CF70 kit (Savyon Diagnostic, Israel), a microarray assay, and compared it with the Innolipa kits (Innogenetics, Belgium) **Methods** From January to July 2012 we analyzed 392 blood samples with Innolipa and Nanochip technologies that identify respectively 70 and 56 CFTR mutations. Both tests include the most common Italian mutations and the poly-T screening. Discordant results were analyzed with the Devyser CFTR Core Kit (Devyser, AB, Sweden) based on PCR allele specific technology, MLPA (MRC Holland), Direct Sequencing (DS) on the 3730 DNA Analyzer (AppliedBiosystems), and Sequenom's MassArray system (Diatechpharmacogenetics, Italy) **Results** Innolipa and NanoChip were concordant for 371/392 samples although some samples were repeated several times on Nanochip because of no call and low signal results. 21/392 (0.5%) discordant results were tested with the aforementioned technologies: DS confirmed Innolipa results in 18/21 samples and Nanochip results in 1/21, while Devyser and Sequenom did not recognize some mutations not included in their panels. DS was essential for the identification of two different homozygous deletions; although they were not present in Innolipa panels, in 2/21 samples Innolipa indicated a mutation with the warning *no interpretation possible* **Conclusions** In this study the Innolipa assay confirmed its reliability and Nanochip showed that it could become competitive with slight changes to the software.

2605F

Next Generation ABO Genomics: the NHLBI Exome Sequencing Project (ESP). *K. Fox[1], I. Stanaway[1], M.B. Bamshad[1,2], P.L. Auer[3,4], A. Gordon[1], D. Crosslin[1], M. Fornage[5], D. Green[6], S. Rich[7], A.P. Reiner[3,8], D.A. Nickerson[1], J. Johnsen[9,10].* 1) University of Washington, School of Medicine, Department of Genome Sciences, Seattle, WA, USA; 2) University of Washington, School of Medicine, Department of Pediatrics, Seattle, WA, USA; 3) Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA, USA; 4) University of Wisconsin-Milwaukee, Zilber School of Public Health, Milwaukee, WI, USA; 5) The University of Texas Health Science Center at Houston, Institute of Molecular Medicine, Research Center for Human Genetics, Houston, TX, USA; 6) Northwestern University, Feinberg School of Medicine, Chicago, IL, USA; 7) University of Virginia, School of Medicine, Center for Public Health Genomics, Charlottesville, VA, USA; 8) University of Washington, School of Public Health, Epidemiology, Seattle, WA, USA; 9) University of Washington, School of Medicine, Hematology Division, Seattle, WA, USA; 10) Puget Sound Blood Center, Research Institute, Seattle WA, USA.

Cardiovascular Disease (CVD) is the leading cause of death in the United States and worldwide. ABO blood type plays a significant role in determining risk for CVD. ABO phenotypes are the result of genetic variation in the coding portion of the ABO gene on 9q34.2. The influence of ABO subtypes (e.g., the A1 and A2 haplotypes) on CVD risk has yet to be fully explored. Previous studies have genotyped SNPs that tag ABO rather than defining the haplotype sequence necessary for subtype detection. We have developed an accurate ranking method to assign ABO haplotype using exome sequence data. This new method uses a matrix-based scoring approach to compare phased sequencing data to reference A, B, and O haplotypes from the Blood Gene Mutation Data Base (BGMUT). We applied our method to phased exome data from ~5,600 individuals of European and African ancestry derived from the ESP. We observed 100% concordance between predicted and actual phenotype in those ESP participants with serological phenotype data (n=80). We identified common variants known to influence function, including a common exon 6 indel that leads to the O genotype, and another common indel that results in the A2 subtype. We have also identified rare coding variants within ABO (single nucleotide/missense variants (n=18), insertion/deletions (n=4) and structural variation spanning multiple exons (n=1)) that segregate on known haplotype blocks. This method has the potential to improve the specificity of blood typing at both the clinical and research level and to reveal novel associations between CVD and previously unidentified rare ABO haplotypes.

2606W

Genome Analysis of iPS Cells for Regenerative Medicine. *A. Watanabe[1], N. Amano[1], M. Nakamura[1], A. Fukuhara[1], P. Unyanee[1], Y. Tokunaga[1,2], M. Yamaguchi[2], T. Aoi[1], K. Okita[1], K. Takahashi[1], S. Yamanaka[1].* 1) Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan; 2) Amelieff.

Pluripotent stem cells are now suggested as an artificial source of tissues, and consequently it is necessary to be able to guarantee their safety in the human body after transplantation. However, both embryonic stem (ES) cells and induced pluripotent stem (iPS) cells are produced after long-term culture, and thus harbor clone-to-clone variations in their DNA sequences and copy numbers as well as epigenetic profiles. It is therefore important to validate the quality of ES and iPS cells by genomic analyses. We aim to establish the standard method for evaluating iPS cells as a clinical-grade cell source by genome and epigenome analysis. We built the pipeline for both single-nucleotide variations (SNVs) and DNA copy number variations (CNVs), and performed exome and whole genome resequencing and compared between the original somatic cells and established iPS cells. We found not only clone-to-clone variations in iPS cell clones but also the heterogeneity in original somatic cells, indicating that comparison between original cells and established cells is essential for evaluating plulipotent stem cells. One of the iPS clones shows no non-synonymous mutations in exonic region. We also performed DNA methylation analysis by sequence-capture-based deep sequencing and identified novel DNA methylation sites specific for iPS cells. We introduce the strategy of genome and epigenome analysis for evaluating iPS cells for regenerative medicine using iPS cells.

2607T

Utility of targeted inherited disease panels for the diagnosis of rare congenital, potentially genetic syndromic disease. *A. Khromykh[1, 2], D. Thach[1, 2], M. George[1, 2], J. McCutcheon[1, 2], W. Wong[1, 2], R. Baveja[3, 4], R. Iyer[1, 2].* 1) Inova Translational Medicine Institute, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 2) Inova Fairfax Hospital, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 3) Inova Fairfax Children's Hospital, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 4) Fairfax Neonatal Associates, 2730-B Prosperity Avenue, Fairfax, VA 22031.

Specific and accurate diagnostic tests are available for almost 3000 genetic disorders. Nevertheless, for many children with rare, potentially genetic syndromic disease, clear diagnosis remains elusive, and the diagnostic odyssey continues. Recently, next generation sequencing has been used to discover genes involved in rare Mendelian disease and for their diagnosis. However, there are significant barriers for equal and widespread accessibility to these tests, including high cost of testing and reimbursement issues, lack of clearly defined, validated, available algorithms to identify clinically relevant results and the need for multidisciplinary expertise to interpret the results. A short term and partial solution to some of these issues may be to implement a tiered testing modality. This includes 1st tier testing on commercially available, cheaper, targeted disease-gene panels which provide less complex and easier to interpret results using algorithms provided by the manufacturer or developed in-house. Though such testing will not identify new genes for Mendelian disease, it could help in the elucidation of new phenotypes associated with targeted genes present in the panel, reduction of further complex and expensive testing, and elimination of the diagnostic odyssey for some families. In this study, we are analyzing our expanding cohort (currently 40+) of families with a proband afflicted with severe multiple congenital malformations that are not consistent with any known clinical diagnosis using targeted gene panels. The probands are negative by currently available standard genetic testing including arrays (when indicated and performed); most of these have not been analyzed on clinically available next generation panels, and have no clearly relevant family history. Targeted sequencing of the coding exons of >800 mendelian disease genes associated with rare and ultra-rare neuromuscular, cardiovascular, developmental, metabolic and other inherited disease, is performed using the Ion Ampliseq and TruSight inherited disease panels (by Life Technologies and Illumina respectively). We expect that our study will provide useful information regarding the analytical and clinical sensitivity, specificity and utility of these panels that could provide support for the deployment of such tiered testing strategies in the clinical laboratory, using these and/or other enhanced targeted gene panels.

2608F

Towards a medical grade exome: Use of a gold standard to evaluate and enhance exome sequencing for diagnosis. *M. Pratt[1], G. Bartha[1], S. Luo[1], J. Harris[1], S. Garcia[1], G. Chandratillake[1], S. Chervitz[1], R. Chen[1,4], M. Clark[1], M. Snyder[1,2,3], J. West[1], R. Chen[1].* 1) Personalis Inc, Menlo Park, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Stanford Center for Genomics and Personalized Medicine, Stanford University, Stanford, CA; 4) Icahn School of Medicine, Mount Sinai, New York, NY.

Exome sequencing is increasingly utilized in clinical genetics practice to diagnose cases where other genetic testing has been unsuccessful or would be cost-inefficient. However, diagnostic yield estimates from clinical exome testing remain low (~25%). Furthermore, standards for exome sequencing have not converged with respect to content inclusion and minimum coverage requirements to achieve clinical sensitivity and specificity on that content. We have assembled a broad content set over which we assess and optimize exome performance to significantly improve coverage of medically interpretable content with the aim of increasing diagnostic yield. To assess absolute performance of exome sequencing approaches, we have developed internal gold standards based on genomic sequencing. As an accuracy standard, we use a reference comprising more than 3 Tb of next generation sequencing data over multiple technologies on a large pedigree to develop a thorough set of confirmed inherited variants comprising single nucleotide, indel, and structural variants. We also compute coverage standards across samples at nucleotide and exon resolution over whole genomes and exomes. Using these standards, we determined sensitivity and specificity over the interpretable exome at varying mean sequencing depths in order to develop coverage targets by region and content type to ensure analytic validity over the most interpretable regions. We further assessed performance on additional functional subsets, e.g. UTRs, intronic variants and regulatory regions. Through examining coverage and accuracy shortfalls over interpretable content, we developed additional pull-down targets and protocols to augment commercial exomes in regions of low (<20x) or absent coverage with the aim of creating a finished medical-grade exome. We also targeted regulatory regions and other non-exonic regions known to contain disease and pharmacogenetic-associated variation. This approach greatly increases the number of biomedically-relevant genes finished at our medical coverage target (>99% of bases at >20X) from 2700 to >3800. Such increased coverage enables sensitive and specific variant detection including at an additional 4854 loci having previously reported deleterious variants. Assessing accuracy on a larger set of biomedical variants including regulatory region variants and those associated with complex disease, we increase sensitivity from 58% to 90% while reducing the error rate.

2609W

Genetic Research and diagnostic using Fluidigm Integrated Fluidic Circuits (IFCs). *D. Bercovich[1], Y. Plostky[2], S. Allon-Shalev[3].* 1) Human Molecular Genetics & Pharmacogenetics, Migal- Tel Hai, Kiryat Shmona, Galile, Israel; 2) Galil Genetic Analysis - GGA lab, Kazarin, Israel; 3) The Institute for Genetics, Ha'Emek Medical Center, Afula 18101, Israel.

The flexibility of the BioMark Real-Time PCR System, allow us to preform genetic research using different types of nano-fluidic (48.48 or 96.96) chips setup, in the thermal cycle of these chips and image the data in real time for quantity determination of DNA copy numbers (CNV) or mRNA expression in multiple loci locations or genes, and can also to be used as an endpoint image reader for analyzing different allelic genotyping frequencies in a panel of 90 common mutations, in the Israeli populations, for 42 different mono genetic disorders revile there frequency of 15% in screening over 250 indeviduals general Israeli population , and 15-14 common mutations in Breast/ Ovarian and Colon cancers revile there frequency of 3-4% in the Israeli population. A panel of 47 novel Canis (dogs) SNPs was used, for the determination of a phylogenetic tree in 45 different Canis DNA samples. Digital-PCR gene chip were used to determine the number of a human gene which were incorporated in a plant cells for the production of this protein to treat human disease.

2610T

Whole Exome Sequencing is a sensitive cost-effective method of detecting mutations in Osteogenesis Imperfecta and Marfan syndrome.
A.M. McInerney-Leo[1], M. Marshall[1], B. Gardiner[1], P. Coucke[2], B. Loeys[2], J. West[3], M. West[3], B.P. Wordsworth[4], A. Zankl[5], P.J. Leo[1], M.A. Brown[1], E.L. Duncan[1]. 1) University of Queensland Diamantina Institute, Woolloongabba, Queensland, Australia; 2) Department of Medical Genetics, MRB, 1st floor Ghent University, Hospital De Pintelaan 185, B-9000 Ghent, Belgium; 3) School of Medicine, The University of Queensland Coordinator, Cardiovascular Genetic Clinic, The Prince Charles Hospital, Rode Road, Chermside QLD, Australia; 4) Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford England; 5) Bone Dysplasia Research Group, UQ Centre for Clinical Research (UQCCR), The University of Queensland, Building 71/918, Level 3 Royal Brisbane and Women's Hospital, Herston, QLD Australia.

Osteogenesis imperfecta (OI) and Marfan syndrome (MFS) are amongst the commonest Mendelian disorders. Both are normally diagnosed clinically without genetic testing, as conventional sequencing is expensive due to the size and number of potentially causative genes and mutations. However, genetic testing benefits patients, at-risk family members and individuals with borderline phenotypes; improves genetic counselling; and allows critical differential diagnoses. We assessed whether whole exome sequencing (WES) is a sensitive method for mutation detection in OI and MFS. Methods: WES was performed on genomic DNA from 13 participants with OI and 10 participants with MFS, all of whom had known mutations, with massive parallel sequencing of multiplexed samples. SNPs and small indels were called using Genome Analysis Toolkit (GATK) and annotated using ANNOVAR. CREST, exomeCopy and ExomeDepth were used to detect larger deletions. Results were compared with the previous data. The target capture of the currently available exome capture platforms was also compared. Results: All 13 mutations in the OI cohort and 9/10 in the MFS cohort were detected (overall sensitivity = 95.6%) which included non-synonymous SNPs, small indels (<10bp), and a large UTR5/exon 1 deletion. One mutation was not detected by GATK due to strand bias. Capture platforms and analysis programs differed considerably in their ability to detect mutations. Costs for WES were lower than the available commercial sequencing services. Conclusion: WES is both sensitive and cost-effective for mutation detection in patients with OI and MFS. Careful selection of platform and analysis programs is necessary to maximise success.

2611F

Rapid and high mutation detection rate using ion torrent technology and inherited disease panel.
N. Al Tassan[1,2], A. Almostafa[1,2], D. Khalil[1,2], J. Shinwari[1,2], R. Kattan[1,2], A. Alissa[1,2], A. Tahir[1,2], M. Abouelhoda[1,2]. 1) Department of Genetics, King Faisal Specialist Hospital, Riyadh,11211, Saudi Arabia; 2) Saudi Genome Project, King Abdulaziz for Science and Technology, P.O Box 6086, Riyadh 11442, Saudi Arabia.

Ion Torrent semiconductor™ is a chip based sequencing technology (Life technologies, Guilford, CT) that performs sequencing-by-synthesis accompanied by electrochemical detection of base incorporation. In addition to Whole genome and exome sequencing, the technology is being applied in a heavily multiplexed PCR -based approach (Ampliseq) for targeted gene panels. The Ion Ampliseq™ Inherited Disease Panel (IDP) uses >10,000 primer pairs divided into 3 PCR pools to amplify the exons of 328 genes associated with approximately 700 inherited diseases. We applied Ion Torrent sequencing and this gene panel to study 100 samples having a Sanger sequencing validated mutation in one of the genes covered by the panel. 10ng of each sample was used to the Ampliseq PCR, the 3 PCR products were pooled and the library was purified, amplified and quantified. 20 pmoles of each library was used for emulsion PCR, followed by enrichment and sequencing on the medium scale 316 Chip for single samples and the high scale 318Chip for multiplexed samples. Specificity and sensitivity was 100% for all homozygous single point mutations. Several indels required manual review of to identify the mutation which was not identified using the current variant caller. Given consanguinity of the study population and consequent frequency of homoallelic mutations, we also examined exclusion of genes based upon presence of heteroallelic SNPs. Ion Torrent technology and sequencing of highly multiplexed panels offers an efficient screening/diagnostic option for inherited diseases commonplace in consanguineous populations.

2612W

High-throughput molecular genetic analysis in 92 patients with steroid resistant nephrotic syndrome applying Fluidigm Access Array™ technology.
E.A. Otto, V. Vega-Warner, M. Sampson. Dept Ped, 8220A, MSRB III, Univ Michigan, Ann Arbor, MI.

Background: Steroid resistant nephrotic syndrome (SRNS) is a genetically heterogeneous kidney disease characterized by heavy proteinuria, hypoalbuminemia, and edema. In most SRNS cases the kidney function declines over time, resulting in end-stage renal disease which necessitates renal replacement therapy. Mutations in more than 20 genes are implicated in the pathogenesis of SRNS rendering mutational analysis tedious and costly when applying conventional Sanger sequencing. Here, we screened for mutations in 21 established SRNS genes in 92 mostly pediatric patients recruited at the University of Michigan with SRNS. **Methods:** We implementing the PCR-based Fluidigm Access Array™ approach established previously in our lab [1]. We screened for mutations in all coding regions (424 exons) of 21 established monogenic SRNS genes (*NPHS1, NPHS2, PLCE1, COQ2, COQ6, LAMB2, WT1, PDSS2, CD2AP, SMARCAL1, PTPRO, CFH, CUBN, SCARB2, INF2, MYO1E, NEIL1, TRPC6, ACTN4, LMX1B,* and *ITGA3*) in a cohort of 92 mostly sporadic childhood cases with SRNS. We applied the Fluidigm Access Array™ nanofluidic technology followed by barcoding and next-generation sequencing on an Illumina HiSeq2000 platform. Bioinformatics analysis was performed using CLC Genomics Workbench™ software and significant variants were validated by Sanger sequencing. **Results:** We established the molecular diagnosis in 12 out of 92 patients (13%) and detected pathogenic mutations in the genes *NPHS1, NPHS2, WT1, PLCE1, LAMB2,* and *ACTN4*. Moreover, we discovered novel mutations in the genes *NPHS1* (p.Tyr638Cys, p.Gly601Ala), *WT1* (p.Thr225Met, p.Cys393Tyr), *PLCE1* (p.Ser1190Gly), *LAMB2* (p.Ile992Thr, p.Arg697Trp), and *ACTN4* (p.Arg65Gln). **Conclusion:** Fluidigm Access Array high-throughput mutation analysis allows screening of a large patient cohort for multiple genes in parallel at low cost and helps to find mutations in those genes which are rarely implicated in SRNS. [1]Halbritter et al., Hum Genet 2013, Apr 5 [Epub ahead of print].

2613T

Mutaome Profiling and Retrospective Mutaome Profiling Using Archived Bone Marrow or Peripheral Blood Smear in AML.
H. Liu[1], F. Wang[1], W. Teng[1], Y. Wang[1], X. Chen[2], Q. Yin[1], M. Wang[1], P. Zhu[2]. 1) Molecular Medicine Lab., Hebei Yanda Hospital, Sanhe, Hebei, China; 2) Department of Hematology, Peking University First Hospital, Beijing, China.

Background: Acute myeloid leukemia (AML) is a heterogeneous disease with respect to presentation and clinical outcome. In recent years, more and more somatic mutations and their clinical significance were identified in AML. Herein we recommend using the novel word "mutaome' for representing the repertoire of somatic gene mutations in a specific tumor tissue, and aimed to establish a panel of mutation profiling protocol for clinical use in AML. Methods and Cases: Mutation profiling protocol for CEBPA, DNMT3A, FLT3, IDH1, IDH2, KIT, NPM1, PHF6 and TET2 by PCR and Sanger sequencing was established, with the detection sensitivity about 15% to 20%. Bone Marrow (BM) or peripheral blood (PB) was collected form patients with newly diagnosed or relapsed. Archived BM or PB smear on glass slides at the time of newly diagnosed were used for retrospective mutation analysis. Results: 1) Totally 46 archived smear, 50 fresh BM or PB and 1 paraffin-embedded pathological specimen from 91 patients were analyzed. Age ranged from 1 to 77 years old, with the median age of 32 years old. 2) 68 samples carrying at least one mutation, 33 of them each carrying 2 or more mutations. The number of mutated sample for each gene is 18 for CEBPA, 6 DNMT3A, 25 FLT3, 5 IDH1, 6 IDH2, 2KIT, 15 NPM1 , 3 PHF6, and 14 TET2. 3) Both archived smear and relapsed samples were analyzed for 5 cases, 2 of them showing a difference. One patient carrying FLT3 H811Q, FLT3 D839G, IDH1 A51D and CEBPA A295T in archived BM smear, but only carrying FLT3 D839G in relapsed BM and swelling testicular biopsy specimens, another patient carrying FLT3 G822W, TET2 P208H and TET2 R447S in archived BM smear, but only carrying FLT3 G822W in relapsed BM sample. Conclusions: Panel testing for mutations is effective method for detection of AML molecular markers. Archived BM or PB smear can be used for retrospective mutation analysis.

2614F

Mutaome Profiling and Retrospective Mutaome Profiling Using Archived Bone Marrow or Peripheral Blood Smear in B-ALL. *F. Wang[1], H. Liu[1], W. Teng[1], Y. Wang[1], X. Chen[2], L. Guo[1], M. Wang[1], Q. Yin[1], H. Yang[1], P. Zhu[2].* 1) Molecular Medicine Lab., Hebei Yanda Hospital, Sanhe, Hebei, China; 2) Department of Hematology, Peking Univerity First Hospital, Beijing, China.

Background: B-cell acute lymphoblastic leukemia (B-ALL) is a heterogeneous disease with respect to presentation and clinical outcome. In recent years, more and more somatic mutations and their clinical significance were identified in B-ALL, including point mutations and large genomic sequencing deletion mutations. Herein we recommend using the novel word "mutaome" for representing the repertoire of somatic gene mutations in a specific tumor tissue, and aimed to establish a panel of mutation profiling protocol for clinical use in B-ALL. Cases and Methods: Bone Marrow (BM) or peripheral blood (PB) was collected form patients with newly diagnosed or relapsed. Archived BM or PB smear on glass slides at the time of newly diagnosed were used for retrospective mutation analysis. Mutation profiling protocol for JAK1, JAK2 (Exon16, 20), PAX5, PHF6 and TP53 by PCR and Sanger sequencing was established, with the detection sensitivity about 15% to 20%;. IKZF1 exon deletion, PAX5 exon deletion and MLL partial tandem duplication (MLL-PTD) were detected by RT-PCR, electrophoresis and sequencing of the aberrant bands, which didn't applied to archived smear specimens. Results: 1) Totally 24 archived smear, 29 fresh BM or PB samples from 53 patients were analyzed. Age ranged from 2 to 81 years old, with the median age of 14 years old. 2) Totally 15 out of 29 fresh BM or PB samples carrying one or more mutations. The number of mutated cases for each gene is 5 for TP53, 3 PHF6, 1 FLT3, 3 PAX5 exon deletions and 4 IK6 exon deletions. Totally 6 out of 24 archived smear specimens carrying one or more mutation, 2 with TP53 mutations, 3 with PAX5 point mutations, 1 with FLT3 and 1 with PHF6 mutations. Conclusions: Panel testing for mutations is effective method for detection of B-ALL molecular markers. Archived BM or PB smear samples can be used for retrospective mutation analysis.

2615W

Molecular characterization of children with severe autism spectrum disorders. *A.C. Tsai.* OHSU, Oregon, OR.

From April 2012 to May 1013, 10 pts with severe autism spectrum disorder ascertained in the autism and general genetic clinic in the CDRC/DCH were characterized by CMA and molecular panels. the clincial features of the pateitns in his cohort fall in the Angelman/Rett syndrome-like phenotype: patient either present with severe seizure at birth, severe hypotonic, being non-verbal after 7 years of age, some with history of sleep apnea, unique hand mannerism, ataxia, fascination about the water or inappropriate laughter. Method: All patients receive Rett/Angelman-like panel followed by microarray if negative. The panel includes CDKL5 , SLC9A6, TCF4 and FOXG1, UBE3A and MECP2. When the results of the above tests are normal, additional testing based on clinical finding were added including FLNA, RAI1 and seizure panels, as well as SNP plus 400K CMA. Results: 7 Patient were found to have remarkable findings which include c.409T>C heterozygousmutation of UBE3A, MECP2 (Zappella variant), duplication of MECP2, FOXG1 1-bp deletion, PQBP1 1-bp insertion, and 2 with 1p36 deletion, sized from 2.0-2.4 Mb and a 13Kb deletion of 14q13.2 with iatrogenic deletion of the RALGAPA1 (aka TULIP1; aka GARNL1) Discussion: the significance of c.409T>C heterozygous mutation of UBE3A is still unknown ; this girl is non-verbal but has significant signing, the rest of the testing were normal MOC carries this mutation but cannot determine the parental origin as maternal grandfather is not available for testing. The TULIP1 deletion is inherited from the father who has autistic features but clinically is less involved than the proband. This child also received a seizure panel which were negative. The child with PQBP1 mutation was originally thought to have angelman syndrome but detected by the seiszure panel which include the AS/Rett panel. the diagnosis was subsequently changed. Conclusion:This small study highlight the clinical utility of panel assay which is perfect to characterize conditions with a spectrum of presentation in a gene and overlapping phenotype among the genes on the pathway. This study also demonstrate the importance of proper pre-testing genetic counseling, as the test result might detect unexpected genetic syndromes.

2616T

Establishment of a next-generation sequencing protocol for genetic testing of tuberous sclerosis complex. *P. Chen[1,2,3,4].* 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan, Taiwan; 2) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan.

Tuberous sclerosis complex (TSC) is a characteristic disease drawing high clinical and research attention. It is an autosomal dominant disorder characterized by tumors in different body organs, including brain, kidney, liver, skin, lung, heart and eye. TSC is caused by dysfunction of the critical biological pathways related to mammalian target of rapamycin (mTOR). About 400 different causative variants have been identified in the two known genes (*TSC1* at chromosome 9q34 and *TSC2* at chromosome 16p13). However, routine genetic testing for TSC patients using traditional Sanger sequencing method is too expensive and labor intensive because there are a total of 64 exons of these 2 genes. We decided to apply the NGS technology for the cost-effective diagnosis of TSC. We have focused on the whole genomic regions containing all exons, all introns and 10 Kb upstream and 10 Kb downstream of the *TSC1* (hg19, chr9:135761735-135830020, 68285 bp) and *TSC2* (hg19, chr16:2087990-2143712, 55722 bp) genes. A Roche NimbleGen customized capture library was designed for enrichment of the targeted sequences. A reasonably good coverage (97% of the intended regions; and 100% of the exons) can be achieved. The enriched libraries were then pair-end sequenced on the Illumina HiSeq2000 Genome Analyzer. The whole bioinformatics analysis was processed in a computer cluster with more than 50 nodes at the High Performance Computing (HPC) center at National Taiwan University. We used BWA for initial read mapping, GATK pipeline for realignment and SNP/indel calling. Aggressive filtering was applied to further reduce false positive results. Biological significance of the genetic variants was then predicted using bioinformatics tools, mostly SIFT and PolyPhen2. The integrative genomics viewer (IGV) was used for visualization of the results. Among the 61 TSC patients we tested, 50 (82%) of them could be assigned causative variants. We demonstrated that our method could detect not only simple variants (such as single nucleotide substitution and small indels) but also difficult genetic variants (such as a big deletion of ~15 Kb size, and mosaicism). This serves as a single stop for TSC genotyping and waive the necessity of both Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA). We consider this new protocol a cost-effective diagnostic test for TSC, and may be applied to other diseases in the future.

2617F

Magnitude of effect of reported common allele risk from clinical genome sequencing and conventional decision factors in therapeutic equipoise: initial comparison in The MedSeq Project. *C.A. MacRae[1, 2], S. Kong[3, 8], J. Krier[4, 5], I. Leshchiner[1, 5, 7], I.H. Lee[3, 8], H.M. Maclaughlin[4, 7], W.J. Lane[4, 5], D. Metterville[7], A.L. Hernandez[7], BH. Funke[4, 7, 8], M. Lebo[4, 5, 7], P. Kraft[9], I.S. Kohane[3, 6], R.C. Green[1, 2], H. Rehm[4, 7] for the MedSeq Project.* 1) Medicine, Brigham and Women's Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; 3) Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA; 4) Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA; 5) Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; 6) Department of Pediatrics, Childrens Hospital Boston, Boston, Massachusetts, USA; 7) Laboratory for Molecular Medicine, Partners Center for Personalized Genetic Medicine, Cambridge, Massachusetts, USA; 8) Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts, USA; 9) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA.

Genome wide association studies have defined the common risk alleles for numerous human traits, but to date these findings have not impacted clinical practice. The MedSeq Project is a randomized clinical trial that aims to develop standards and procedures for the evaluation and reporting of genome sequencing data and will directly assess the impact of integrating genome sequencing into clinical medicine. MedSeq participants in the sequencing arm of the study will receive both a General Genome Report and a Cardiac Risk Report (CRR), the latter featuring a summary of variants in a targeted panel of 102 monogenic cardiac disease-associated genes, a polygenic predicted fasting lipid profile, and common allele risk information for eight core cardiovascular phenotypes. To define a rational basis for including common alleles in the CRR, we estimated the effect sizes for rigorously validated common risk alleles across all published loci for all phenotypes directly or indirectly related to cardiovascular disease and compared these with the estimated effect sizes of non-genetic risk factors commonly used in clinical decision-making, but which have no randomized clinical trial support. For the eight phenotypes in the current CRR, (abdominal aortic aneurysm, atrial fibrillation, coronary heart disease, type 2 diabetes, hypertension, obesity, platelet aggregation and QT prolongation) the genetic risk, even from a limited number of small effect loci, was comparable in magnitude to that of non-genetic factors used in routine therapeutic decision making in situations of equipoise. We will present a summary of the quantitative approach taken in the evaluation of traits for inclusion in the CRR, discuss the role of small effect-size risk in clinical decision making and describe the evaluation of the effects of CRR disclosure on additional testing and equipoise in both primary care and cardiology clinics within the MedSeq study. Our experience highlights important parallels between the utilization of traditional non-genetic evidence in clinical practice and the practical implementation of complex trait risk results from clinical genome sequencing.

2618W

Translating results of the Australian and New Zealand Registry of Advanced Glaucoma (ANZRAG) into clinical practice. *E. Souzeau[1], K.P. Burdon[1], B. Ridge[1], A. Dubowsky[2], J.E. Craig[1].* 1) Dept of Ophthalmology, Flinders University, Flinders Medical Centre, Adelaide, SA, Australia; 2) SA Pathology, Flinders Medical Centre, Adelaide, SA, Australia.

Purpose: Glaucoma is the leading cause of irreversible blindness worldwide, and is a complex disorder with genetics playing a crucial role. Adequate monitoring and interventions at the early stages can prevent glaucoma blindness. Genetic testing is a promising strategy to identify at risk individuals and reduce the impact of glaucoma blindness through prevention. The Australian and New Zealand Registry of Advanced Glaucoma (ANZRAG) has established a biobank of severe glaucoma cases to identify novel genetic risk factors for the worst glaucoma outcomes, and to establish genetic testing protocols for known glaucoma genes. Methods: Advanced Open-Angle Glaucoma (OAG) cases defined by central visual field loss or severe peripheral vision loss were recruited. Secondary glaucoma cases were also recruited regardless of severity. Cases with advanced glaucoma and Primary Congenital Glaucoma (PCG) were tested in an accredited pathology laboratory by direct DNA sequencing for known glaucoma genes, *Myocilin* and *CYP1B1* respectively, and results were provided back to the participants with the provision of genetic counseling. Cascade genetic testing was made available for relatives of participants with pathogenic mutations in these genes when clinically relevant. Results: 1570 participants with severe OAG and 960 with other glaucoma subtypes have been recruited. Using genome-wide association studies, new glaucoma risk alleles were identified: *TMCO1* in advanced primary OAG cases, and *CDKN2B-AS1* associated with normal-tension glaucoma. *Myocilin* mutations were identified in 4.2% of advanced primary OAG individuals, compared with 1.6% in non-advanced primary OAG. These mutations were found in individuals with younger age at diagnosis and high-tension glaucoma. *CYP1B1* mutations were found in 18% of PCG cases. Cascade genetic testing has identified highly penetrant *Myocilin* mutations in 35 as yet asymptomatic individuals. The experience of tested asymptomatic individuals has been evaluated to help health professionals in providing better support to patients. Outcome: The ANZRAG provides a comprehensive approach to reducing glaucoma blindness by unraveling the genetic complexity of glaucoma, providing accredited genetic testing and genetic counseling, and translating research findings in better clinical care.

2619T

ClinVar: Improving Access to Clinically Relevant Variants for the Research and Clinical Genomics Communities. *M.J. Landrum, J. Lee, G. Riley, R. Tully, S. Chitipiralla, M. Halavi, D. Hoffman, J.B. Holmes, W. Jang, K. Katz, M. Ovetsky, A. Sethi, R. Villamarin, D.M. Church, W.S. Rubinstein, D.R. Maglott.* National Center for Biotechnology Information, NLM, NIH, Bethesda, MD.

Advances in technology for genomic testing, particularly whole exome and genome sequencing, are identifying thousands of new variants. However, determining the clinical relevance of variants can be challenging, especially for rare variants. To increase the access to such data, ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) maintains a freely available, public repository for the relationships between variants and phenotypes along with supporting evidence. The database can be used interactively or incorporated into variant analysis pipelines. The public release of the database in April 2013 largely included curated data from OMIM and GeneReviews, as well as submissions from some LSDBs, testing labs, and ClinSeq. Within the first two months, submitters also included additional testing labs, research groups, and expert panels and professional societies. Each ClinVar record includes a variant and an assertion of clinical significance for the variant in the context of an observed phenotype, with a review status indicating the level of support. Submitters retain ownership of submitted data, and each record is assigned an accession and version number to allow updates by the submitter as the variant interpretation is refined over time. For submitters who maintain their own website for variants, such as LSDBs, ClinVar links to the submitter's site for each submitted variant, allowing users who start at ClinVar an awareness of the LSDB's curated variants and access to more information on the variant that may be available at the LSDB. ClinVar supports standardization of terms, such as those for variant nomenclature, phenotypes, and pathogenicity, to lessen data ambiguity and aid comparison of information from multiple sources. ClinVar also provides related variant data, such as HGVS expressions mapped across molecule types and allele frequencies. As part of the submission process, ClinVar provides feedback to submitters. This feedback includes HGVS expressions that are invalid and whether a submission has a conflict in clinical significance with an existing record for the same variant and phenotype, alerting submitters to records that may warrant further curation. When multiple submitters provide an assertion for the same variant/phenotype pair, ClinVar collects the data into an aggregate record, allowing clinical labs to more easily identify high-confidence variant associations and to organize curation efforts around variants with conflicting interpretations.

2620F

Molecular diagnostic approach for limb-girdle muscular dystrophy using both multi-gene panel sequencing and Sanger sequencing. *H. Park[1], S. Lee[1], S.H. Seo[1], S. Park[2], S.I. Cho[1], M.W. Seong[1], S.S. Park[1].* 1) Department of Laboratory Medicine, College of Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of Laboratory Medicine, College of Medicine, Konkuk University Medical center, Seoul, Korea.

Introduction: Limb-girdle muscular dystrophy is genetically heterogeneous disease, with clinical involvement typically limited to skeletal muscle. Because more than 20 genes identified, it is not easy to find the causative mutations by conventional PCR and direct sequencing for individual patient. Here we present a molecular diagnostic strategy combing multi-gene panel sequencing and Sanger sequencing. Methods: Seventeen patients who were suspected of limb-girdle muscular dystrophy were screened for sixteen genes using TruSeq Custom Enrichment Kit (Illumina) and MiSeq (Illumina). All previously reported mutations and probable pathogenic variants including novel nonsynonymous ones were confirmed by Sanger sequencing. Also all low-coverage regions with coverage depth under 10X were resequenced by Sanger sequencing. Results: Among 17 patients, five ones were molecularly confirmed as limb-girdle muscular dystrophy. Two patients had a *LMNA* gene mutation (c.1357C>T and c.1366A>C), which was inherited autosomal dominant manner. Another two patients had compound heterozygous mutations in *DYSF* gene (c.[2248C>T(;)5668-7G>A] and c.[2494C>T(;)4200delC]) and the other one patient had two mutations in *CAPN3* gene (c.[439C>T(;)1076C>T]) which were inherited autosomal recessive manner. Conclusion: Our multi-gene panel detected pathogenic mutation(s) in five limb-girdle muscular dystrophy patients. Multi-gene panel sequencing using next generation sequencing technology can be a cost efficient and fast method for diagnosis of genetically heterogeneous disease like limb-girdle muscular dystrophy.

2621W

Coding mutations and variations in the 3'UTR of *CYP21A2* gene in heterozygous females associate with hyperandrogenism. *V. Neocleous[1], C. Shammas[1], AAP. Phedonos[1], M. Picolos[2], TC. Kyriakides[3], M. Toumba[4], N. Skordis[5], LA. Phylactou[1].* 1) Molecular Genetics, Function & Therapy, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) Alithias Endocrinology Center, Nicosia, Cyprus; 3) Department of Epidemiology & Public Health, Yale University, USA; 4) Iasis Hospital, Paphos, Cyprus; 5) Paediatric Endocrine Unit, Makarios III Hospital, Nicosia, Cyprus.

Congenital adrenal hyperplasia (CAH) is a common autosomal recessive disorder primarily caused by mutations in the *CYP21A2* gene. Heterozygosity for *CYP21A2* mutations in females increases their risk of clinically manifesting hyperandrogenism. The present study was designed to seek evidence on the association between the mutations in the *CYP21A2* gene and the biochemical/clinical findings on heterozygous children, adolescents and women with hyperandrogenemia. Moreover, the implication of variants in the 3'UTR region of the *CYP21A2* gene was investigated. The hormonal response to ACTH was evaluated in heterozygous females with clinical signs of hyperandrogenism along with direct DNA sequencing and MLPA analysis for mutations in the *CYP21A2* gene. The suspicion of heterozygote state was based on the median plasma 17-OHP before and 60 minutes after ACTH stimulation. The most frequent mutations among the 66 carriers were the mild p.V281L (53.0%), followed by p.Q318stop (18.2%), p.P482S (10.6%), p.V304M (6.1%), p.P453S (6.1%), p.A391T (1.5%), large deletion/conversion exons 1-4 (1.5%), large deletion/conversion exons 6-8 and 8bpdelE3 (1.5%). Higher values of stimulated 17-OHP levels were found in the carriers of the p.V281L mutation, compared with carriers of other mutations (mean = 22.1 nmol/L vs 15.9 nmol/L). The haplotype of the *52 C>T, *440 C>T, *443 T>C in the 3'UTR was found *in cis* in all heterozygous females with p.V281L. In a similar fashion the haplotype *12 C>T, *52 C>T was found *in cis* in all heterozygous females with the p.Q318stop. Conclusions: Females with heterozygous *CYP21A2* mutations may develop hyperandrogenism. Therefore, systematic evaluation of 17-OHP values in combination with the molecular testing of *CYP21A2* gene may be beneficial. Finally, the identification of variants in the 3'UTR of the *CYP21A2* gene in combination with the heterozygous mutation may be associated with the mild form of the disease and reveal the importance of analyzing the *CYP21A2* untranslated regions to better characterize and treat this category of patients.

2622T

Detection of deleted D4Z4 locus in turkish patients with facioscapulohumeral muscular dystrophy. *S. Berker Karauzum[1], OB. Sahan[1], H. Uysal[2].* 1) Med Biol& Gen, Akdeniz University Faculty Med, Antalya, Turkey; 2) Dept. of Neurology Akdeniz University Faculty Med, Antalya, Turkey.

Facioscapulohumeral Muscular Dystrophy(FSHD) which is characterized by progressive weakness of muscles in the face, schoulder girdle and upper arms is the third must common muscular disease. FSHD is caused by deletions of the D4Z4 repeats in the 4q35 region. In this autosomal dominantly inherited disease, more than 95% patents have only 1-10 repeats instead of 11-100 repeats observed in healthy controls. In this study, 24 individuals (11 male, 13 female) from 6 unrelated Turkish families with FSHD were handled after the preclinical diagnosis by the Akdeniz university medical faculty, department of neurology. Three healthy individuals were studied as a control group. In order to show the deletion of D4Z4 tandem repeats at the q35 locus on chromosome 4, Southern blot method was performed. While D4Z4 repeats of control group was found to be in the normal range(more than 11 repeats), 20 FSHD patients' repeat numbers were observed to be the under the normal range(less than 11 repeats) as concordant to clinical findings. Our results show that, Southern blot was observed as a suitable method for determining the D4Z4 repeat deletions on chromosome 4 and also, because of in all the affected cases displayed D4Z4 repeat deletions, this analysis could be performed in patients with myopathies similiar to FSHD, especially for prenatal diagnosis and genetic councelling.

2623F

Investigating the carrier screening potential of the MPL c.79+2T>A transversion, a known cause of congenital amegakaryocytic thrombocytopenia for individuals of AJ descent. *S.R. Birkeland, J. Sugalski, C. Holland, J. Stoerker, J. Buis.* Progenity, Ann Arbor, MI.

Mutations in the thrombopoietin receptor c-MPL cause a rare inherited disease known as congenital amegakaryocytic thrombocytopenia (CAMT). This disease presents as severe thrombocytopenia at birth, reduced megakaryocytes and progression to bone marrow failure. The only known effective treatment for this is a bone marrow transplant, which presents significant risk. Recently, a founder mutation in the c-MPL gene known as the c.79+2T>A transversion has been shown in the Ashkenazi Jewish population. Carrier screening for other disease alleles within this community has met with great success, reducing the numbers of new disease-affected individuals across a number of syndromes. Here, we present data from a blinded and random screening of individuals of AJ descent and compare this with a cohort of non-AJ individuals. In concordance with the previously studies, we find a significant carrier rate for the c.79+2T>A mutation amongst the AJ population. Our preliminary findings indicate that out of the first 358 tested, 6 are carriers for c.79+2T>A which have been confirmed via Sanger sequencing. We will plan on screening a total cohort of ~1000 individuals. The ability to screen for the c.79+2T>A mutation has been included in our validated AJPNxt carrier screen. Our data presented here along with previous studies and the severity of CAMT supports the value of including testing for this mutation in future AJ population screening.

2624W

Screening of gene mutation in hyperphenylalaninemia using Ion Torrent sequencing. *Y. Cao, F. Song, Y. Qu, J. Bai, Y. Jin, H. Wang.* Department of Medical Genetics, Capital Institute of Pediatrics, Beijing, China.

Hyperphenylalaninemia (HPA) is a very common autosomal recessive genetic disease. It has been verified that the metabolic pathway of phenylalanine involves five pathogenic genes: PAH, GCH1, PTPS, QDPR, and PCBD1. Rapid and accurate genetic diagnosis is very crucial for a clear diagnosis of disease types, choosing the right treatments in a timely manner, and for genetic counseling and prenatal diagnosis. To evaluate the feasibility of gene screening in children with HPA using Ion Torrent Personal Genome Machine (PGM), the CDS and UTR regions of PAH, GCH1, PTPS, QDPR, and PCBD1 in 3 patients with HPA and one healthy control were amplified using Ion AmpliseqTM PCR and then sequenced by PGM. Meanwhile, a set of mutant sample mixture representing common mutations in patients with HPA in China was used as a positive control to determine the accuracy of the PGM. All of known mutation sites were correctly identified in the positive control. In addition, we detected 22 variations in the patients and the healthy control. Compared with database and verified by the Sanger sequencing method, it was confirmed that 6 were pathogenic mutations, 18 were polymorphisms and 4 were false-positive calls. Based on our study, the PGM sequencing might be suitable for screening gene mutation in HPA via metabolic pathways, which could meet the medical need for individualized diagnosis and treatment. However, mutations suspected for false-positive calls should be discriminated by other method.

2625T

Comparison of Seven Commercial DNA Extraction Kits for the Isolation of Listeria monocytogenes DNA from Whole Blood Samples. *M. El-Mogy[1], M.AK Abdalla[2], L. Graziano[3], T.A Haj-Ahmad[4], Y. Haj-Ahmad[4,5].* 1) Molecular Biology Department, National Research Center, Dokki, Cairo, Egypt; 2) Department of Biochemistry, Faculty of Science, Alexandria University, Egypt; 3) Department of Biomedical Sciences, University of Guelph, Guelph, Ontario, Canada; 4) Brock University, 500 Glenridge Avenue, St Catharines, ON, L2S 3A1; 5) Norgen Biotek Corp., 3430 Schmon Parkway, Thorold, Ontario, Canada, L2V 4Y6.

Listeriosis is an important public health concern in North America; it is mainly transferred by food contaminated with the bacterium Listeria monocytogenes. The majority of susceptible individuals are pregnant women, newborns, older adults and those who are immunocompromised, with a mortality rate of about 30% among infected individuals. In this study we compared the recovery as well as the limit of detection of L. monocytogenes DNA from whole blood using seven commercial blood DNA isolation kits (QiaAmp DNA Blood Mini Kit, Norgen Genomic DNA Isolation Mini Kit, MoBio UltraClean DNA BloodSpin Kit, Macherey-Nagel NucleoSpin Blood Kit, Gentra Puregene Blood Kit, Norgen Non-enzymatic DNA Blood Kit and Norgen Enzymatic DNA Blood Kit). Human whole blood samples were spiked with known copy numbers of L. monocytogenes. Total genomic DNA was isolated and compared both qualitatively and quantitatively among the various kits; the concentration as well as OD260/280 and OD260/230 were measured spectrophotometerically. The isolated DNA was used in a real-time PCR reaction using specific primers. We determined the limit of detection for every kit as well as the linearity of pathogen recovery at different concentrations. Also we investigated the presence or absence of contamination and ease of handling of the used methods. Our results show that the column-based methods have higher consistency than alcohol precipitation methods with higher yield and purity from procedures utilizing proteinase K.

2626F

A comprehensive low-cost diagnostic test for hundreds of inherited conditions. *J. Garcia, J. Sorenson, M. Sommargren, J. Westbrook, E. Hare, Y. Kobayashi, M. Anderson, J. Major, R. Hart, K. Jacobs, E. Oliveras, J. Hagenkord, S. Lincoln, M. Cargill, R. Scott.* InVitae, San Francisco, CA.

Historically, diagnostic tests using DNA sequencing have only been offered for a limited set of genes to patients with specific clinical indications. The high cost of de facto standard assays (Sanger sequencing, MLPA, etc.), and more importantly, the high cost and challenges in clinical data interpretation have been cited among the reasons for this. Thus, many genetic diseases and clinically important genetic conditions often go undiagnosed. We have developed an in-house infrastructure for NGS-based diagnostic assay development, validation, and operation in a CLIA environment. To date, we have conducted a thorough scientific review of the literature for over 500 genes and their associated conditions, storing validated gene sequences, transcripts, risk models, and over 32,000 clinically characterized variants in a database used both to optimize assay design and to help interpret results. We have a hybrid calling and data QC pipeline employing GATK, Freebayes, and custom algorithms for different variant classes. Preliminary reports for known and novel SNVs, indels, and CNVs are automatically generated for review, and a team of medical specialists then classifies variants according to ACMG guidelines given the patient's indication and signs-out finalized clinical reports. To date, validation has been performed on over 200 clinical, Get-RM and HapMap samples. 100% of clinical SNV and indel genotypes reported agreed with the results of established, traditional diagnostic assays on those samples. Importantly, pathogenicity assessments in the clinical reports agreed as well. In collaboration with other labs and patient advocates, we have launched an effort to expand the publicly available set of unpublished clinical variants that we believe will be critical in diagnostic settings. Most importantly, we believe these processes are highly scalable, allowing the assay to grow to report on the vast majority of genetic conditions with high accuracy. For the HapMap samples specifically we performed a comparison against a combination of 1000 Genomes Project and Complete Genomics data. For all 2172 coding sequence SNVs in the 211 genes assayed in the HapMap samples, 99.7% sensitivity and greater than 99.99% specificity was demonstrated. For all 66 coding indels, 98.3% sensitivity and greater than 99.99% specificity was demonstrated.

2627W

A Case in Point: When is Extended Genotyping of AAT (SERPINA1) Indicated? *S. Kwong[1, 2], J. Stoller[2], F. Mularo[2], F. Lacbawan[2].* 1) Case Western Reserve University School of Medicine, Cleveland, OH; 2) Cleveland Clinic, Cleveland, OH.

Alpha-1 antitrypsin deficiency (AATD) (OMIM #613490) affects 1 in 2,000 to 1 in 5,000 individuals, but is clinically under-recognized. AATD is associated with mutations of SERPINA1 (or alpha-1 antitrypsin), a serine protein-ase inhibitor that functions as a major inhibitor of neutrophil elastase. AATD typically manifests with liver disease and/or early-onset emphysema. Patients who present with early-onset or unexplained (without typical risk factors) emphysema, unexplained liver disease, or more rare symptoms such as necrotizing panniculitis or vasculitis have alpha 1-antitrypsin serum levels measured. Abnormally low levels are followed up with S/Z genotyping or phenotyping with thin-layer isoelectrofocusing (IEF), the gold standard, to look for aberrant SERPINA1 alleles. The most well-understood and most common alleles are M (wild-type); S and Z (dysfunctional mutant alleles). However, the Q0 (or null alleles), which have nonsense or frameshift mutations leading to absent or truncated mRNA transcripts or proteins that are incompletely characterized. Null alleles are typically not detected with conventional gold-standard methods and may require direct DNA sequencing. Even then, the function of only a few null alleles has been well described, thus making predictions about their phenotypic manifestations difficult. The current report presents a case of an asymptomatic 48-year old female patient who had only chronically elevated AST and was found to have an AAT enzyme level of 54 mg/dL (normal range: 100-220 mg/dL). DNA sequencing showed heterozygous M1V (c.710 T>C) and S (c.863 A>T) alleles with heterozygous Q0_cairo (c.847A>T). Extended genotyping by subcloning demonstrates a trans relationship between the alleles. This explains how the compound heterozygote state produces the unexpected phenotype. To our knowledge, the Q0_cairo and S combination has never been described to date. Neither allele has been associated with liver manifestations, only with lung manifestations. This case demonstrates that direct DNA sequencing and extended genotyping are necessary to help uncover rare alleles associated with AATD and to offer accurate genetic counseling information to the proband and family members. This newly described allelic combination extends the approximately 120 alleles described to date and permits better phenotypic characterization of individuals with unusual variants of AATD.

2628T

The NIH Genetic Testing Registry: 2013 status report on genetic testing. *W.S. Rubinstein[1], B.L. Kattman[1], A.J. Malheiro[1], J.M. Lee[1], D.R. Maglott[1], V. Hem[1], M. Ovetsky[1], G. Song[1], C. Wallin[1], K.S. Katz[1], R. Villamarin-Salomon[1], C. Fomous[2], J.M. Ostell[1].* 1) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 2) Office of Biotechnology Activities, National Institutes of Health.

The NIH Genetic Testing Registry (GTR; http://www.ncbi.nlm.nih.gov/gtr/) is a free, centralized, international registry of comprehensive genetic test information covering clinical and research tests for Mendelian disorders and drug responses. As of early June 2013, submitters in 37 countries have voluntarily provided detailed information on over 6400 registered tests for over 2700 conditions and 5600 genes. Active submitter participation coupled with GTR's rich data structure enables many questions to be answered about the current status of genetic testing. Methods: Extensive stakeholder input and consultation with advisory boards determined concepts to be represented and ongoing feedback is elicited to enhance utility of GTR. All registered tests provide details about methodology, tested conditions, and targets (genes and variants, chromosomal or mitochondrial regions, proteins or analytes). Clinical tests have fields for purpose, test performance characteristics, target population, ordering information, AMA molecular pathology CPT codes, proficiency testing, and regulatory information. Research tests provide a study description, eligibility criteria, consent form, and enrollment information. GTR staff augment content with practice guidelines, position statements and recommendations from expert sources and the website integrates information from authoritative sources. Results: As of June 2013, a total of 352 laboratories have registered 6465 tests employing molecular (N=6146; 95%), cytogenetic (N=154; 2.4%) and biochemical (N=287; 4.4%) methods (combinations included). Next-generation sequencing is a component of 6.0% of molecular tests. Eighteen registered labs offer whole genome or whole exome sequencing services. All clinical tests have analytical validity statements. Submitted data include statements with supporting citations [in brackets] about clinical validity (for 686 tests [281]), target population (for 1005 tests [305]) and clinical utility (632 total statements [329]). Among tests registered by US labs, 0.2% report FDA status cleared/approved. A total of 158 practice guidelines, position statements and recommendations pertaining to 289 conditions have been assembled by GTR staff. These results will be updated at the time of presentation. Conclusions: GTR enables informed selection of tests by clinicians, provides a snapshot of genetic testing, and helps identify knowledge gaps of interest to professional societies and regulatory agencies.

2629F

Diagnostic exome sequencing beneficial among patients with a prior diagnosis. *L.M. Shahmirzadi, K.D Gonzalez, S. Tang, E. Chao, S. Gandomi, B. Trippin, S. Nahas, W. Zeng.* Ambry Genetics, Aliso Viejo, CA.

Clinical diagnostic exome sequencing (DES) provides data on all the coding exons of the genome and is currently indicated when prior tests have been negative, when the phenotype fits with the clinical spectrum of multiple genes, and/or when the phenotype is not consistent with any known clinical picture. DES has not only identified pathogenic alterations in genes not previously associated with the disease, but it has also revealed broader phenotypes not previously considered as part of the clinical spectrum within genes with well-established disease-associations. Even in cases where the patient has been provided with a clinical diagnosis or appropriate differential diagnosis, DES may still be pursued to identify the underlying molecular etiology when a single gene test was not available or when DES is more cost- and time-effective than the sequential gene-by-gene approach. A retrospective analysis of the first 200 reported patients undergoing DES at one laboratory revealed that prior to testing, 10% had a clinical diagnosis. Among the 20 clinical diagnoses made prior to DES testing, 3 (15%) were associated with 1-3 genes, while multiple genes could be implicated in the rest (85%). DES provided a definitive molecular diagnosis (including characterized and novel genes) in 53% of patients with a prior clinical diagnosis. The detection rate was lower (33%) in cases those with only 1-3 genes associated with the diagnosis, than those with multiple suspected genes (75%), thought to be attributed to the low overall clinical detection rate of the suspected gene(s). Gene coverage was greater than 95% for all of the cases associated with 1-3 genes. These results highlight the clinical utility of DES, even among patients with a prior diagnosis, as it may reveal the underlying molecular etiology of the disease of interest when multiple genes may be involved.

2630W

A retrospective analysis of discrepancies between genotypes and phenotypes on next generation sequencing colon cancer panels (ColoNext NGS): Implications for clinical diagnosis. *S. Tandy, A. Stuenkel, T. Pesaran, H. LaDuca, E. Chen, S. Keiles, V. Speare, C. Radford, W. Zeng.* Ambry Genetics, 15 Argonaut, Aliso Viejo , CA 92656.

The launch of Ambry's cancer NGS panels in early 2012 has led to unexpected genotype-phenotype discrepancies among patient results. We reviewed 38 ColoNext NGS cancer panels that had mutations detected in genes associated with cancer syndromes with established clinical criteria. Of those, a significant percentage of patients with pathogenic mutations did not meet the correlating clinical criteria based on their personal and family history: 50% (3/6) with biallelic MUTYH mutations, 50% (2/4) with SMAD4 mutations, 50% (1/2) with PTEN mutations, 14% (1/7) with APC mutations, 40% (2/5) with MSH6 mutations, 28.5% (2/7) with MSH2 mutations, and 57% (4/7) with PMS2 mutations. Here we discuss specifics in twelve of these cases. Biallelic MUTYH gene mutations were detected in three unrelated individuals with early onset colon cancer and less than 20 adenomatous colon polyps. A SMAD4 mutation was detected in a proband with gastric cancer at age 35 who was later found to have 20-99 colon/GI polyps of varying pathology types, none of which were juvenile, by age 50. Another SMAD4 mutation was found in a proband diagnosed with colon cancer at age 30 with normal tumor MSI and IHC testing. A PTEN mutation was identified in a 65 year old individual with over 100 adenomatous colon polyps and no family history of cancer. An individual with 2-5 adenomatous polyps and colon cancer at age 39 was found to carry an APC mutation. A 16 year old with colon cancer and a reported heavy polyp load was found to carry two MSH6 mutations. A 36 year old unaffected individual with a MSH2 mutation was tested due to a paternal family history of sarcoma, endometrial, and late onset colon cancers. An intronic MSH2 mutation was found in a patient with endometrial cancer at age 48 but whose family history was not suggestive of Lynch Syndrome. A single exon PMS2 deletion was found in a proband with 2-5 colon polyps, colon cancer at age 49, and normal MSI and IHC tumor testing. A proband meeting Cowden syndrome clinical criteria was found to carry a PMS2 mutation. Since these results reveal that genotype-phenotype discrepancies clearly exist among individuals carrying mutations in well-known cancer syndrome genes, we propose continued evaluation of the clinical features associated with each condition. We encourage the continued reporting of additional cases with uncharacteristic correlations in order to further expand upon the clinical criteria utilized in current diagnostic processes.

2631T

Pre-publication sharing of exome/genome variant and phenotype data to resolve rare disease. *J. Den Dunnen, I.F.A.C. Fokkema, M. Vermaat, J.F.J. Laros, M. Kriek, P.E.M. Taschner, G.W.E. Santen.* Center for Human and Clinical Genetics, Leiden University Medical Center, Netherlands.

We have built a system facilitating sharing of unpublished data from whole exome (WES) and whole genome sequencing (WGS) studies. The system should help to speed up gene discovery by assisting researchers to find the critical 'second case'. The main hurdle in current studies using WES/WGS is to find sufficient evidence to prove causality. Roughly studies give a yield of 1/3 proven causality, 1/3 likely causality and 1/3 unresolved cases. An obvious way to improve overall yield would be reaching out to colleagues world-wide to find additional positive or negative evidence for the candidate variants remaining. An important obstacle here is, for several reasons, the desire and demand to publish in peer-reviewed journals which works against pre-publication data sharing. The approach is built around the gene variant databases (LSDBs) using version 3 of the LOVD platform (Leiden Open-source Variation Database, http://www.LOVD.nl) to collect and share information about genes, variants and phenotypes (diseases) and facilitate the analysis of exome and genome sequence data. Within the databases curated by us we have implemented a so-called **VIP-status** for both variants and phenotypes demanding specific attention. People can submit both phenotype descriptions and/or gene variants with the request to assign these a **VIP status**. **VIP-phenotypes** are those for which WES/WGS studies were performed but for which insufficient evidence was gathered to prove causality with variants in a specific gene. **VIP-variants** are those remaining after stringent filtering that could not be linked to nor excluded as causative in WES/WGS. The submitter decides whether the submission is named or anonymous. Others studying **VIP-phenotypes** or identifying **VIP-variants** are invited to contact the submitter to collaborate. For anonymous submissions the LOVD-curator will act as an intermediate to bring researchers into contact. We believe this option is effective regarding sharing data while at the same time protecting peoples' interest to ultimately publish their findings.

The Human Variome Project has granted LOVD the recommended system status for variant collection.

2632F

Multiplex mutation panel for molecular diagnostics of increased nuchal translucency with normal karyotype. *P. Tavares[1], J. Sá[2], A. Lopes[2], L. Lameiras[2], L. Dias[2], A. Palmeiro[2], P. Rendeiro[2].* 1) CEO/Clinical Dir, CGC Genetics, Porto, Porto, Portugal; 2) CGC Genetics, Porto, Portugal.

Introduction: Increased fetal nuchal translucency (NT) is usually associated with chromosomal abnormalities, therefore, cytogenetics in amniotic fluid or CVS are offered. In the absence of chromosomal abnormalities, increased NT is associated with fetal malformations, dysplasias, deformations and genetic syndromes resulting in poor perinatal outcome. The most prevalent genetic abnormality associated with increased NT and normal karyotype is Noonan Syndrome. The molecular characterization of genes involved in Noonan syndrome is extremely important for establishing a precise diagnostic evaluation, specifically during the prenatal period. Method (max 200 words) We developed a multiplex mutation panel that allows the molecular identification of Noonan Syndrome as well as the major diseases in the same metabolic pathway - Costello, Cardiofaciocutaneous and LEOPARD Syndrome. This panel test for 80 point mutations on genes PTPN11, RAF1, SOS1, KRAS, HRAS, BRAF, MAP2K1 and MAP2K2. Results: Between 2009-2012 we analyzed 53 prenatal cases with increased nuchal translucency and normal karyotype using the described multiplex mutation panel. Two cases were positive for mutations on PTPN11 gene, establishing the molecular diagnosis of Noonan Syndrome. Results were obtained, in average 10 days, between 11-16 weeks gestation, allowing early detection of an affected fetus. Conclusions: In the presence of ultrasonographic findings such as cystic hygroma, increased nuchal translucency or hydrops fetalis, Noonan syndrome may be suspected. This new molecular diagnostics panel can be applied in prenatal diagnosis, especially in situations with increased nuchal translucency and normal karyotype, allowing an early molecular diagnostics (before week 14) and thus reducing the couple's anxiety and allowing a faster decision process. A broader clinical spectrum and a 10 days diagnosis window are achieved, at a small fraction of the cost of the traditional approach.

2633W

Clinical Interpretation Accuracy of CytoScan® Dx Assay.

J. Tepperberg[1], S. Schwartz[1], A. Roter[2], C. Du[2], R. Duttagupta[2], G. Mamtora[2], J. Danzer[2], J. Wallace[2], S. Close[3], K. Kwiatkowski[2], E. Fung[2], R. Pfundt[4]. 1) Dept Cytogenetics, Lab Corp America, Res Triangle Pk, NC; 2) Clinical Applications, Affymetrix, Santa Clara, CA; 3) GenEngine, Carlsbad, CA; 4) Radboud University Medical Centre, Nijmegen, Netherlands.

Background: Identification of pathogenic copy number variants (CNVs) by cytogenetic microarrays has demonstrated a higher diagnostic yield as compared to conventional methods, such as karyotyping and FISH and are recommended by the American College of Medical Genetics for use as first tier tests in the assessment of patients exhibiting developmental delay, intellectual disability, and congenital anomalies/dysmorphisms. This study's objective was to characterize the clinical interpretation accuracy of CytoScan® Dx assay compared to samples which were assessed for developmental delay and/or intellectual disability using routine patient care (RPC) methods (excluding any Affymetrix array) in post-natal samples. The study also assessed the average number and size of copy number variation (CNV) regions uncovered in patients with a clinical laboratory diagnosis, which included a syndrome, as well as in phenotypically normal individuals. Materials & Methods: The clinical performance of the assay was evaluated in 153 samples across a broad range of syndromes, of which 149 were included in the analysis, and in 110 presumed phenotypically normal samples. The samples were run on CytoScan Dx assay and the results were sent to a cytogeneticist, not involved in the original analysis for interpretation. The results were compared to the original clinical laboratory diagnosis that accompanied the patient's sample. Results: CytoScan Dx assay identified 15.4 ± 11.5 copy number variation regions per syndromic sample and 10.4 ± 3.40 CNVs per phenotypically normal sample. Diagnostic accuracy in pathogenic samples using the clinical interpretation based upon routine patient care as reference was measured as Positive Percent Agreement (PPA). The Positive Percent Agreement was 145/149 = 97.32% (95%CI 93.30-98.95%). Conclusions: This study demonstrated the ability of the CytoScan Dx assay to achieve high clinical accuracy, as measured by PPA at 97.32% as compared to routine patient care methods.

2634T

Clinical application for gene disorders in children using bench top sequencer.

T. Naruto[1,2], Y. Kuroda[1], I. Ohashi[1], K. Kurosawa[1,2]. 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Japan Science and Technology Agency, CREST.

Objectives: Next-generation sequencing (NGS) has transformed genomic research by decreasing the cost of sequencing and increasing the throughput which overcome the limitations of Sanger sequencing methods used to sequence the disease causing genes. Routine clinical use of NGS technologies is appealing, but there are a number of the genes for analysis. We have performed mutation analysis of single gene disorders using the long range PCR based method, and multi gene disorders using the target enrichment method. Methods: The PCR products were purified using a QIAquick column. Libraries were created using the Nextera DNA Sample Prep kit (Illumina) or HaloPlex Custom Kits (Agilent Technologies), then sequenced on a MiSeq. The alignments were performed and compared the Miseq Reporter, SureCall and BWA + GATK based pipeline. Detected mutations were confirmed by Sangar sequence. HGMD, ANNOVAR and SnpEff were used for annotating variants. Results: We identified the novel and new mutation in gene disorders. (1) Single gene disorders analysis can be identified pathogenic mosaic mutations present at low allele frequencies. The mean coverage is over 1,000 of MLL2, TCOF1 and NIPBL multiplex sequencing. (2) Multi gene disorders analysis can be identified CNVs. Sequencing data were scanned for gains and losses using a comparison of normalized coverage data between samples. The 5× and 20 × coverage were 98.4% and 95.6% in 29genes (target size was 140,417 bp) Conclusion: Our method detected mutations, mosaics and CNV, and had a low false-positive rate. In this study, we showed that these gene multiplexing sequence and multi-targeted sequence will be invaluable in small-scale experiments can be done even for small scale individual projects with bench top sequencer.

2635F

Cancer Risk Assessment Using Genetic Panel Testing: Considerations for Clinical Application.

S. Hiraki[1], E. Rinella[2], F. Schnabel[2], R. Oratz[2], H. Ostrer[1]. 1) Albert Einstein College of Medicine, Bronx, NY; 2) NYU Langone Medical Center, New York, NY.

Background: With the completion of the Human Genome Project and the development of high throughput technologies, such as next-generation sequencing, the use of multiplex genetic testing is growing rapidly. The development of genetic cancer panels to assess multiple cancer risks represents one way in which multiplex testing is being applied clinically. There are a number of unique issues to consider when conducting genetic panel testing for cancer risk assessment that differ from the traditional single-gene approach. **Methods:** To address an emerging need for multiplex sequencing panels, we set out to design a cancer panel comprised of genes that confer high or moderately increased risks for breast, ovarian, and colon cancers. Through this process we addressed the following issues: determination of genes to include in the panel, determination of risk estimates and how to convey them to the patient, review of surveillance and management guidelines for increased risk, the genetic counseling process, and return of results. Literature review, cancer gene databases, and existing cancer gene panels were used to inform the development of this panel. **Results:** Our cancer gene panel consists of 26 genes that confer high or moderate risks for breast, ovarian, and colon cancers, with some genes conferring risks for additional cancers. Many of the genes share molecular pathways such as the FANC-BRCA, CHEK2, and MMR pathways. Risks were generally found to cluster at 2 levels: Moderate risk genes, such as those in the FANC-BRCA and CHEK2 pathways for breast cancer, confer a 2-4 fold increased risk while high risk genes, such as CDH1 and PTEN, confer a 10-20 fold increased risk. Established management guidelines for those with increased risk based on family history and clinical factors can be applied to those with comparable risk levels conferred by moderate penetrance genes. **Conclusions:** Cancer panel testing enhances the benefits of genetic risk assessment by 1) extending testing to a wider population beyond those who meet standard genetic testing criteria and 2) broadening the number of gene targets to assess risk, providing a more comprehensive risk assessment. By examining and integrating the data and tools currently available, we can maximize the clinical utility of panel testing and identify the gaps in our knowledge that warrant further investigation.

2636W

Sex determination using free fetal DNA at early gestational ages: A comparison between a modified mini-STR genotyping method and Real-Time PCR.

HReza. Goodarzi[1], MReza. Aghanouri[2], Yasaman. Yazdani[3], G. Mohammadzadeh Shahriary[4], Salman. Abbas Zadeh[5], Saman. Yazdani[6]. 1) Department of Medical Genetics, Shiraz University of Medical Sciences, Shiraz, Iran, MD Ph.D; 2) Department of Medical Genetics, Shiraz University of Medical Sciences, Shiraz, Iran, M.Sc; 3) Department of Medical Genetics, Shiraz University of Medical Sciences, Shiraz, Iran; 4) Department of Genetics, Chamran University, Ahvaz, Iran; 5) Cancer Research Institute, Medical University of Vienna, Vienna, Austria; 6) Department of Laboratory Medicine, Lund University, Sweden.

Introduction: Free fetal DNA in maternal blood can be utilized for noninvasive prenatal genetic screenings. In this study, a new algorithmic base Conventional Polymerase Chain Reaction (PCR) genotyping method and also Real-Time (RT) PCR technique for detecting fetal X and Y chromosome sequences in maternal plasma was applied to determine fetal sex in pregnant women in early gestational period (5-13 weeks). Finally the efficiency of the methods in sex determination was compared to verify which method can be more applicable in clinical settings. Materials and methods: DNA was extracted from 106 pregnant women and their husband's blood samples. Fetus mini-Short Tandem Repeats (STR) genotyping was accomplished through amplification of 19 mini-STRs and three non-STR markers using conventional PCR followed by Poly Acrylamide Gel Electrophoresis (PAGE) analysis. Simultaneously, RT-PCR was carried out by using DYS14 specific primers and probe. Results: In conventional PCR method, 47 cases were diagnosed as male and 49 as female. In comparison, RT-PCR amplified DYS14 (Y-marker) sequences in 45 pregnant women plasma samples. Sensitivity and specificity were calculated as 95.9% and 98% for conventional PCR and 91.8% and 100% for RT-PCR respectively. Conclusion: Conventional PCR technique has revealed a higher level of sensitivity rather than RT-PCR technique and could be employed in future clinical diagnosis singly or in combination with RT-PCR. Prospectively, our future studies will be focused on designing and optimizing an algorithm-based RT-PCR method to avoid any non-accurate results and also on advanced methods for extracting and separating high yields of cell-free fetal DNA (cffDNA) from maternal plasma. Keywords: Sex determination, Free-fetal DNA, Maternal plasma, Mini-STR genotyping, Real-Time PCR, DYS14.

2637T

Exome sequence of genetic disorders in consanguineal family. *G. Atzmon[1], D. Ben Avraham[1], B. Pode-Shakked[2], Y. Anikster[2].* 1) Departments of Medicine and Genetics, Albert Einstein College of medicine, Bronx, NY, USA; 2) Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Israel.

Consanguinity, is common in certain minority populations in Israel, naturally encompasses a greater risk for autosomal recessive disorders. In order to reveal the molecular mechanism and polygenic predisposition of two cases (same family) with autosomal recessive disorders (Osteogenesis imperfecta and Congenital Disorders of Glycosylation (CDG)), we employed an unbiased screening approach by using Exome sequence capture arrays followed by next generation sequencing (HiSeq2000). We Exome sequenced (due to the polygenic anomaly of the investigated diseases) of 6 members of the family, the 2 cases kids their parents and one sibling of each of the parent, as the primary diagnosis of both disorders is of great significance for future family planning. Of the 154,432 sequence variants (available from the 6 samples) 28,487 variants passed the various QC filtering stages with 15 reads/variant. Of which 316 and 216 sequence variants, displayed heterozygosity in the unaffected samples and homozygosity in the effected ones (Osteogenesis imperfect and CDG, respectively). While, forty missense mutation in the Osteogenesis imperfect case demonstrated the homozygosity pattern of inheritance, were the OR4 (olfactory receptor, family 4) gene's subtype exhibit the highest number of variants, among the CDG case 24 were of missense mutation with substantial prevalence of the FANCA gene (which encoded Fanconi anemia). In conclusion, Exome sequencing was demonstrated herein as a powerful tool for Molecular diagnosis of specific mutations and is therefore extremely necessary in such cases. Primary care physicians should be alert to the possibility of more than one autosomal recessive disorder in consanguineous families and evaluate them promptly with the use of such tool.

2638F

Whole exome sequencing for cancer - is there evidence of clinical utility? *A. Malhotra, L. Cushman-Spock, L. Wieselquist, S. Levine, D. Allingham-Hawkins.* Hayes, Inc., Lansdale, PA.

Objective: To assess the evidence behind the use of whole exome sequencing (WES) to identify genetic changes in cancer. **Context:** Next-generation sequencing technologies are advancing at a rapid rate, allowing the generation of a large amount of data in a relatively short period of time. Genetic changes in cancer are increasingly used for diagnosis and may guide treatment decisions. There are often more than 1, and sometimes many, different genes contributing to the clinical presentation of the disease. Given that the exome encompasses < 2% of the genome, variants affecting the cancer being studied may potentially be missed by WES; however, WES allows detection of 85% of disease-causing variants, thereby providing a less expensive method to detect variants when compared to whole genome sequencing. Nevertheless, the question remains: is there evidence that WES impacts patient outcomes in cancer? **Methods:** Proprietary methodology that combines the ACCE (Analytical validity; Clinical validity; Clinical utility; Ethical, legal and social implications) model for genetic test evaluations with internationally accepted health technology assessment methodology was used. Conclusions were based on peer-reviewed published studies of > 10 patients. **Results:** While WES has been conducted in > 10 patients for a number of different cancers (e.g., colon, prostate, and ovarian cancer), breast cancer has been evaluated most extensively (7 studies to date). Studies evaluating somatic alterations showed high intratumor and intertumor heterogeneity. In addition, both novel and previously implicated variants were identified, with varying frequencies based on the breast cancer subtype (e.g., luminal A or B, basal-like, or HER2-enriched). To date, only 2 studies with > 10 individuals (with breast cancer or ovarian cancer) have shown potential for clinical utility of WES, whereby variants identified through WES may determine response to drug treatment. **Conclusions:** Despite evidence for clinical *validity* of WES in cancers, clinical *utility* is very limited and needs to be further evaluated in large clinical studies. In addition, WES results may have ethical implications (e.g., incidental findings); while recent recommendations suggest that such findings must be provided to physicians/ patients, the appropriate way to communicate this information is not completely clear.

2639W

Recognition of Disease-Associated Alleles in the Reference Sequence is Critical for Accurate Disease-Risk Assessment through Genome Sequencing. *G. Chandratillake[1], S. Garcia[1], R. Chen[1, 2], M. Clark[1], S. Chervitz[1], D. Newburger[1], H. Lam[1], J. West[1], R. Chen[1].* 1) Personalis, Inc., Menlo Park, CA; 2) Icahn School of Medicine, Mount Sinai, New York, NY.

The public reference genome sequence (GRCh37) contains minor alleles at >1 million positions. The presence of minor alleles in the reference negatively impacts both sequence read alignment and variant calling. For example, an individual who is homozygous for a minor allele present in the reference will not be reported as variant at that locus, resulting in failure to apply any medical interpretation relevant to that variant to the individual. On the other hand, individuals heterozygous or homozygous for major alleles absent in the reference are reported as variant at such loci, resulting in an increased burden of variant interpretation. To address the issue of rare alleles in the reference, we extended previous work (PMID 21935354) to create an enhanced human reference sequence. We revised 1.1 million positions in GRCh37 where the reference allele is the minor allele by frequency in four different populations. To assess the impact of this enhanced reference on disease-risk assessment, we interrogated GRCh37 for the presence of medically relevant minor alleles, identifying rare/minor alleles associated with Mendelian disease, pharmacogenetics, and complex disease. We found 38 variants in GRCh37 previously reported to be involved in Mendelian disease (HGMD designation DM/DM?) e.g. rs4784677 in *BBS2* associated with Bardet-Biedl Syndrome and rs1529927 in *SLC12A3* associated with Gitelman Syndrome and hypertension. An additional 217 variants designated disease-associated polymorphisms with functional evidence (DFP) were identified, e.g. the Factor V Leiden allele, rs6025. Furthermore, we identified 4 novel presumed-deleterious alleles (nonsense, frameshift, splice-site) in HGMD genes e.g. rs276936 in *DSC3* and rs9959632 in *PIGN*. Whilst it is unlikely that all of these variants are pathogenic, they warrant in-depth review. Such variants would be completely missed in homozygous individuals and likely filtered out due to population frequency in heterozygous individuals due to recognition of the major allele as variant. In addition, we identified 77 variants with pharmacogenetic associations listed in PharmGKB, e.g. rs1954787 involved in citalopram response, and 985 variants associated with complex disease in our extensive manually curated Disease Variant Database. The use of our enhanced human reference sequence facilitates more complete variant discovery~critical for accurate disease-, carrier- and pharmacogenetic-risk assessment through exome/ genome sequencing.

2640W

Autism associated with an Xq12 deletion involving the gene OPHN1 - importance of pursuing a genetic etiology for an Autism Spectrum Disorder (ASD). *E. Elias*. Dept Genetics & Pediatrics, Children's Hosp Colorado, Aurora, CO.

Introduction: ASD's are prevalent, occurring at a rate of 1:88 children per recent CDC report. Many children with an ASD never undergo genetic evaluation. However,current technology allows the determination of a more definitive underlying genetic diagnosis for children with an ASD, especially if accompanied by intellectual disability (ID).

Case Report: A 9 year old boy presented for genetic evaluation of an ASD, and severe ID. Past medical history was notable for strabismus requiring surgical repair and a seizure disorder requiring medication. His exam revealed mildly dysmorphic features, including down-slanting palpebral fissures and a prominent chin. Head MRI was notable for inferior cerebellar hypoplasia, a Dandy-Walker variant and mild ventriculomegaly. Previous genetic testing included normal karyotype and negative Fragile X testing. Family history was notable for a normal brother, and healthy parents.

Genetic Testing: A 191 kb region of loss involving chromosome Xq12 was detected on a Cytochip 180K Oligo microarray panel, which partially overlapped the OPHN1 gene. Similar patients have been reported with either deletions of Xq12, or molecular mutations of the OPHN1 gene. Oligophrenin-1 (OPHN1) encodes a Rho-GTPase-activating protein, the first Rho-linked protein identified in patients with X-linked ID. Rho-GTPase proteins are critical for normal neuronal development and function, particularly regarding dendritic growth and genes affecting glutamate receptors.

Conclusions: This patient had a constellation of problems including strabismus, seizures, and cerebellar anomalies, markedly similar to other patients who have been reported with this rare chromosome deletion syndrome affecting the OPHN1 gene, and is further confirmation of the importance of this gene in CNS morphogenesis. Updated genetic evaluations can be revealing for patients with ASD, especially for those presenting with ID, dysmorphic features, and anomalies. Establishing a definite genetic diagnosis also allows appropriate genetic counseling for this X-linked ID disorder.

2641T

CELL MALIGNITATION ASSOCIATED TO CHROMOSOME TRANSLOCATIONS. CLINICAL MANIFESTATIONS IN TWO PEDIATRIC PATIENTS 46,XY,t(1;4)(q11q11) AND 46,XY,t(6;9)(p21;q34). *J. Aparicio[1, 3], M.L. Hurtado H[2], S. Chatelain M[4].* 1) Dept Gen; 2) Cytogenetics, Hosp para el Nino Poblano, Puebla; 3) Estomatology, benemerita Universidad Autonoma de Puebla; 4) Biotechnology, Universidad Autónoma Metropolitana Mexico DF.

Chromosome aberrations are considered alterations in the chromosome number or structure. They are mainly considered due to gametogenesis inborn error or during the zygote first cellular divisions. Among 4617 chromosomal studies performed during 19 years (from 1992 to 2011), at Hospital Para El Niño Poblano in México, 34.6% (1596 patients) had chromosomal alterations. Among these studies population, 0.23% (11) chromosome translocations were observed. From this data, two male pediatric patients are described, with 1;4 and 6;9 chromosome translocations. Chromosome changes are classified as structural or numeric alterations respectively, and abnormal cell development has been associated with these two specific chromosomal translocations. Both cases were described in this study analyzing their hematological, clinical features, medical treatments and prognosis.

2642F

A Balanced Chromosome Translocation Reveals Involvement of a Predicted Lipase in Weight Gain, Hearing Loss and Tumor Suppression. *B.B. Currall[1,2], K.E. Wong[1], N.G. Robertson[1], T. Hoyos[3,2], A.E. Hickox[4], Y. Yin[5,2], B.J. Quade[1,2], M.C. Liberman[5,2], E.C. Liao[3,2], C.C. Morton[1,2].* 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Purdue University, West Lafayette, IN; 5) Massachusetts Eye and Ear Infirmary, Boston, MA.

The Developmental Genome Anatomy Project (DGAP; www.dgap.harvard.edu) systematically examines subjects with balanced chromosomal rearrangements to identify genes involved in congenital disorders. DGAP056 is an archetypal case with a constellation of symptoms including profound congenital hearing loss, early onset prostate cancer, mild craniofacial abnormalities (coloboma, exotropia, blepharophimosis, and low-set posteriorly rotated ears), mitral valve prolapse and hypospadias. Genomic analyses of DGAP056, using karyotyping, fluorescence *in situ* hybridization (FISH) and genomic sequencing, revealed a balanced translocation involving chromosomes 2 and 13, t(2;13)(p24.1;q22.3)dn, which disrupts a poorly annotated gene designated *C2orf43*. Linkage and genome-wide studies have associated SNPs in and around *C2orf43* with prostate cancer, coronary heart disease and other phenotypes associated with lipid deregulation. Phylogeny and tertiary structure modeling indicate that *C2orf43* encodes a conserved protein related to the alpha/beta hydrolase clan of proteins and likely functions as a serine-based ester hydrolase involved in lipid metabolism. In animal models, *C2orf43* orthologs are expressed in a variety of tissues including adipose tissues, inner ear and prostate. Mouse knockout models have increased weight, high frequency hearing loss and increased rates of tumors (particularly in the prostate). *C2orf43* expression is also down-regulated in human prostate tumors. While it is yet unclear the biochemical role of *C2orf43*, this gene appears to be necessary for proper lipid metabolism and, when disrupted, leads to a variety of diseases including obesity, hearing loss and prostate cancer. DGAP056 demonstrates the power of the DGAP approach using *de novo* genetic lesions, well characterized phenotypes and convergent data to annotate clinically important regions of the genome.

2643W

Genitourinary Defects Associated with Genomic Deletions in 2p15 Involving OTX1. *C. Jorgez[1,2], J. Rosenfeld[6], N. Wilken[2], V. Vangapandu[2], A. Sahin[2], D. Pham[2], C. Carvalho[4], A. Bandholz[6], A. Miller[7], D. Weaver[7], B. Burton[8], D. Babu[9], J. Bamforth[9], T. Wilks[10], D. Flynn[11], E. Roeder[12], S. Cheung[4], J. Lupski[4,5], D. Lamb[1,2,3].* 1) Center for Reproductive Medicine, Baylor College of Medicine, Houston, TX; 2) Scott Department of Urology, Baylor College of Medicine, Houston, TX; 3) Department of Molecular and Cell Biology, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 7) Department of Medical and Molecular Genetics Indiana University School of Medicine, Indianapolis, IN; 8) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 9) University of Alberta, Edmonton, Alberta, Canada; 10) Madigan Army Medical Center Department of Pediatrics, Tacoma, WA; 11) Department of Children's Endocrinology, St. Luke's Children's Specialty Center, Boise, ID; 12) Department of Pediatrics University of Texas Health Science Center at San Antonio, San Antonio, TX.

Normal development of the genitourinary (GU) tract is a highly complex process that frequently goes awry, causing malformations of GU structures. In male children the most frequent congenital anomalies are GU defects such as cryptorchidism (1% to 4% of full term newborns), hypospadias (1%), micropenis (0.35%) and vesicoureteral reflux (VUR) (1%). Other congenital urologic malformations, such as bladder-exstrophy-epispadias (BEE) (1:47000), occur less frequently but significantly impact patients' lives. Array comparative genomic hybridization (aCGH) identified seven individuals with deletions in the 2p15 region (66.0kb-5.6Mb). These deletions encompass the transcription factor orthodenticle-homolog-1 (OTX1) gene. Subject 1 (with BEE and VUR) had the smallest deletion (66kb), encompassing only OTX1, and was identified among 30 BEE patients screened by aCGH. Male subjects 2-5 were identified among 30,183 subjects submitted to Signature Genomics for clinical aCGH testing, and male subjects 6 and 7 were identified among 18,734 subjects analyzed at Baylor Genetic Laboratories for clinical aCGH testing. Subjects 2-7 had large de novo CNVs (2.39-6.31 Mb) and exhibited features similar to the 2p15p16.1 and 2p15p14 microdeletion syndromes including developmental delay, short stature, facial abnormalities, and variable GU defects. Breakpoint analysis indicated that the deletion in subject 1 was at chr2: 63,130,672-63,196,654, and thus it only included the gene OTX1. Our subjects with GU defects included four with testicular, three with penile, two with kidney, one with bladder, and one with VUR abnormalities. Otx1-/- null mice suffer from seizures and have prepubescent transient growth retardation and gonadal defects. Similarly, two of our subjects have short stature, two suffer from seizures, and five have gonadal defects. The presence of GU defects in six of our cohort and in eight patients of the thirteen reported with deletions within 2p14p16.1 (including two with deletion of OTX1) suggests that this region encodes genes important for GU development. Thus, impairment of OTX1 function or expression may lead to GU abnormalities that range from the commonly observed cryptorchidism to the rarely observed BEE.

2644T

6q24.3-q25.1 deletion syndrome. *Y. Nishi[1], M. Tominaga[1], H. Ueda[2], Y. Kuroda[1], I. Ohashi[1], T. Saito[3], J. Nagai[3], K. Kurosawa[1,4].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Cardiology, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Clinical Labolatory, Kanagawa Children's Medical Center, Yokohama, Japan; 4) Japan Science and Technology Agency, CREST.

6q24-q25 deletion syndrome is a rare and newly recognized syndrome characterized by congenital heart defects, growth retardation, and variable degree of intellectual disability. The phenotype varies with the range of deletion and the genes involved. Nowaczyk et al. (2008) reported three patients with paternal deletion of 6q24.4. Two of the patients shared a 2.5 Mb region of overlap and similar facial features including a triangular face, frontal bossing, short and up-slanting palpebral fissures, asymmetry of upper eyelids, shallow orbits, and long and flat philtrum. The cardiac defects include the anomaly in the outflow tract and atrioventricular defect. Thienpont et al. [2010] demonstrated that the cardiac defect of the syndrome was caused by haploinsufficiency of TAB2 gene harbored on 6q25.1. Here we present an additional case of 13-year-old girl with the 6q24.3-q25.1 deletion. She was born to nonconsanguineous healthy parents at 40 weeks of gestational age after uneventful pregnancy. Her birth weight was 2780 g. Recurrent episodes of pneumonia were noted during the infantile period. Her developmental mile stone was delayed; head control at 4 months, walking alone at 1 year and 6 months, speaking comprehensive words at 2 years and 6 months. At age of 6 years, G-banded chromosome analysis was performed due to her mild developmental delay, but the result was normal karyotype. Her intelligence quotient was 51-75 at age of 12 years. She had short stature due to the deficiency of growth hormone and the GH therapy was started at her age of 7 years. Since the age of 8 years, she had precocious puberty and treated with GnRH antagonist. At age of 12 years, atrial premature contraction and severe mitral regurgitation was noted. Because of uncontrollable attacks of arrhythmia, she had catheter ablation therapy repeatedly. To elucidate the cause of these clinical episodes, cytogenetic microarray was performed. The results showed interstitial deletion of 4.2 Mb from 6q24.3-q25.1. The present patient shared several features including the facial appearance, the degree of intellectual disability, growth phenotype, and characteristic cardiac defect. Together with these results and previous reports, the deletion of 6q24.3-q25.1 represents distinctive microdeletion syndrome caused by haploinsufficiency of the genes in the interval. Further analysis is required to elucidate the phenotype-genotype correlation in the syndrome.

2645F

An infant with 49,XXXXY syndrome and congenital cataract. *C. Vinkler[1], A. Ben Sasson[2], A. Singer[3].* 1) Inst Med Genetics, Wolfson Med Ctr, Holon, Israel; 2) Child Developmental Center, Maccabi Health Service; 3) Inst Med.-Genetics Barzilai Medical Center , Ashkelon, Israel.

49,XXXXY syndrome is one of the rarest sex chromosome abnormality in humans, showing an incidence of 1: 85,000 to 1: 100,000. It is described to be the most severe variant of Klinefelter syndrome however it is categorized as a separate syndrome by many authors. Clinical features of the syndrome are: coarse face, microcephaly, distinct dysmorphic features, short stature and an IQ ranging between 20 to 60 points. Recently, increased rates of brain anomalies were also described in this group of patients. We present a patient with 49,XXXXY syndrome and congenital cataract. The proband is a 2y and 3m old boy referred to our clinic because of severe hypotonia, developmenatal delay and dysmorphic features. This male patient was born at term to healthy parents. There is no description of cataracts in the family. Pregnancy and delivery were normal with a birth weight of 2.7 kg. Right after birth hypotonia and dysmorphic features were noticed. Cardiac echo revealed small PFO and PDA Eye examination revealed Lt anterior polar cataract of 1 mm and another posterior subcapsular cataract. Brain ultrasound was normal. Chromosome karyotyping showed a karyotype of 49,XXXXY. On examination at the age of 2y and 3m his head circumference is 46cm (3rd centile) height is 75cm (-4 SD) and weight 9kg (-3.5 SD). He has a coarse facial appearance with hypertelorism, epicanthal folds, upslanting palpebral fissures and a broad nasal bridge. He has small penis and small retractile testes. He is diagnosed with global developmental delay. This is the first report of congenital cataract in a patient with 49,XXXXY syndrome. Congenital cataract has been described previously, in one case of 46,XXY syndrome. The molecular basis for the combination of congenital cataract in sex chromosome polysomy is not known. It had been previously suggested that it may be explained by an increased BCOR gene dosage due to the patient's extra X chromosome. Mapped to chromosome X, mutations of the BCOR gene can result in cataracts, and other ocular abnormalities. X inactivation may explain the variable phenotype and rare expression of cataracts in these patients. Further investigation in model systems is needed to verify this hypothesis. Eye examination is recommended in all patients with sex chromosome polysomy.

2646W

Aiding the interpretation of CNV and sequence variation in DECIPHER using the Genoverse genome browser. *E. Bragin, E.A. Chatzimichali, G.J. Swaminathan, A.P. Bevan, C.F. Wright, M.E. Hurles, H.V. Firth.* Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom.

DECIPHER (https://decipher.sanger.ac.uk) is a web-accessible database and consortium that facilitates the identification and interpretation of genomic variation in patients with developmental disorders. Over 250 academic departments in genetic medicine contribute phenotype-linked variation data into the DECIPHER database for analysis and interpretation. Following informed consent, shared anonymized patient data enables the identification of clusters of patients with similar phenotype-linked genomic findings and encourages collaboration and contact between member centers. DECIPHER also facilitates contact between external users and consortium members making it an invaluable collaborative resource for genomic research and clinical diagnosis. Driven by decreasing costs and improved technologies, sequencing is now increasingly being used in clinical diagnosis alongside arrayCGH. In order to facilitate the combined analysis and interpretation of phenotype-linked copy-number (CNV) and/or sequence variation we have extended and improved DECIPHER to encompass all forms of genomic variation. New features include informative summary and gene tables, variant-effect predictions as well as a purpose-built genome browser (Genoverse: http://genoverse.org) to visualize and interpret copy-number and sequence variation. Genoverse was developed at the Wellcome Trust Sanger Institute in collaboration with Ensembl. The Genoverse browser utilizes modern web technologies to visualize data 'on-the-fly'. Some of the salient features include customizable views, interactive scroll and zoom, as well as visualization of local files by 'drag n' drop'. Genoverse is currently the visualization engine behind DECIPHER and Ensembl. The design of Genoverse allows easy integration into any website and connection with various data formats and sources. In our presentation we demonstrate our implementation of all recent developments in DECIPHER including the integration of sequence and CNV data, managed data access, inclusion of research data and the new genome browser.

2647T

The Diagnostic Yield of Chromosomal Microarray Analysis in a Large Multidisciplinary Craniofacial Clinic. *K. Dipple[1], J. Peredo[1], J.P. Bradley[2], R. Jarrahy[2], F. Quintero-Rivera[3].* 1) Dept Hum Gen, Gonda 5506B, Univ California, Los Angeles, Los Angeles, CA; 2) Dept Surgery, Division of Plastic and Reconstructive Surgery, Univ California, Los Angeles, Los Angeles, CA; 3) Dept Pathology and Laboratory Medicine, Univ California, Los Angeles, Los Angeles, CA.

We report on 100 children tested in the UCLA Craniofacial Clinic using chromosomal microarray in an effort to locate significant areas of the genome and candidate genes associated with craniofacial anomalies. Abnormalities included isolated defects (cleft lip/palate, cleft palate alone, hemifacial microsomia, microtia, Pierre Robin sequence and Tessier clefts), known complex conditions (VATER, CHARGE [without CHD7 mutations or deletions], Nager, OAV/Goldenhar) and individuals with the above isolated defects plus multiple abnormalities, developmental delay and/or intellectual disability. We excluded those patients who had previously identified chromosomal abnormalities by standard G-banded karyotyping. Patients were tested from September 2007 to May 2013 using BAC arrays and later high density oligonucleotides (with and without single nucleotide polymorphism, SNP) arrays once that platform became available. The clinical positive rate of microarray testing in the UCLA Craniofacial Clinic was 10% (n=10/100). Four patients had large regions of homozygosity (ROH) that ranged from 10 Mb to 389 Mb, (one case was 2nd degree relationship). One patient had a duplication of 224 kb, partly overlapping ARL13B, which has been linked to Joubert syndrome type 8 in single-copy form. One patient had a partial tetrasomy/deletion where the gain was not associated with a supernumerary chromosome; the deletion, in Xp11.1, has been linked with autism. Four patients had deletions that ranged from 335 kb to 2.38 Mb. One patient had a microdeletion of 17q24.3 upstream of SOX9 associated with isolated Pierre Robin Sequence (Amarillo et al., AJMG, 2013). Another patient has a 5q35.2-q35.3 microdeletion associated with Sotos syndrome. (Peredo J, et al., Cleft Palate Craniofac J. 2012). We did not find any overlap with previously reported candidate genes for isolated cleft lip/palate or Goldenhar syndrome. Our experience with microarray testing indicates that those children with cleft palate and hemifacial microsomia plus multiple abnormalities (congenital heart defects, encephalopathies, seizures, muscle tone abnormalities) and developmental delay are the most likely to have deletions/duplications that can be detected by CGH array. We did not find deletion/duplications in patients whose craniofacial abnormalities were isolated.

2648F

Effects of up-regulation of the SHH pathway on Ts65Dn, a mouse model of Down Syndrome. *T. Dutka[1], N. Singh[2], J.T. Richtsmeier[2], R. Reeves[1,3].* 1) Institute Of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Department of Anthropology, Pennsylvania State University, University Park, PA; 3) Department of Physiology, Johns Hopkins School of Medicine, Baltimore, MD.

Down Syndrome (DS) is caused by a triplication of human chromosome 21 (Hsa21). DS consists of a constellation of over 80 phenotypes. Three characteristics found in all individuals with DS are some level of craniofacial dysmorphology, brain structural and functional changes, and cognitive impairment. Ts65Dn, a mouse model of DS, contains a freely segregating extra chromosome consisting mostly of the distal portion of mouse chromosome 16 (Mmu16), a region orthologous to part of Hsa21. Previous work has demonstrated that Ts65Dn recapitulates aspects of the craniofacial defects, brain dysmorphology, and intellectual disability seen in DS. Moreover, some aspects of each of these 3 traits have been linked to a reduced response to Sonic Hedgehog (SHH). If all trisomic cells show a similarly reduced response to SHH, then up-regulation of the pathway in affected cells might ameliorate the phenotypes in multiple tissues. To investigate this hypothesis, Ts65Dn mice were crossed with $Ptch1^{tm1Mps/+}$; mice in which the canonical SHH pathway is up-regulated in every SHH-responsive cell through the loss of function of one *Ptch1* allele; this should affect every cell that signals through the canonical HH pathway throughout development. Ts65Dn; $Ptch1^{tm1Mps/+}$ mice were compared to Eu; $Ptch1^{tm1Mps/+}$, Eu; $Ptch1^{+/+}$ and Ts65Dn; $Ptch1^{+/+}$ for craniofacial, behavioral and brain phenotypes. Morphometric assessment of the skull demonstrated specific affects on craniofacial morphology. Evaluation of motor learning was accomplished using an increasing speed rotarod trial; working memory was assessed with a Y-maze; and hippocampal function (contextual learning and visiospatial integration) was explored using fear conditioning and Morris Water Maze. The brain morphology was assessed for cerebellar granule cell density, cerebellar size compared to brain size, foliation and Purkinje cell density with mid-sagittal sections. Comparisons of the effects of reduced Ptch1/increased SHH pathway signaling on all of these phenotypes will indicate if consistently lowering the threshold for SHH stimulation can improve multiple DS phenotypes and clarify roles of the SHH pathway in the overall presentation of DS.

2649W

46,X,del(X)(p11.2) Turner syndrome patient with severe keratoconus. *L. Gabriel[1, 2], L. Junior[4], T. Oliveira[1], I. Silva[1], L. Chaves[1], C. Sousa[1], L. Elias[1], L. Mendonça[1], R. Filho[1], L. Lavigne[1, 2, 5], J. Filho[3], J. Jaime[3], M. Avila[1, 2].* 1) Ophthalmologia, CEROF-HC-UFG, Goiania, GO, Brazil; 2) Brazilian Center for Eye Surgery, BCES, Goiania, GO, Brazil; 3) APAE, Anapolis, GO, Brazil; 4) Centro de Diagnostico por Imagem, CDI, Goiania, GO, Brazil; 5) Hospital Geral Roberto Santos, Salvador, BA, Brazil.

A 19-year-old female with amenorrhea, short stature, and severe keratoconus on both eyes was investigated for the diagnostic hypothesis of Turner syndrome. After an extensive laboratory work-up we found: uterus and ovaries with reduced dimensions, anteverted and anteflexed uterus, elevated follicle-stimulating hormone (FSH), elevated dehydroepiandrosterone (DHEA), and a G-banding karyotype showing the X chromosome partial deletion 46, X, del(X)(p11.2) confirming our suspicion of Turner syndrome.

2650T

Heterotaxy in a woman with mosaic Turner syndrome. *P. Kannan[1], A.E. Lin[1], N.S. Scott[2], I. Sahai[1].* 1) Medical Genetics, MassGeneral Hospital for Children, Boston, MA; 2) Department of Cardiology, Massachusetts General Hospital, Boston, MA.

BACKGROUND: Heterotaxy refers to an abnormal arrangement of the thoracic and abdominal organs due to a failure to establish left-right asymmetry during embryogenesis. Important clinical findings include complex cardiovascular malformations, abnormal inferior vena cava (IVC) relationship to the spine, spleen defects, transverse liver, and intestinal malrotation. Heterotaxy is genetically heterogenous. Not all causative genes are known. Chromosomal abnormalities are uncommon. Turner syndrome is a chromosomal disorder occurring in 1 in 2500 live female births. Short stature, ovarian dysgenesis and heart defects are common findings. Turner syndrome is caused when an entire X chromosome is missing (45,X) or is structurally abnormal. In mosaic Turner syndrome 45,X/46,XX, a milder clinical presentation is typically observed. CASE: We report a 29 year old woman with a history of Total Anomalous Pulmonary Venous return into coronary sinus status post surgical repair in infancy, left sided IVC, polysplenia, left of midline liver and intestinal malrotation. Her karyotype is 45,X[3]/46,XX[27]. Sequencing of the 10 known genes associated with heterotaxy did not reveal any variants. Heterotaxy in Turner syndrome is extremely rare and has been reported in two cases previously; our case is the only one in which known heterotaxy genes were analyzed.

2651F

Dilated cardiomyopathy in a patient with Pallister Killian Syndrome while on a ketogenic diet. *J. Lazier[1], J. Harder[2,3], M.A. Thomas[1,2].* 1) Department of Medical Genetics, Alberta Children's Hospital, Calgary, Alberta, Canada; 2) Department of Pediatrics, Alberta Children's Hospital, Calgary, Alberta, Canada; 3) Department of Pediatric Cardiology, Alberta Children's Hospital, Calgary, Alberta, Canada.

Pallister Killian Syndrome (PKS) is caused by mosaic tetrasomy of chromosome 12p. It is characterized by multiple congenital anomalies, characteristic craniofacial anomalies, and intellectual disability. Seizures are common and often poorly controlled with medication. Up to 40% of patients with PKS have congenital heart anomalies, including atrial septal defects, patent ductus arteriosus and bicuspid aortic valves. Cardiomyopathy is a rare association, with only one reported case of hypertrophic cardiomyopathy. There are no previously reported PKS patients with dilated cardiomyopathy as seen in our patient. *Vogel et. al.* (2009) reported a case of a combined hypertrophic/dilated cardiomyopathy reported in a patient with hexasomy 12p, and Parasuraman *et. al.* (2011) reported a case of severe dilated cardiomyopathy reported prenatally in a fetus with mosaic trisomy 12. These cases suggest that upregulation of genes in this area may contribute to dilated cardiomyopathy.

Ketogenic diets are commonly used to treat children with severe refractory epilepsy, such as that seen in PKS. While they are generally considered safe, long term use has been associated with metabolic changes, including selenium deficiency, which is present in up to 20% of patients on this diet. Rare cases of dilated cardiomyopathy and long QT syndrome have been seen in these selenium deficient patients.

We present the case of a four year old girl with PKS who presented with a long QT syndrome and acute dilated cardiomyopathy. An echocardiogram done 15 months previously had been normal. At the time of presentation, she was on a ketogenic diet for medically refractory epilepsy and had low selenium levels. Despite cessation of the ketogenic diet, medical treatment of her heart failure, and initial improvement of her cardiomyopathy, the patient died from congestive heart failure two months after the initial presentation. Dilated cardiomyoapthy has not been previously reported in PKS. Given that there is a low risk of cardiomyopathy in all patients who are selenium deficient secondary to a long term ketogenic diet, this was likely the explanation for her cardiomyopathy. It is possible, that based on their genetic background, PKS patients are at increased risk of this rare complication and close monitoring of selenium levels may be even more important.

2652W

Clinical and cytogenomic evaluation in two siblings with an 8.5 Mb 6q24.2q25.2 deletion inherited from a paternal insertion. *M.I. Melaragno[1], S.S. Takeno[1], M. Migliavacca[1], A.L. Pilla[1], N.L.M. Sobreira[2], C.B. Mello[3], V.A. Meloni[1].* 1) Genetics Division, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) McKusick Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, USA; 3) Psychobiology Department, Universidade Federal de Sao Paulo, Sao Paulo, Brazil.

Interstitial deletions of the long arm of chromosome 6 are rare. Recently, a clinical 6q microdeletion syndrome was identified as being associated with intellectual disability (ID), microcephaly, characteristic facies, minor dysmorphic features, and multiple organ anomalies. We present two siblings, male and female, with a 6q24.2q25.2 deletion studied in detail by different molecular methodologies. Both patients had global developmental delay, ID, microcephaly, short stature, characteristic facial dysmorphisms, and multiple organ anomalies. Different from the previous cases described in literature, our patients had moderate skin laxity in hands and the female patient had stenosis of the thoracoabdominal aortic junction and atrial septal defect. They differ concerning ID level since the female has moderate, and the male has mild ID. Both patients' karyotype analyses were normal; however, SNP array showed an 8.5 Mb interstitial deletion on chromosome as follows: .arr 6q24.2q25.2(144,444,361-152,966,111)x1. FISH-BAC using the probe RP11-632J15 (6q24.3 region) confirmed the deletion in both siblings, and revealed normal signal in the mother and the older sister. Furthermore, the father has an insertion of a chromosome 6q segment into the long arm of chromosome 8 (8q13), which results in a high recurrence risk for the disease. Our study confirms SNP-array as a powerful method in the identification of cryptic deletions, and underlies the essential role of BAC-FISH in assessing the position of the chromosomal segment involved in the deletion, and detecting balanced carriers. Thus, in our study both methods were important in defining patients' clinical follow-up and genetic counseling. Financial support: FAPESP and NHGRI (1U54HG006542).

2653T

Developmental and Growth Delays, Dysmorphic Features and Microcephaly in a Child With Microduplication 12q13.12. *C. Melver, A. Archbold.* Pediatrics, Div Gen, Akron Children's Hosp, Akron, OH.

We report the case of a 9-year-old boy with a duplication within band 12q13.12 which is de novo and not previously described. He presented for genetics evaluation for developmental delay and dysmorphic features. His medical history was otherwise significant for small size, microcephaly, and undescended testis. An unknown congenital cardiac condition reportedly resolved in infancy, and he is now followed by cardiology for mildly increased triglycerides and total cholesterol. He had mild motor delays, walking around 18 months, but more significant delays in speech. He said his first words after his second birthday and did not use sentences until age four. At the time of his visit he was in the third grade but doing what his mother described as kindergarten-level work. On review of family history his mother reported removal of a unilateral cataract at age 5 thought to be secondary to toxoplasmosis. The remainder of the family history was significant only for a distant cousin on each parent's side of the family thought to have diagnoses on the autism spectrum. On physical exam his height and weight were at the 3rd percentile and his head circumference was less than the 3rd percentile and 50th percentile for 3 years of age. His face was significant for upslanting palpebral fissures without epicanthic folds, double cowlicks and an overall impression of mild facial asymmetry and mild microagnathia. Previous genetic workup at that time included normal karyotype, FISH for SNRPN and testing for Fragile X syndrome. A whole-genome oligonucleotide array CGH with SNP was done which revealed a de novo duplication of at least 417 kb within band 12q13.12. This duplication has not been previously described to vary in its entirety in the general population. It contains 4 genes with known clinical associations, including TUBA1A and MLL2 associated with neurodevelopmental phenotypes. Heterozygous mutations in these two genes are associated with lissencephaly type 3 and Kabuki syndrome type 1, respectively. The clinical consequences of having three copies of these genes are not known. Based on the de novo occurrence the laboratory has now classified this as a positive result representing a possible new microduplication syndrome.

2654F

DIAGNOSTIC AND MANAGEMENT CHALLENGES OF GENETIC DISEASES IN RWANDA. *L. Mutesa[1,2], A. Uwineza[1,2], J. Hitayezu[1], S. Murorunkwere[1], E.K. Rusingiza[3], L. Tuyisenge[3], R. Teteli[3], J. Mucumbitsi[4], N. Muganga[3], A.C. Hellin[2], M. Jamar[2], V. Bours[2].* 1) Medical Genetics, National University of Rwanda, Huye, Huye-Butare Southern Province, Rwanda; 2) Center for Human Genetics, University of Liege, Belgium; 3) Department of Pediatrics, Kigali University Teaching Hospital, Rwanda; 4) Department of Pediatrics, King Faysal Hospital, Kigali, Rwanda.

Background: Genetic diseases refer to genetic disorders caused by defects or abnormalities in chromosomes or genes. Chromosome disorders include numerical anomalies (e.g. monosomies, trisomies, etc) and chromosomal rearrangements mainly characterized by unbalanced translocations, inversions, duplications or deletions. In addition, molecular defects affecting gene expression represent a high proportion of monogenic or polygenic genetic diseases. Methods: In the present study, we conducted a survey aiming at assessing clinical aspects and genetic diagnostic patterns of Rwandan patients presenting with clinical phenotype suggestive of genetic disease. This survey was a compilation of several genetic studies so far conducted within a six-year-period starting from 2006 to 2012. Most of patients were selected based on clinical features suggesting of a specific genetic disease. Cytogenetic studies including karyotype, Fluorescent In Situ Hybridization (FISH) analysis were performed in the majority of these patients. In addition, molecular tests such as Polymerase Chain Reaction (PCR), gene sequencing or Multiplex Ligand Probe Amplification (MLPA) analysisand CGH-array were done for better diagnostics' characterization of monogenetic or chromosomal disorders. Results: In total, we identified a considerable number of chromosomal abnormalities including trisomy 21, trisomy 13, trisomy 18, Turner syndrome, Cat Eye syndrome, Williams symdrome and others. Furthermore, monogenic disorders such as cystic fibrosis-like disease, sickle cell anemia, spinocerebellar ataxia type 2 (SCA2), Hunter syndrome, Hutchinson-Gilford Progeria syndrome were also found. In addition, seven patients showed microdeletion and microduplication syndromes. Other polygenic disorders including holoprosencephaly, meningomyelocele, cyclopia syndrome, and encephalocele sydrome were also identified. Conclusion: this survey revealed a large number of genetic diseases in Rwanda. The cytogenetic and molecular tests should be mandatory for diagnosis assessment. Nevertheless, their management is still a challenging situation in our African settings. However, genetic counselling should be provided to all families at high risk of a recurrent autosomal dominant, recessive or X-linked diseases.

2655W

Neonatal Management of Trisomy 13: Clinical Details of 12 Patients Receiving Intensive Treatment. *E. Nishi[1,2], M. Takasugi[3], T. Hiroma[3], T. Nakamura[3], Y. Fukushima[2], T. Kosho[1,2].* 1) Department of Medical Genetics, Nagano Children's Hospital, Azumino, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Neonatology, Nagano Children's Hospital, Azumino, Japan.

Trisomy 13 is one of the most common autosomal trisomy syndrome, characterized by multiple congenital anomalies, severe developmental delay, and a short life span with the 1-year survival rate as 5-10 % and the median survival time as 7 days through population-based studies. Management of neonates with trisomy 13 is controversial, palliative care or intensive treatment, supposedly due to the lack of precise clinical information of the syndrome especially on efficacy of treatment. To delineate the natural history of trisomy 13 managed under intensive treatment, we reviewed detailed clinical data of 12 patients with full trisomy 13 admitted to the neonatal intensive care unit of Nagano Children's Hospital, providing intensive treatment to those with trisomy 13, from 2002 to 2012. Two patients had prenatal karyotyping through amniocentesis. Five were delivered by a cesarean. Major clinical findings included congenital heart defects (92%), cryptorchidism (83% of male), cleft lip or cleft palate (75%), polydactyly (67%), tracheomalacia or laryngomalacia(58%), prolonged hypoglycemia(50%), gastroesophageal reflux (50%), and biliary tract disease (50%). 10 patients (83%) received resuscitation by intubation. Mechanical ventilation was required by 10 (83%), two (20%) of whom were extubated and six of whom needed tracheostomy. Surgical operations included tracheostomy, abdominal operation for umbilical hernia, ileostomy, hepatic portenterostomy, plastic operation for cleft lip, plastic operation for polydactyly, cataract surgery, were performed on 10 (83%), of the patients. Enteral feeding was accomplished in all, four of whom were fed orally. Three patients could be discharged home and are still alive. Common final modes of death were congenital heart defects and heart failure (67%). The survival rate at age 1 day, 1 week, 1 month, and 1 year was 100%, 100%, 92%, and 17%, respectively. Median survival time was 229.5 days (range, 22-2334). These date have a resemblance to the date about the patients with trisomy 18 in our hospital. [Kosho et al., 2006] The current study, though the sample size is small, has demonstrated improved survival through neonatal intensive treatment, which would be helpful for clinicians to offer the best information on treatment options to families of patients with trisomy 13.

2656T

Estimation of live birth prevalence of Down syndrome in Japan. *I. Ohashi[1], Y. Kuroda[1], M. Masuno[2], Y. Kuroki[2], K. kurosawa[1].* 1) Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Kawasaki University of Medical Welfare, Kurashiki, Okayama,-Japan.

The live birth prevalence of Down syndrome is determined by several factors including the rates of advanced maternal age, prenatal diagnosis, and policy for the prenatal screening. The maternal ages have steadily increased from 1970s in most of the developed countries. However, the rate of live birth prevalence of Down syndrome varies between countries and geographical status. According to the data from International Clearinghouse for Birth Defects Surveillance and Research (ICBDSR) Annual Report 2011, the peak of the incidence of Down syndrome was observed at 2008 in Japan. The most recent data of the prevalence during the last 4 years has not been presented for the time lag of the epidemiological study. To elucidate the trend of live birth prevalence of Down syndrome in Japan, we used three sets of data from the Kanagawa Birth Defects Monitoring Program (KAMP) during the period 1981-2008, the clinical records of patients with Down syndrome in Kanagawa Children's Medical Center (KCMC), and demographic data in Kanagawa Prefecture. KAMP is a population-based monitoring system, covering about half of total births in Kanagawa Prefecture including 40,000 births annually. KCMC is a tertiary children's hospital in Kanagawa Prefecture. The prevalence of Down syndrome in KAMP was correlated with the numbers of the patients with Down syndrome at KCMC during the study period. The rate of Down syndrome was decreased or stationary from 2009 to 2012. Despite of the fact that the rise of maternal ages would lead to increase of the prevalence of DS, our data showed opposite results. The total number of prenatal diagnosis was stationary or slightly increased, but the rate of mothers over 35 increased two fold from 1998 to 2008 in Japan [Sasaki et al. 2011]. These results suggest that the prenatal diagnosis focused on the advanced maternal ages have had the impact on the birth prevalence of DS since 2009 in Japan. Non-invasive prenatal testing (NIPT) is a screening test to identify pregnancies at risk for common aneuploidies (trisomy 13, 18, and 21). In Europe and United States, the NIPT has been introduced as a routine first screening from 2011. In Japan, NIPT was started as a research framework for clinical trial since April 2013. Further analysis for the prevalence of DS is required for the estimation of the use and spread of prenatal diagnostic techniques and elucidating the impact of elective termination on DS.

2657F

National Institutes of Health Activities in Down syndrome Research: From Creation of a Consortium to Support of a Patient Registry, DS-Connect. *MA. Parisi[1], S. Bardhan[1], L. Kaeser[1], ML. Oster-Granite[1], V. Rangel Miller[2], YT. Maddox[1].* 1) IDD Branch, NICHD, NIH, Bethesda, MD; 2) PatientCrossroads, San Mateo, CA.

Down syndrome (DS) is one of the most common genetic causes of intellectual disability worldwide. The National Institutes of Health (NIH) has a longstanding program of research in DS to improve understanding of this chromosomal disorder, with the goal of developing effective treatments to improve quality of life for those with DS. In 2007, the Trans-NIH Down Syndrome Working Group (DSWG) published *The NIH Research Plan on Down Syndrome* identifying long- and short-term research goals. An NIH DS portfolio review demonstrates progress in DS research; needs remain in several longer-term objectives, and some new goals need development. For example, recognition of the high prevalence of Alzheimer's disease in aging individuals with DS prompted the NIH to sponsor a workshop in April 2013 on *Advancing Treatments for Alzheimer's Disease in Individuals with Down Syndrome*; the outcomes of this workshop will inform the upcoming revision of the research plan.

In December 2010, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) and the Global Down Syndrome Foundation co-sponsored a conference on patient registries, research databases, and biobanks. One of the recommendations was to create a forum for ongoing discussions about research in DS. To that end, the NICHD convened a DS Consortium composed of ~20 members including self-advocates and family members, DS researchers, leaders of professional DS organizations and foundations, and members of the DSWG. The Consortium explored options for creating a Down syndrome patient registry. In September 2012, NICHD awarded a contract to create *DS-Connect: The NIH Down Syndrome Consortium Registry* to facilitate information sharing among families, persons with DS, researchers, and parent groups. Family members and those with DS will be able to enter contact and health information into an online, confidential, and secure database. A researcher and professional portal willl allow evaluation of aggregated de-identified data. If a participant is eligible for a research study, Registry staff will invite them to contact the investigator for potential participation in the study. DS-Connect is planned for launch within the coming year. These activities highlight progress and remaining challenges in DS research. With new therapeutics being developed to treat the cognitive deficits in DS, the need for these resources is timely.

2658W

Phenotype correlation of a patient with a large 16q23.2 to 16q24.3 duplication and a patient with a 16q23.3 to 16q24.3 duplication and small 16p13.3 deletion. *J. Richer[1,2], J. McGowan-Jordan[1,2].* 1) Genetics Department, CHEO, Ottawa, Canada; 2) University of Ottawa, Department of Pediatrics, Ottawa, Canada.

Patients with large duplications involving the long arm of chromosome 16 are extremely rare. A limited number of patients with large duplications of chromosome 16 with additional rearrangements are described in the literature; descriptions in the literature of patients with isolated large interstitial duplications are rare. We present a girl who has a de novo pure duplication of the long arm of chromosome 16 arising from a derivative chromosome 15. We also present an older patient with a similar duplication in conjunction with a deletion of the short arm of chromosome 16 consistent with the better described 16p13.3 deletion syndrome. Case 1: The patient, now 21 months of age, presented with dysplastic kidneys, failure to thrive with all growth parameters below the 5th percentile, dysmorphic features and cognitive impairment. There were also concerns about recurrent upper respiratory infections and about the child's ability to tolerate vaccination. Genomic 180,000 oligonucleotide array showed a 9.8 Mb duplication of 16q23.2 to the terminus at 16q24.3. Case 2: The patient is a 16 year old female. She has failure to thrive (height at the 5th percentile and a weight below the 5th percentile, but head circumference just above the 50th percentile), severe cognitive impairments (still non-verbal and non-continent), idiopathic thrombocytopenic purpura requiring a splenectomy, low immunoglobulins (IgG), alpha-thalassemia trait and dysmorphic features. Genomic SNP microarray showed a large terminal 16q duplication of 16q23.3 to 16q24.3, as well as a 16p13.3 terminal microdeletion. In conclusion, the dysmorphic features associated with the 16q duplications in our patients are similar with those described in the literature in patients with duplication of the entire long arm of chromosome 16. Based on our results, we postulate that the phenotype of pure large duplications of 16q23.2-q24.3 includes heights and weights which remain at or below the 5th percentile and head circumference near or just above the 2nd percentile. Severe developmental disabilities with more pronounced delays in verbal skills than in motor skills characterize our older patient. Finally, it remains to be established whether low IgG may be part of the phenotype.

2659T

Predicting Obstructive Sleep Apnea in People with Down Syndrome.
B. Skotko[1, 5], M. McDonough[1], L. Voelz[2], D. Rosen[2, 5], A. Ozonoff[2, 5], E. Davidson[2, 5], V. Allareddy[2, 5], N. Jayaratne[2], R. Bruun[2, 5], N. Ching[2], G. Weintraub[4], L. Albers Prock[2, 5], R. Becker[2, 5], D. Gozal[3]. 1) Medical Genetics, Massachusetts General Hospital, Boston, MA; 2) Boston Children's Hospital, Boston, MA; 3) The University of Chicago, Chicago, IL; 4) David Geffen School of Medicine at University of California, Los Angeles, CA; 5) Harvard Medical School, Boston, MA.

Obstructive sleep apnea (OSA) in individuals with Down syndrome is associated with multiple morbidities: systemic and pulmonary hypertension, glucose intolerance, cardiovascular and cerebrovascular disease, and behavioral problems. The prevalence of OSA in this population is very high, with estimates ranging between 55-97%. Currently, an overnight polysomnogram (sleep study) is the gold-standard diagnostic test for patients with Down syndrome. Yet, this testing is cumbersome, poorly tolerated by these children, costly, and not widely available around the country. In this study, we looked to identify predictive factors for OSA in persons with Down syndrome. We enrolled ~100 subjects, ages 3-35 years, who already participate in the Down syndrome Program at Boston Children's Hospital. For each patient, we collected subjective and objective measurements using validated parental survey instruments, standardized physical exams, lateral cephalograms, 3D-digital photogrammetry, and urine samples. Afterwards, all participants underwent standardized polysomnography at the Boston Children's Hospital Sleep Laboratory where objective measurements were collected on OSA. We analyzed which combination of our assessment methods best predicted OSA, as ultimately determined by polysomnography. This will be the first time presenting the results of our data. Our final screening tool will hopefully allow physicians to avoid ordering polysomnograms for those individuals with Down syndrome at lowest risk of OSA. Further, those patients with Down syndrome and clear predictors for OSA can proceed directly toward adenotonsillectomy, the current treatment.

2660F

Diamond Blackfan Anemia due to a RPS17 gene deletion in a child with previously detected `balanced' 15,16 translocation.
J. Sullivan, S. Barry, J. Luty, M-A. Abbott. Baystate Medical Ctr, Springfield, MA.

Diamond-Blackfan anemia (DBA) is a rare congenital anemia due to arrested maturation of erythrocytes. Classic cases present as profound anemia in the first 1-2 years of life. Diagnosis is made with complete blood count demonstrating macrocytic anemia and reticulocytopenia in the setting of bone marrow deficient in erythroid progenitors. There may be associated physical malformations including thumb abnormalities, urogenital defects, craniofacial dysmorphisms, low birth weight, growth and developmental delay. There is also a modest increase risk of acute myeloid leukemia, osteogenic sarcoma, and other malignancies. In the majority cases of DBA there is no family history and the causative gene mutations are de novo; however, in 10-25 percent of cases there is a positive family history with autosomal dominant inheritance and variable expressivity. Approximately 50 percent of DBA individuals possess mutations in genes coding for ribosomal proteins (DBA1-10); DBA4 due to mutations in the RPS17 gene at 15q25.2, accounts for less than 1 percent of cases. We report a 4 year old boy who initially presented with upper respiratory illness and low blood counts thought to be related to viral suppression. Symptoms (pallor, shortness of breath and leg pain) continued and he was referred to pediatric hematology where a repeat CBC showed persistent low Hgb and elevated ferritin. A previous work-up for developmental delays revealed a de novo 46,XY, t(15:16)(q26.1;q13) karyotype. Although this reciprocal translocation appeared cytogenetically balanced, whole genome oligonucleotide array was performed to rule out a cryptic imbalance of genetic material and revealed a 3.3 Mb de novo intersitial deletion extending from cytogenetic band 15q25.1 to 15q25.2. The deleted region contains more than 25 genes including the RPS17 gene associated with DBA4. Recurrent microdeletion of 15q25.2 has been described in association with a variable presentation that can include DBA; anomalies including congenital diaphragmatic hernia, VSD, and cryptorchidismm; developmental and behavioral abnormalities; and short stature. Our patient with DBA4 has mild physical dysmorphisms, no birth defects, and resolving developmental delays.

2661W

The Developmental Genome Anatomy Project (DGAP): Annotating the Human Genome from Balanced Chromosomal Rearrangements.
K.E. Wong[1], I. Blumenthal[2], H. Brand[2,3], B. Currall[1], C. Hanscom[2], T. Hoyos[4], D. Lucente[2], Z. Ordulu[1], M.R. Stone[2], S. Pereira[1], V. Pillalamarri[2], L.P. Yuan[5], J.F. Gusella[2,3,6], D.J. Harris[7], E.C. Liao[4], R.L. Maas[8], B.J. Quade[5], M.E. Talkowski[2,4,6], C.C. Morton[1,5,6]. 1) Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Departments of Psychiatry, Neurology, and Genetics, Harvard Medical School, Boston, MA; 4) Department of Plastic and Reconstructive Surgery, Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA, USA; 5) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 6) Program in Medical and Population Genetics, Broad Institute, Boston, MA; 7) Division of Genetics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; 8) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

DGAP (www.DGAP.Harvard.edu) is a collaborative effort to identify genes critical in human development and disease pathogenesis through study of genes disrupted by balanced chromosomal rearrangements. Since 2001, 263 subjects have been enrolled, with recurring phenotypes that include craniofacial abnormalities (64%), symptoms of neurodevelopmental disorder (NDD) (70%) - 45%; comorbidity of both traits -, and other anomalies, including hearing loss (11%). Detection of breakpoints have migrated from FISH mapping and Southern blotting to next-gen sequencing, and 165 candidate genes and 21 non-coding RNAs have been elucidated. Sequencing of 99 genomes from different individuals harboring a *de novo* balanced rearrangement has identified a rich landscape of chromosomal reorganization previously undetectable by cytogenetic methods. We detected complex chromosomal exchanges in ~20% of breakpoints at sequence resolution (previous estimates using cytogenetics predicted a rate of 2.8% complex exchanges), 90% of which contained microinversions, predominantly repaired through non-homologous end joining, and defined that highly complex localized chromosome 'shattering', or chromothripsis, resolves to largely balanced germline states in individuals with congenital anomalies. To date, 26 animal models have been acquired or created for functional analysis of candidate genes. DGAP provides a unique opportunity to annotate the human genome from an 'n of one' perspective with exciting gene discoveries including a component of the BRAF-histone deacetylase complex as the critical intellectual disability gene in the Potocki-Shaffer syndrome region (*PHF21A*), a histone/lysine demethylase/methyl-transferase in a profound debilitating phenotype (*KDM6A*), genes for traits such as height (*HMGA2*), a noncoding RNA in a severe developmental disorder (*LINC00299*), a chromodomain protein associated with autism spectrum disorder (*CHD8*), and single gene contributors to previously defined regions of genomic disorders (*CDK6*, *MBD5*, *KIRREL3*, *EHMT1*, *SATB2*, and *SNURF-SRPN*). Although largely a basic research endeavor, DGAP methods have recently been implemented as clinical diagnostic tests, particularly for counseling dilemmas encountered with prenatal detection of *de novo* balanced rearrangements. With research results returned to physicians, DGAP offers a wealth of information on poorly annotated, clinically relevant areas of the genome and is actively seeking enrollments and collaborators.

2662T

Establishing a reference group for distal 18q-: clinical description and molecular basis. *J. Cody[1,3], M. Hasi[1], B. Soileau[1], P. Heard[1], E. Carter[1], C. Sebold[1], L. O'Donnell[1,2], B. Perry[4], R. Stratton[1], D. Hale[1].* 1) Dept Pediatrics, Univ Texas Hlth Sci Ctr, San Antonio, TX; 2) Dept. Psychiatry, Univ Texas Hlth Sci Ctr, San Antonio, TX; 3) The Chromosome 18 Registry and Research Society, San Antonio, TX; 4) Ear Medical Group, San Antonio, TX.

Although constitutional chromosome abnormalities have been recognized since the 1960s, clinical characterization and development of treatment options have been hampered by their obvious genetic complexity and relative rarity. Additionally, deletions of 18q are particularly heterogeneous, with no two people having the same breakpoints. We identified sixteen individuals with deletions that, despite unique breakpoints, encompass the same set of genes within a 17.6 Mb region. This group represents the most genotypically similar group yet identified with distal 18q deletions. As the deletion is of average size when compared with other 18q deletions, this group can serve as a reference point for the clinical and molecular description of this condition. We performed a thorough medical record review as well as a series of clinical evaluations. Common functional findings included developmental delays, hypotonia, growth hormone deficiency, and hearing loss. Structural anomalies included foot anomalies, ear canal atresia/stenosis, and hypospadias. The majority of individuals performed within the low normal range of cognitive ability but had more serious deficits in adaptive abilities. Of interest, the hemizygous region contains 38 known genes, 26 of which are sufficiently understood to tentatively determine dosage sensitivity. The data suggest that 20 are unlikely to cause an abnormal phenotype in the hemizygous state and five are likely to be dosage sensitive: *TNX3, NETO1, ZNF407, TSHZ1*, and *NFATC*. *ATP9B* may be conditionally dosage sensitive. Not all distal 18q- phenotypes can be attributed to these six genes; however, this is an important advance in the molecular characterization of 18q deletions.

2663F

Dysmorphological Characteristics of Mosaic 4q31 Terminal Deletion: A Case Report. *D. Torun, Y. Tunca.* Dept. of Medical Genetics, Gulhane Military Medical Faculty, Ankara, Turkey.

Terminal deletion of the long arm of chromosome 4 is a rare condition and the clinical findings may vary depending on the region that is missing. Developmental, craniofacial, digital, skeletal and cardiac involvements are the most common anomalies observed in these patients. Most of the common clinical findings which described for 4q terminal deletion syndrome are restricted to the 4q33----qter. A 7 year old male patient from non-consanguineous healthy parents has been evaluated for the psychomotor developmental delay and multiple congenital anomalies. He had been operated for cleft soft palate, macroglossia, undescended testicles and inguinal hernia. Broad forehead, arched eyebrows, upslanted palpebral fissures, telecanthus, depressed nasal bridge, anteverted nostrils, down-turned corners of the mouth, retrognathia, micrognathia, high palate, bifid uvula, protruding and dysplastic ears and clinodactyly has been observed on his physical examination. Abdominal ultrasound, brain magnetic resonance imaging and hearing tests were normal. Echocardiography showed atrial septal defect. Cytogenetical analysis of patient revealed 46, XY [20%]/46, XY, del (4)(q31qter) [80%]. In this study, we made a literature review and compared the clinical findings between mosaic and complete 4q31 terminal deletion. As we know, this study is the first report of deletion of chromosome 4 (q31qter) in mosaic pattern and gives a further delineation for the clinical features.

2664W

Monozygotic twins with Turner syndrome and discordant phenotypes but concordant genotypes due to in utero chimerism. *A. O'Donnell[1], T.L. Toler[2], J. Krier[1], S.E. Dukhovny[3], S. Eggert[4], S. Weremowicz[4], F. Bieber[4], A.E. Lin[2].* 1) Genetics, Boston Children's Hospital, Boston, MA; 2) Genetics, Massachusetts General Hospital for Children, Boston, MA; 3) OB/GYN, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

The chromosome basis of Turner syndrome includes complete 45,X in 40% and mosaicism in 20-30% of affected women, usually with phenotypic correlation. We report on monozygotic twins diagnosed with mosaic Turner syndrome (45,X / 46,XX) with discordant phenotypes, where only one of the twins has features of Turner syndrome. The twins were the product of a triplet IVF pregnancy where two embryos were implanted and resulted in live births of one male and twin female infants.

The phenotypically affected twin had characteristics of classic Turner syndrome *in utero* and postnatally, including coarctation of the aorta requiring surgical repair, webbed neck, and hand and pedal lymphedema. A buccal karyotype on DOL 2 showed 92% 45,X cells. She developed growth delay for which she receives growth hormone. The phenotypically unaffected twin had a blood karyotype performed at 3 months of age that unexpectedly showed 33% 45,X cells. She did not manifest any obvious features of Turner syndrome. Blood karyotyping repeated at 9 years of age showed approximately 1/3 of the lymphocytes were 45,X in both of the twins. We performed skin biopsies and fibroblast karyotypes on the now 11-year-old twins. The affected twin's karyotype was 100% 45,X while the unaffected twin's karyotype was 100% 46,XX.

We hypothesize that twin-to-twin blood mixing *in utero* led to seeding of the hematopoietic stem cell pools during development and resulted in both twins having lymphocytic mosaicism for Turner syndrome. Twin-to-twin blood mixing through placental anastomoses is a common occurrence for monozygotic twins. If the loss of the X chromosome occurred at or after the twinning event, both twins would have isolated hematopoietic chimerism. This is distinct from mosaicism where the loss of the X chromosome occurred prior to the twinning events, in which case the unaffected twin received a smaller proportion of 45,X cell. In mosaicism, there are expected to be some 45,X cells in the fibroblast lineage of both twins. The mechanism of chimerism, especially in contrast to mosaicism, has implications for the risk of ovarian competency for the unaffected twin. This case highlights the importance of using a non-hematopoietic cell line for any genetic testing of discordant twins, given the risk of both false positive results for the unaffected twin and false negative results for the affected twin.

2665T

Detection of an FMR1 Deletion by Chromosomal Microarray Analysis. *J. Smith[1], A.E. Scheuerle[2], L. Ellis[1], S. Peacock[1], P. Fang[1], A. Patel[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Tesserae Genetics, Dallas TX.

Fragile X syndrome, one of the most common etiologies of intellectual disability, is typically caused by expansion of CGG repeats silencing the FMR1 gene. However, silencing of FMR1 by deletion has been reported in a few cases with variation in size and gene content; the phenotypic picture is still evolving. A 2 year 4 month old male was referred for microarray analysis and Fragile X testing due to moderate developmental delay and mildly dysmorphic features. The patient and his dizygotic twin brother, whose development is reportedly normal, were conceived through in vitro fertilization due to unexplained infertility. General health is normal. Motor development was delayed; he sat at 9 months, crawled at 1 year, walked at 19 months and had 2-4 words at 28 months. He was minimally interactive. He was non-verbal but had constant loud vocalizations, tonation changing with mood: happy/angry/sad. The patient was reported to have repetitive habits of rocking and pacing. Physical exam revealed length at 50th centile and head circumference of 25th centile. He had mild dolicocephaly, long palpebral fissures with mild lateral ectropion, flattened midface and periorbital area, large ears, wide nose, patulous lips with downturned mouth, and thick, short toes and fingers. MRI of the brain at 18 months was normal except for a small pineal cyst. Cytoscan HD (Affymetrix) array revealed an 89.48 kb deletion including the FMR1 and FMR1-AS1 genes; the proximal breakpoint was approximately 47 kb upstream of FMR1 and the distal breakpoint approximately 3.5 Mb downstream. The deletion was confirmed by PCR and Southern analyses, which detected no FMR1 allele. While microarray analysis showed three other CNVs, none are considered causative of disease. Thus the patient's phenotypic features are most likely due only to the small deletion affecting FMR1. The presence of moderate developmental and speech delay is in agreement with a previously published report [PMID:22890812 subject 3] of a male patient with a 265 kb deletion; however, at age 4, that patient had relative macrocephaly (75th centile). The couple's infertility is of interest in light of premature ovarian failure suffered by some Fragile X carriers; microarray analysis of both parents to assess the status of all CNVs is pending. This case further shows haploinsufficiency of FMR1 leads to altered dosage resulting in developmental delay. CMA is a sensitive method to detect copy number changes in FMR1.

2666F

Microcephaly associated with duplication of chromosome 15q24. *Y. Wang[1], Y. Zhou[2], D. Shrestha[1], E. Carter[1]*. 1) Dept Pathology, Univ South Alabama, 600 Clinic Dr. TRP IV, 278, Mobile, AL 36688; 2) Dept. of Biology, Baylor University, Waco, TX 76706.

An 11-month-old African American female was born with microcephaly, and she has developmental delay. The physical examination shows an OFC of 43.5 cm (2nd percentile); height is 76 cm (90th percentile); and weight 11.9 kg (above the 95th percentile). The features previously noted for a child with craniofacial dysmorphism including narrow bitemporal diameter and sloping forehead with associated microcephaly. She has no associated anomalies or significant dysmorphisms, otherwise. A brain MRI revealed evidence of simplified gyral pattern with shallow sulci and a fewer gyri as well as thin corpus callosum. These findings are consistent with the diagnosis of microcephaly with simplified gyral pattern (MSGP). A detailed analysis of the constitutional chromosomal changes in the patient was conducted by microarray comparative genomic hybridization (array CGH), and linkage analysis using 6 short tandem repeat (STR) markers on chromosome 15. Array-CGH revealed microduplication with 1.265 Mb on chromosome 15q23-q24.1 (chr15: 72023070-73288923 hg19 coordinates). Linkage analysis confirms this duplication. Human chromosomal region 15q24-26 harbors a high density of chromosome-specific duplicons. Consistent and recognizable clinical phenotypes of microduplication for distal 15q are minor craniofacial anomalies, congenital heart disease, mental retardation, and genital anomalies. Clinical phenotypes may vary in patients, depending on the size and location of duplication portion of chromosome 15q. Duplications in this region had been reported to be associated with neurodevelopmental disorders. However, microcephaly or brain malformation had not been reported in the 15q24 microduplication syndrome. This patient presents microcephaly and developmental delay caused by 15q24 microduplication.

2667W

Elevation of Insulin-Like growth Factor Binding Protein 2 Level in Pallister-Killian Syndrome: Implications for the Postnatal Growth Retardation Phenotype. *K. Izumi[1,2], E. Kellogg[1], M. Kaur[1], A. Wilkens[1], I. Krantz[1]*. 1) Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Research Center for Epigenetic Disease, Institute of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo, Japan.

Pallister-Killian syndrome (PKS) is a multi-system developmental disorder caused by tetrasomy 12p that exhibits tissue-limited mosaicism. The spectrum of clinical manifestations in PKS includes craniofacial dysmorphia, cleft palate, ophthalmologic, audiologic, cardiac, musculoskeletal, diaphragmatic, gastrointestinal, gentiourinary, cutaneous anomalies, intellectual disability and seizures. Probands with PKS often demonstrate a unique growth profile consisting of macrosomia at birth with deceleration of growth postnatally. Since the insulin-like growth factor signaling (IGF) pathway plays a critical role in normal growth physiology, we hypothesized that dysregulation of IGF pathway is involved in the pathogenesis of PKS. Genome-wide expression array analysis performed on cultured skin fibroblasts from PKS probands revealed significantly elevated expression of insulin-like growth factor binding protein 2 (IGFBP2). The amount of IGFBP2 secreted from cultured skin fibroblast cell lines was measured. About 60% of PKS fibroblast cell lines secreted higher levels of IGFBP2 compared to the control fibroblasts, although the remaining 40% of PKS samples produced comparable level of IGFBP2 to that of control fibroblasts. Serum IGFBP2 level was measured in PKS probands, and in 40% of PKS probands, IGFBP2 level was elevated. No correlation was seen between mosaic ratio and the IGFBP2 level. PKS probands with elevated IGFBP2 manifested with severe postnatal growth retardation. IGFBPs are the family of related proteins that bind IGFs with high affinity, and IGFBPs are typically thought to attenuate IGF action, and mouse models of IGFBP2 overexpression demonstrate a postnatal growth attenuation phenotype. Hence, we suggest the possibility that elevated IGFBP2 levels might play a role in the growth retardation phenotype of PKS. Since IGFBPs functions in an endocrine/paracrine fashion, dysregulated IGF signaling pathway should manifest its effects not only on tetrasomic cells, but also on the chromosomally normal cells in PKS probands.

2668T

Aetiology of Moderate Mental Retardation. *A. KUMAR, M. VASHIST.* DEPARTMENT OF GENETICS, M.D.UNIVERSITY, ROHTAK, INDIA.

Mental retardation is a common disorder which imposes a large medical, psychological and social burden .It is grouped into four classes on the basis of IQ scores by ICD-10. Individuals with IQ (35-50) were placed in the moderate mental retardation (MMR) category. Two hundred moderately mentally challenged patients were studied from Pandit Bhagwat Dayal Sharma University of Health Sciences, Rohtak and other rehabilitation centers across Haryana. In our study the outdoor patients reaching the local medical care services were between age group from 1 month of child to 50 years. Age group showed peak values below 16 years and a sharp decline after 50 years of age. Nearly 80% excess of affected males in the present study strongly suggested contribution of non specific X linked mental retardation. Higher percentage of males can also be an indication of the community based bias towards a male child. There were 44% cases from lower socio economic classes like agriculture laborers, construction laborers and other daily wages workers. 72% cases of urban population with lower socio-economic status is an indication of some certain specific risk factors associated with moderate mental retardation. Poverty and nutritional deprivation are environmental conditions which also seem to be relevant predisposing risk factors. Various Non Government Organization and Social Welfare Departments must be intervened to produce more impactful programs in addition to existing schemes in the State of Haryana. The initiation of various health care programmes may limit occurrence of moderate mental retardation. . Retrospective analysis revealed prenatal risk factors in 60% patients, neonatal in 40.35% and postnatal risk factors in 30.53% MR patients. Ninety four percent MR patients showed delayed developmental history. Malnutrition at prenatal stage was the main risk factor.

2669F

A dyslexia case with de novo der(14)t(Y;14)(q10;q10). *E. Kirat, G. Güven, M. Seven, M. Özen, E. Yosunkaya, H. Ulucan.* Medical Genetics, Istanbul University Cerrahpaşa Medical Faculty, Fatih, Istanbul, Turkey.

Dyslexia is a neuropsychiatric syndrome which has historically been called as ‘congenital word blindness’ and contemporarily it is defined as developmental analogue of acquired selective impairment in reading and writing ability due to neuronal damage in certain regions of the brain. The etiology of dyslexia, whether it is a single trait or a cluster of subtypes which are likely to involve different subsets of genes, is an ongoing debate. Also in some of chromosomal abnormalities like XYY syndrome, dyslexia could be a symptom. XYY syndrome is a chromosomal aneuploidy with a predisposition to learning disability, tall stature, behavior disorder and long ear. Here we report a case that has a minor facial dysmorphism and dyslexia with an extra Y chromosome, translocated to the 14th chromosome with a karyotype of 46, XY, der(14) t (Y;14)(q10;q10). The relationship between this translocation and this XYY-like phenotype is under investigation.

2670W

Computer-aided facial recognition of individuals with Angelman Syndrome.
L. Wolf[1,2], W.H. Tan[3,4], L. Karlinsky[1], M. Shohat[1,2,5], L.M. Bird[3,6]. 1) FDNA Ltd., Herzlyia, Israel; 2) Tel Aviv University, Tel Aviv, Israel; 3) 3.NIH Rare Diseases Clinical Research Network~Angelman, Rett, & Prader-Willi Syndromes Consortium, USA; 4) Division of Genetics, Boston Children's Hospital, Boston, MA, USA; 5) 5.Schneider Children's Medical Center of Israel, Raphael Recanati Genetics Institute, Rabin Medical Ctr and Felsenstein Medical Research Center, Petah Tikva, Israel; 6) Division of Genetics/Dysmorphology, Rady Children's Hospital San Diego, San Diego, CA, USA.

Angelman syndrome (AS) is a neurodevelopmental disorder characterized by severe developmental delay, intellectual disability, speech impairment and a unique behavior with an inappropriate happy demeanor. AS patients also commonly present unusual facies that may include deep set eyes, prognathism, thin vermilion of the upper lip, wide mouth, and strabismus.

In this study we examined whether a computer-based dysmorphologic analysis can help distinguish between AS patients and non-AS patients. We used a collection frontal facial images of genetically and clinically verified AS cases (n=210) and compared them to two separate control sets: (i) normal individuals (n=520); and (ii) individuals affected with other syndromes with dysmorphic facial features (n=808). The latter control set is added to verify that the system distinguishes not only between dysmorphic and non-dysmorphic populations, but can also identify the specific facial abnormality associated with AS.

A novel facial dysmorphology analysis system based on 2D photographs (FDNA®) was used. The system is fully automatic and starts by localizing hundreds of facial fiducial points and taking various measurements. The final classification is based on these measurements, as well as on a global "gestalt" detector that estimates the probability of the subject having AS based on the appearance of the entire facial image. A cross validation technique was used to estimate the recognition capability of the computer system. At each one of 20 rounds, the data was split randomly to training and testing data, each comprising of 50% of the individuals.

The system was able to distinguish AS from normal controls with 95.67% specificity and 95% sensitivity, but it was able to distinguish AS from dysmorphic controls with 98.02% specificity and 95% sensitivity. The gestalt of the face as captured by multiple local patterns of facial texture contributed significantly to the correct recognition of the individuals affected with AS.

Overall, we have demonstrated that computer-based analysis can be successfully to support the correct recognition of patients affected with AS.

2671T

Computer-aided detection of Down syndrome from facial photography.
Q. Zhao[1], K. Rosenbaum[2], D. Zand[2], L. Kehoe[2], R. Sze[1,3], M. Summar[2], M. Linguraru[1]. 1) Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington, DC; 2) Division of Genetics and Metabolism, Children's National Medical Center, Washington, DC; 3) Department of Radiology, Children's National Medical Center, Washington, DC.

Purpose To develop an automatic computer-aided detection (CADe) tool from simple facial photography to allow for instant, remote and accurate diagnosis of Down syndrome. **Methods** A method of automated detection of Down syndrome was developed based on 2D facial image analysis. A dataset consisting of 130 frontal facial photographs with 50 Down syndrome pediatric patients and 80 healthy individuals was collected with a variety of cameras and lighting conditions. The subjects are from multiple ethnicities including 98 Caucasian, 20 African American and 12 Asian and both genders (88 males and 42 females). The age of patients varies from 0 to 3 years. After image acquisition, 44 clinically-relevant anatomical landmarks are defined in the face region. A hierarchical statistical face model is built to locate these landmarks automatically on patient images. Patients of Down syndrome present with special morphology that relates both to geometry (upward slanting eyes and small nose) and texture (flattened philtrum and prominent epicanthal folds). To analyze the facial information, geometric and texture features are extracted. Geometric features are defined via interrelationships among anatomical landmarks to incorporate morphological clinical criteria. There are a total of 27 geometric features, including horizontal and vertical distances normalized by the face size, and corner angles between landmarks. The 132-dimensional local texture features are extracted from square patches around each inner facial landmark to describe the pattern of the facial region such as epicanthal folds and philtrum. Then the most representative 97 features are selected by maximizing the area under ROC curve (AUC) with respect to diagnosis accuracy. Finally, a support vector machine classifier is employed to discriminate Down syndrome and the healthy group by analyzing the selected features. The performance is evaluated using accuracy, sensitivity and specificity with a leave-one-patient-out strategy. **Results** For Down syndrome detection, the CADe system achieved 95.9% accuracy, 100% sensitivity and 93.6% specificity. The AUC was 0.99. **Conclusion** The automatic CADe system demonstrated high accuracy for detecting Down syndrome from simple home photography of variable quality. While access to modern testing for Down syndrome remains limited by geography, cost and physician access, this tool could potentially create an affordable solution for doctors worldwide to detect genetic disorders.

2672F

Tetraploid/diploid mosaicism as a potential cause of hypospadias.
J.C. Giltay[1,3], A.J. Klijn[2,3], M. van Breugel[1,3], L. van der Veken[1,3], R. Hochstenbach[1,3]. 1) Department of Medical Genetics; 2) Pediatric Renal Center; 3) University Medical Center Utrecht, Netherlands.

Both environmental and genetic factors have been implicated in the etiology of hypospadias. Environmental factors include placental insufficiency and pre-existing diabetes, genetic factors include several genes involved in the formation of male external genitalia. Monogenic etiology has mainly been proposed in cases with posterior hypospadias whereas the functional consequences of the mutations identified usually remain unclear. Since in the vast majority of the published reports genomic DNA isolated from blood has been used we hypothesized that mosaic chromosomal abnormalities present in genital tissues might be responsible for hypospadias in some cases. Thus, in a patient with penoscrotal (posterior) hypospadias, micropenis and scrotal testes, who was initially suspected to have partial androgen insensitivity syndrome (PAIS) we cultured fibroblasts from a genital skin biopsy, obtained during surgical correction of the hypospadias. Karyotyping revealed a 92,XXYY/46,XY mosaicism in 53 and 43 cells respectively. Interfase fluorescence in situ hybridization (FISH) with probes for chromosomes 18, X and Y confirmed this distribution of tetraploid and diploid cells. SNP array analysis on DNA from these genital cells gave no indication of chimerism. Karyotyping and/or interfase FISH in blood lymphocytes (n= 600), buccal mucosa cells (n=100) and cells from a urine sediment (n=100) was consistent with 46,XY in all cells. Tetraploidy mosaicism is an extremely rare chromosomal abnormality reported in at least 15 live born patients with growth retardation and developmental delay. Only one of them had renal/urinary tract abnormalities (Stefanova et al 2010). Complete tetraploidy is even rarer and reported in 9 live born children who had a very limited life expectancy (Stefanova et al 2010). In at least 7 out of 9 cases abnormalities of the urogenital tract were reported so these abnormalities seem to be a near consistent finding in tetraploidy. The lack of these abnormalities in patients with mosaicism may be due to ascertainment bias since tetraploid cells in our patient were only found in genital skin fibroblasts. We conclude that tetraploid/diploid mosaicism is a rare cause of hypospadias and perhaps other congenital anomalies of the urogenital tract which can only be diagnosed if cells from these organs are karyotyped.

2673W

Domains of histone marks in monozygotic twins discordant for trisomy 21.
X. Bonilla[1,2], A. Letourneau[1,2], F.A. Santoni[1], M.R. Sailani[1,2], M. Guipponi[3], C. Gehrig[3], S.E. Antonarakis[1,2,4]. 1) Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospitals of Geneva, Geneva, Switzerland; 2) National Center for Competence in Research Frontiers in Genetics, Geneva Switzerland; 3) Department of Genetic Medicine and Laboratories, University Hospitals of Geneva, Switzerland; 4) iGE3 institute of Genetics and Genomics of Geneva, Switzerland.

The use of samples from twins offers the opportunity to study epigenetic modifications that may be associated with phenotypic discordance. The availability of samples of monozygotic twins discordant for trisomy 21 has allowed us to study the effects of the trisomy in gene expression and genome organization in a unique model free of inter individual variability. When the expression level of all protein-coding genes was compared between fibroblasts of the twins, we found that the differential expression is organized in chromosomal domains with regions of upregulated or downregulated expression in the T21 twin. These domains correlated well with previously described replication domains and lamina associated domains (LADs). However we demonstrated that LAD topology is not altered in T21 cells, suggesting the possibility that the differences in expression between the twins are the result of changes in histone marks density or DNA methylation. In order to test the eventual effect of histone methylation on differential gene expression, ChIP seq for H3K4me3 was performed in the twins' fibroblasts. HTS libraries for input and IP chromatin of both twins were sequenced in a HiSeq2500 36bp single read, 41 million reads per sample on average. Peaks were called using HOMER. 18,895 and 22,017 peaks were identified for the T21 and the euploidic twin, respectively. The twins have 18,219 peaks in common. Those significantly enriched in the euploidic (46) are within or near genes related to embryonic development and morphogenesis according to DAVID. Notably, 3,788 peaks are unique to the euploidic twin. They are significantly enriched within or near genes involved in morphogenesis, axogenesis, axon guidance and neuron morphogenesis. Furthermore, we estimated the fold change difference in H3K4me3 density between the twins for each gene. Remarkably, we found a strong correlation between the H3K4me3 differential distribution and the chromosomal gene expression domains. We have studied the DNA methylation differences between the twins and they do not correlate with expression domains. These results suggest that the differences in H3K4me3 levels and not the LADs topology or overall DNA methylation are likely responsible for the changes in gene expression pattern between the twins. More histone marks are currently being investigated to better understand the dynamics of gene expression dysregulation.

2674T

Wolf-Hirschhorn syndrome: natural history into adulthood. A preliminary study of twenty-one individuals. *A. Battaglia[1], V. Doccini[1], T. Filippi[1], A. Lortz[2], JC. Carey[3].* 1) Dev Neurosciences, Stella Maris nst/Univ Pisa, Pisa, Italy; 2) 4p- Support Group, USA; 3) Division of Medical Genetics, Dept. of Pediatrics, University of Utah Health Sciences Center, Salt Lake City, UT, USA.

Information on long-term survival in Wolf-Hirschhorn syndrome (WHS) is very limited. Only five adult WHS are described in the medical literature, with age range between 24 and 37 years. We report on twenty-one WHS individuals (17 females; 4 males), aged between 21ys 7 m and 48ys 9m. Fourteen were personally observed at the Stella Maris Clinical Research Institute by A.B., and 7 were recruited from the 4p- Support Group-USA. All are described as being happy, friendly, outgoing adults; liking to be around family and friends. 11/21 are close to total care with needs to be assisted in feeding, diapering, bathing, dressing/undressing, and when walking needing assistance on unlevel ground or transitioning from carpet to tile. 9/21 are partly self independent, being able to feed and dress themselves, but needing supervision and some assistance with personal hygiene, and, at times, with walking on unlevel ground. 1/21 is fully independent. All had variable degrees of developmental delays, with a moderate-severe cognitive deficit. 18/21 enjoy a good health, whereas 1/21 has diabetes type II, diagnosed at age 13; 1/21 has Raynaud disease, diagnosed at age 18; and 1/21 had esophagitis at age 19. All but one are seizures free. All are enrolled in a personalized rehabilitation program. Knowledge of the medical and developmental aspects of adults with WHS will inform health supervision.

2675F

Prader-Willi syndrome and oculocutaneous albinism due to a 5;15 translocation and hemizygous OCA2 mutation. *A.C.E. Hurst[1], C.R. Haldeman-Englert[2], T.H. Stamper[2], M. Hanna[3], M.J. Pettenati[2], P.P. Koty[2].* 1) Dept. of Pediatrics, Wake Forest School of Medicine , Winston-Salem, NC; 2) Department of Pediatrics, Section on Medical Genetics, Wake Forest School of Medicine, Winston-Salem, NC; 3) Wake Forest School of Medicine, Winston-Salem, NC.

Prader-Willi syndrome (PWS), characterized by infantile hypotonia and feeding difficulties followed by later onset of hyperphagia and obesity with delayed motor milestones, distinctive behavioral phenotype, hypogonadism, short stature, and characteristic facial features, is caused by the absence of imprinted genes on the paternal 15q11.2-q13 region. Oculocutaneous albinism type 2 is diagnosed based on the clinical findings of skin and hair hypopigmentation, nystagmus, and reduced iris and retinal pigment with decreased visual acuity. This condition is due to mutations of OCA2 (previously called the P gene), located within the same region of 15q11.2-12. These two conditions have been described together previously in a limited number of patients. We present here an African-American female, born at 41 weeks gestation, who was noted at birth to have decreased cutaneous pigment, fine curly yellow hair, blue eyes, and hypotonia. Her karyotype revealed a 5;15 translocation: 45,XX,der(5)t(5;15)(q35.3;q11.2). FISH analysis for a PWS deletion was positive, and confirmed the deletion to be on the translocated chromosome 15. A cytogenomic microarray analysis clarified the loss of chromosome 15 material, showing a heterozygous 5.8-Mb deletion of 15q11.2q13.1 that involved OCA2 (chr15:22,770,421-28,547,716; Build 37/hg19), with no gain of material from chromosome 5. Sequencing of OCA2 revealed a previously reported hemizygous mutation: c.819_822delCTGGinsGGTC (p.Asn273_Trp274delinsLysVal). Parental studies are pending, but it is likely that the translocation involves the paternal chromosome 15 and is de novo, and the hemizygous OCA2 mutation involves the maternal chromosome 15 and is carried by her mother. Review of the literature revealed three reports of a chromosome 5;15 translocation associated with PWS, but none of those patients had albinism. Therefore, our patient's findings are a novel cause of the PWS plus albinism phenotype.

2676W

Comparing serum calcium tests in 22q11DS and other genetic conditions: ionized vs. total calcium. *E. Chow[1,2], T. Leung[1], M. Torsan[1], C. Stefan[3].* 1) Clinical Genetics Service, Centre for Addiction & Mental Health, Toronto, ON, Canada; 2) Department of Psychiatry, University of Toronto; 3) Clinical Laboratories and Diagnostics Services, Centre for Addiction & Mental Health, Toronto, ON, Canada.

Background: Hypocalcemia is common in 22q11.2 deletion syndrome (22q11DS) and can occur or recur throughout the life, making its detection an important part of routine monitoring. While the more expensive serum ionized calcium test is often recommended, no scientific research has been conducted on whether it has advantages over the cheaper and more easily performed total serum calcium test for patients with 22q11DS or other genetic conditions. This study aims to compare the detection of hypocalcemia by these two tests. Methods: A simultaneous set of serum calcium tests (pH-normalized ionized calcium and albumin-corrected total calcium) was performed 47 times in 26 individuals (9 male) with 22q11DS and 17 times in 16 individuals (9 male) mostly with other genetic conditions. All subjects were patients at the Clinical Genetics Service of CAMH in Toronto. Mean age at testing was younger for the 22q11DS group [24.9 (SD 6.6) years vs. 31.5 (SD 8.2) years, p=0.006] but the two groups did not differ in the proportion of male subjects or subjects with psychosis. Measured hypocalcemia was defined as having one or both calcium test results below the lower limit of the reference range for the test. The number of times results from the two tests were discordant for measured hypocalcemia was recorded and compared. Results: Rate of measured hypocalcemia was higher in the 22q11DS group (36.2% vs. 11.8%), but at a trend level (p=0.07). For the 22q11DS group, test results were discordant in 9 of the 17 instances of measured hypocalcemia: 7 times only in the ionized calcium test and twice in the total calcium test. 22q11DS subjects had neuromuscular symptoms in 2 out the 7 times when only the ionized calcium was abnormal, but none when only the total calcium was abnormal. On the other hand, both instances of measured hypocalcemia in the other group were found only in one test, and both times in the ionized calcium test. Although neither of these two subjects was symptomatic, the one with Rett Syndrome had a significantly elevated PTH. Conclusions: The increased instances of hypocalcemia detected by pH-normalized ionized calcium appear to represent true hypocalcemia as subjects were more likely to have symptoms of hypocalcemia and/or an elevated PTH. Thus the pH-corrected ionized calcium test was a more sensitive and superior test in detecting hypocalcemia in 22q11DS and other genetic conditions than the albumin-corrected total calcium test.

2677T

Scoliosis and Vertebral Anomalies: Additional Abnormal Phenotypes Associated with Chromosome 16p11.2 Rearrangement. *H. Al-Kateb[1], G. Khanna[1], I. Filges[2], S. Kulkarni[1], M. Shinawi[1].* 1) Pathology and Immunology, Washington University School of Medicine, St. Louis, MO; 2) University of British Columbia, Vancouver, BC.

16p11.2 rearrangements are estimated to occur at a frequency of approximately 0.6% of all samples tested clinically and have been identified as a major cause of autism spectrum disorders (ASD), developmental delay, behavioral abnormalities, and seizures. Careful examination of patients with these rearrangements revealed association with abnormal head size, obesity, dysmorphism and congenital abnormalities. In this report, we extend this list of phenotypic abnormalities to include scoliosis and vertebral anomalies through detailed characterization of phenotypic and radiological data of seven new patients, 5 with 16p11.2 deletion and 2 with duplication corresponding to chromosomal coordinates: chr16: 29 528 190-30 107 184 (hg19) with a minimal size of 579 kb. We discuss the phenotypic and radiological findings in our patients and review 5 previously reported patients with 16p11.2 rearrangement and similar skeletal abnormalities. The results show male to female (M:F) ratio of 2:1, strikingly different from the known 1:10 ratio in the general population for idiopathic scoliosis. Interestingly, we found that males have more severe spinal deformities than females. We discuss the implications of these findings on the diagnosis, surveillance and genetic counseling of patients with 16p11.2 rearrangement.

2678F

MED13L haploinsufficiency in isolated delayed neuromotor development: further delineation of the phenotypic spectrum. *B. Callewaert[1], A. Dheedene[1], B. Menten[1], B. Delle-Chiaie[1], E. Snauwaert[2], K. Decaestecker[3], F. Roelens[4], O. Vanakker[1].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Ghent, Belgium; 2) Department of Pediatrics, Ghent University Hospital, Ghent, Belgium; 3) Department of Pediatrics, Stedelijk Ziekenhuis Roeselare, Roeselare, Belgium; 4) Department of Pediatrics, Heilig Hart Ziekenhuis Roeselare, Roeselare, Belgium.

A decade ago, a chromosomal translocation disrupting the MED13L (Mediator complex subunit 13-like) gene was found in a patient with transposition of the great arteries (TGA) and intellectual disability (ID). Three rare MED13L variants were subsequently identified in a cohort with isolated TGA. More recently, MED13L haploinsufficiency was detected in 2 patients with ID and conotruncal heart defects, while a copy number gain resulted in a ventricular septal defect, hypotonia and learning problems. We report 2 novel patients with a de novo intragenic MED13L deletion and further delineate the phenotype of this rare ID syndrome. Results: Both patients, respectively a 2-year old girl (patient 1) and a 3-year old boy (patient 2), presented with moderately delayed neuromotor development, generalized hypotonia, hypermetropia and a typical facial gestalt including a broad forehead, telecanthus, epicanthal folds, a bulbous nasal tip, rather small and low-set ears with a slight uplift of the earlobe. Staturoponderal evolution and head circumference were normal as was echocardiography. Brain MRI showed slightly delayed myelinisation in both patients. In addition, patient 1 had an episode of acute pyelonephritis. Abdominal ultrasound revealed a normal urogenital tract, but three small subdiaphragmatic liver cysts were noted. Patient 2 had a surgically corrected unilateral cryptorchidism. ArrayCGH analysis (180k Agilent array) revealed a 300kb deletion on chr. 12q24.21, deleting exons 2-22 of the MED13L gene in patient 1 and a 150kb deletion on chr. 12q24.21 deleting exons 2 and 3, likely resulting in a shift of the open reading frame in patient 2. Conclusion: We further document the clinical phenotype of intragenic MED13L deletions with moderate ID and consistent dysmorphic features. Particularly the morphology of the ear and earlobe should prompt MED13L analysis, even in the absence of cardiovascular defects. Hypermetropia in both patients suggests that refractive errors may also be part of the phenotypic spectrum. Besides this clinically recognizable ID syndrome, the mutation rate of the MED13L gene in patients with isolated neurodevelopmental delay remains to be determined.

2679W

Ear anomalies, mild intellectual disability, delayed growth and orodental development in a boy with a 5 Mb microduplication within 20q13.12q13.2. *J.R. Helle[1,2], M. Fannemel[2], T. Barøy[1,2], S. Axelsson[3], A.B. Karstensen[4], E. Frengen[1,2], D. Misceo[1,2].* 1) Department of Medical Genetics, University of Oslo, Oslo, Norway; 2) Department of Medical Genetics, Oslo University Hospital, Ullevål, Oslo, Norway; 3) TAKO Center, Lovisenberg Hospital, Oslo, Norway; 4) Department of Pediatrics, Akershus University Hospital, Norway.

Duplications within 20q13.12q13.2 are rare, and most imbalances previously reported were associated with deletions, resulting from an unbalanced segregation of translocation chromosomes from one parent. In these cases, the contribution to the phenotype of the duplication and the deletion was challenging to assess. More recently, isolated micro-duplications were reported in this region. However, mapping the boundaries and the gene content in these duplicated regions was difficult, mainly due to limited resolution of the investigation methods used, even in recent reports. We present a boy born post-term, presenting with large, cupped ears. He showed delay of developmental milestones, intellectual skills, growth and dental development. From childhood on he showed a mild intellectual disability, delayed loss of deciduous teeth and growth of permanent teeth, height and weight 2,5 standard deviations below mean and protrusion of the tongue. By the age of 20 years, dental status was normal. Array comparative genomic hybridization revealed a de novo 5 Mb duplication on chromosome 20q13.12q13.2, which was shown by FISH to be in situ. The duplication contains 30 RefSeq genes. To the best of our knowledge, the present micro-duplication is one of few duplications in the 20q13.12q13.2 region characterized at high resolution, allowing a precise identification of its gene content. We provide the first clinical report on the long-term development from birth until young adulthood in a patient with a 20q13.12q13.2 duplication, as followed by a multi-disciplinary team of clinicians. We compare the clinical findings in our patient with previously reported patients with overlapping duplications. Our patient's phenotype appears as relatively mild, considering the large number of genes duplicated. This contrasts with the severe phenotype in previously reported patients and suggests a high degree of clinical variability. Further clinical studies and expression studies will likely give information about the phenotypic impact of altered expression of specific genes in the 20q13.12q13.2 region.

2680T

A novel interstitial microdeletion in 2q37 refines critical region and candidate genes for microcephaly, myelination and developmental delay expressed in human neural progenitors. *J. Imitola[1], D. Khurana[2], A. Legido[2], K. Carvalho[2].* 1) Brigham and Women's Hospital Harvard Medical School; 2) 2Section of Pediatric Neurology, St Christopher Hospital for Children's, Drexel School of Medicine. Philaldelphia.

2q37 microdeletion syndrome is characterized by bone, cardiovascular alterations, neurodevelopmental delay, microcephaly and seizures. This syndrome is associated with loss of genetic material approximately 100 genes in the 2q37 band. However, the genes associated with neurodevelopmental phenotype in this syndrome are still unknown. A novel 496 kb deletion was discovered by whole genome array CGH in a patient who fulfilled the criteria for 2q37 microdeletion syndrome with additional features of hand wringing, toe walking and seizures. The abnormal segment contains 10 genes based on UCSC 2006 hg18, including SEPT2, FARP2, HDLBP, STK25. Pathway analysis of these genes revealed links to neural development,myelination, ciliogenesis and interaction with genes associated with neural function, gene expression analysis revealed that these genes are highly expressed in cortical plate, human neural progenitors areas in vivo and in vitro. Our report narrows the genomic region for 2q37 microcephaly, myelination and neurodevelopmental delay to 10 candidates genes and suggests that this segment may represent an important locus for human forebrain development. Haploinsufficiency of this region may lead to human microcephaly, seizures and neurodevelopmental delay.

2681F

Deep White Matter Brain Abnormalities in a Patient with Chromosome 15q11-q13 Deletion and Angelman Syndrome Phenotype. *N. Sekhri[1], G. Scharer[1,2].* 1) Dept. of Pediatrics, Genetics Center, Children's Hospital of Wisconsin, Milwaukee, WI; 2) Advanced Genomics Laboratory, Medical College of Wisconsin, Milwaukee, WI.

While Angelman syndrome (AS) is a neurodevelopmental disorder with a known genetic basis that is characterized by severe intellectual disability, lack of speech, ataxia, seizures, and specific behavioral patterns, leukodystrophy is not a typical finding. Subtle changes to white matter pathways detected by DTI (diffusion tensor imaging) have been described by *Peters et al, 2011*. We describe a 22-month old male patient with more overt leukodystrophy (detected on standard 1.5T brain MRI), childhood onset complex partial seizures and borderline microcephaly. MRI showed symmetric long TR hyperintensity in the periventricular and deep white matter involving the fronto-temporal and parieto-occipital white matter without evidence of abnormal enhancement. Diagnostic work-up initially focused on the white matter abnormalities and seizures; Angelman syndrome was not considered high in the differential diagnosis. Chromosomal microarray analysis was included in the work-up, but not methylation studies. Subsequently, the patient was found to have a 4.9 megabase deletion on microarray analysis (Affymetrix®6.0 oligo/SNP array) at chromosome 15q11.2-q13.1 encompassing the Prader-Willi syndrome (PWS)/(AS) region, while the remainder of genetic tests were negative. After review, the patient's behavioral pattern and clinical presentation was felt to be consistent with a diagnosis of AS. This case highlights a potential emerging problem in the approach to confirming a suspected genetic diagnosis. The availability of clinical exome/genome analysis and sequencing of disease-targeted gene panels based on Next-Generation Sequencing (NGS) technologies still leave gaps in the detection of common disorders due to methylation defects, triple repeat expansion or small CNVs. Clinicians should keep these limitations in mind when evaluating patients with varied genetic disorders.

2682W

Deletion 2q37: Cognitive-Behavioral Profiles, Developmental Trajectories, and IQ Related to Deletion Size. *G.S. Fisch[1], S.T. South[2,3], A. Rutherford[3], R. Falk[4], J. Carey[2].* 1) New York University Dept. Epidemiology & Health Promotion New York, NY; 2) University of Utah Dept. Pediatrics Salt Lake City, UT; 3) Institute for Clinical and Experimental Pathology ARUP Laboratories Salt Lake City, UT; 4) Medical Genetics Institute Cedars-Sinai Medical Center Los Angeles, CA.

2q37 deletion syndrome is typically described as encompassing, among other clinical features, mild to moderate intellectual disability, behavioral problems, short stature, hypotonia, characteristic craniofacial anomalies, and a high proportion of affected individuals with autism. Deletions involving 2q37 occur most often as a de novo mutation although instances of inversions and unbalanced translocations have been observed. We report on 8 cases with deletion of 2q37, the largest single multicenter study to date of individuals with this genetic anomaly. We examined clinical features, cognitive and behavioral abilities from 5 females and 3 males, previously diagnosed cytogenetically with del2q37, and ranging in age from 4 - 18 years at the time of their first cognitive-behavioral assessment. Each participant was administered a comprehensive neurocognitive battery consisting of the Stanford-Binet (4th Ed; SBFE), Vineland Adaptive Behavior Scales [VABS], Conners Parent Rating Scale [CPRS], Child Behavior Checklist [CBCL], and the Child Autism Rating Scale [CARS]. Of the 8 participants, 3 lacked expressive language and could not be assessed by the SBFE. Intellectual disability for the 5 tested with the SBFE ranged from severe to mild [IQ Range: 36 - 59], with relative strength in Quantitative Reasoning. Adaptive behavior composite scores from the VABS were much below adequate levels, ranging from a floor value ['19'] to 55. Based on the CPRS, 7/8 [87%] could receive a diagnosis consistent with ADHD. Scores on the CARS ranged from 22 [non-autistic] to 56 [extremely autistic]; 5/8 [63%] children received scores in the autism range. Deletion size based on microarray analysis using Cytoscan HD or Affymatrix 6.0 was determined for 4 subjects from whom DNA was obtained. Curiously, participants with the largest deletions, 10.1 Mb and 9.5 Mb, achieved the highest IQ and DQ scores; whereas, those with the smallest deletions, 7.9 Mb and 6.6 Mb, attained the lowest IQ and DQ scores. Previously, the CENT2 gene [now known as AGAP1] was thought to be a candidate gene for autism. In our study, the 3 individuals with the largest deletions were haploinsufficient for AGAP1 but not autistic. However, the subject with the smallest deletion in which AGAP1 was extant, was extremely autistic. Conventional doctrine relating deletion size and phenotype severity is discussed further.

2683T

Identification of 22q11.2 deletion in patients from adult congenital heart disease clinic - a missed burden in the transition care in Hong Kong. *B. Chung[1], P. Chow[1], A. Liu[1], P. Lee[1], K. Chan[2], M. Tang[2], E. Lau[2], Y.F. Cheung[1], K.T. Chau[1], Y.L. Lau[1].* 1) Paediatrics & Adol Med, The University of Hong Kong, Hong Kong, N/A, Hong Kong; 2) Prenatal diagnosis unit, Dept of O&G, Tsan Yuk Hospital, The University of Hong Kong.

22q11.2 deletion syndrome (22q11DS) is a common genetic diagnosis in patients with congenital heart disease (CHD). It is multi-systemic with both congenital and later-onset features with lifelong consequences. Variable clinical expression and limited awareness contribute to its under-diagnosis. With low childhood mortality, there is an increasing number of diagnosed/undiagnosed adults, posing a hidden challenge in the transition care for patients with CHD. Our objective is to determine the prevalence of 22q11DS in adult patients with conotruncal defects and to delineate their extra-cardiac manifestations. We enrolled patients through an adult CHD clinic by active screening, using fluorescence-PCR and FISH. We have recruited 156 with conotruncal defects in 2012-2013(on-going recruitment). Results: Eighteen patients are diagnosed with 22q11DS, which translates into a missed diagnosis of 1 in every 10 adults with conotruncal defects. Eleven had the cardiac diagnosis of tetralogy, 6 had pulmonary atresia and ventricular septal defect, 1 had interrupted aortic arch. Our approach of screening by molecular testing has increased the detection rate of 22q11.2DS by at least 4 times.Screening by cardiologists using dysmorphology assessment only may miss 40 percent of patients. Our cohort of patients with 22q11.2DS previously undiagnosed have less extra-cardiac manifestations that the other reported series, presumably due to lack of medical surveillance appropriate for patients with 22q11.2DS.Conclusion: This study will provide important information on the disease burden of 22q11DS and may highlight an important and actionable gap in the transitional care of patients with CHD.

2684F

Diagnosis of atypical 22q11.2 deletion and a recurrent Herpes Zoster virus infection in a man with T cell immunodeficiency. *L.P. Barragan Osorio, G. Giraldo, J.C. Prieto.* Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogota, Bogota, Colombia.

Individuals with the 22q11.2 deletion have a range of findings, including cardiovascular malformations consisting predominantly of conotruncal heart defects, palatal abnormalities characteristic facial dysmorphic features, learning difficulties, mild to moderate immune deficiency, hypoparathyroidism, among other. About 90% of patients with the 22q11.2 microdeletion have a common ~3 Mb deletion, whereas 7% of the patients have a smaller, proximally nested ~1.5 Mb recurrent deletion and other atypical deletions, nested, overlapping or adjacent to the typically deleted region (TDR). However, to date, only a limited number of atypical deletions in 22q11.2 have been reported and only a few are located distally to TDR. In this paper we report a male 47 years old patient who had been studied for primary immunodeficiency with recurrent Herpes Zoster infection which onset 9 years ago. The patient had a history of delayed developmental milestones, mild learning difficulty at school, hypothyroidism and unilateral sensorineural hearing loss. At the first consultation the patient display slight facial dysmorphism, including hooding of the lower lids, long pointed nose, velopharyngeal voice, left ocular Herpes Zoster infection with fourth nerve palsy and mild cognitive deficiency. He had a 46, XY normal karyotype. Immunological profile reported remarkable decrease of immunoglobulin and CD3, CD4 and CD8 deficiency. His IQ total score was 72.These finding suggested a DiGeorge Syndrome, therefore was performed an array-based comparative genomic hybridization (aCGH). It reported a ~1.8 Mb deletion within chromosomal band 22q11.21-22q11.23 that matched with distal 22q11.2 microdeletion. This result was confirmed with an MLPA analysis (kit SALSA MLPA probemix P250 B1 DiGeorge).Here we report one of the few atypical microdeletions in the distal 22q11.2 deletion region with a barely phenotype and an important cellular immunodeficiency leading to recurrent Herpes Zoster infection that has not been described previously. The phenotypic spectrum of 22q11.2 deletion depends on multiple dosage-sensitive genes located in this region. Non-overlapping atypical deletions have showed significantly overlapping phenotypes, suggesting either, a positional gene effect in 22q11.2, the involvement of several candidate genes or a common developmental pathway.

2685W

Clinical characteristics of a newly identified microdeletion syndrome involving chromosome region 16q22.1. *L. Dupuis[1,3], M. Helal[1,3], D.J. Stavropoulos[2,3], R. Mendoza-Londono[1,3].* 1) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) University of Toronto, Toronto, Ontario, Canada.

We describe two unrelated cases of a de novo microdeletion of chromosome 16 at band q22.1 that present with distinctive and recognizable features. Patient 1 was first evaluated at 17 months of age because of a history of intrauterine growth retardation, failure to thrive, developmental delay, and congenital heart defect. At her last assessment at 4 years of age, the patient met criteria for autism spectrum disorder, had significant global intellectual disability and was G-tube fed. She also had hearing and visual impairment. Microarray analysis (44,000 oligos platform) revealed an estimated 0.444 Mb deletion of chromosome 16q at band 22.1, which included 25 RefSeq genes. Patient 2 presented at birth with microcephaly, dysmorphic features and hypoplastic deep set nails. At 14 months of age he had mild expressive language delay and distinctive facial features. Microarray analysis (180,000 oligos platform) revealed a larger but overlapping deletion of chromosome 16q22.1. The deletion was approximately 0.833 Mb in size and included 39 RefSeq genes. In both patients, the deletion comprised 4 OMIM genes (*HSF4, HSD11B2, NOL3 and AGRP*) (UCSC Genome Browser, hg 18). Mutations in *HSF4* have been implicated in autosomal dominant and recessive forms of cataracts. Mutations in *NOL3* have been shown to cause adult onset familial cortical myoclonus syndrome in the Menonite population. *HSD11B2* mutations and polymorphisms have been reported in apparent mineralocorticoid excess syndrome and as a risk factor for hypertension. *AGRP* is involved in the regulation of body weight. A thorough review of the medical literature revealed that these two cases are the first report of this particular microdeletion of chromosome region 16q22.1. The patients presented with distinct and recognizable symptoms characterized by failure to thrive and dysmorphic facial features. The two patients had between 25 and 39 genes deleted in this region but less than 4 have been associated with recognized disorders. Deletion of the disease-associated genes does not explain all the features present in our patients. Further work to understand the function of the remaining deleted genes will provide further insight into their role in this newly identified microdeletion syndrome.

2686T

Microarray Comparative Genomic Hybridization analysis (array-CGH) as a diagnostic tool for the investigation of patients with ID/DD/ASD with or without MCA. *M. TZETIS, S. KITSIOU-TZELI, H. FRYSIRA, V. OIKONOMAKIS, K. KOSMA, K. GIANNIKOU, A. SYRMOU, E. KANAVAKIS.* Medical Genetics, University of Athens, Athens, Attica, Greece.

Clinical characteristics of patients are not always related to specific syndromes. Array comparative genomic hybridization (aCGH) is used to detect submicroscopic copy number variants (CNVs) within the genome not visible by conventional karyotyping. The clinical application of aCGH has helped the genetic diagnosis of patients with unexplained developmental delay (DD)/ intellectual disability (ID), autism spectrum disorders (ASD), with or without multiple congenital anomalies (MCA). We have implemented aCGH using the 244K & 4x180K Agilent platform, since 2008 on 581 patients with various degrees of DD/ID, seizures, ASD, MCA and normal previous conventional karyotype. Many of the patients had also received a variety of other genetic tests (FRAX, RETT, single FISH tests or metabolic screens), which were normal. Clinically significant submicroscopic imbalances with aCGH were detected in 164 (28.15%) patients. CNV sizes ranged from 10Kb to>10Mb (155 deletions & 67 duplications). In 120/164 patients one CNV contributed to their phenotype while in 44 two or more explained their clinical phenotype (34: with 2CNVs, 9: with 3CNVs, 1: with 4CNVs and 1: with 5CNVs). A total of 51 patients presented with ASD, some of which also presented: seizures (8/51), hypotonia (1/51), obesity (2/51), mental retardation (9/51), hearing loss (1/51) and dysmorphic features (2/51). CNVs detected in ASD patients were in loci at high risk for autism: 2q37.2, 2p16.3, 5p15.32p14.3, 7p22.3, 10q26.3, 12q24.22q24.3, 15q11.2q11.3, 16p11.2, 18q22.3q23, 21q22.3, Ψq21.31. Genes associated with autism or ASD which were identified included: NRXN1, SHANK3, DOCK8, ZNF92, ASMT, HSFX1, KCNH7, CHRFAM7A, CHRNA7, KCND2, CNTNAP3, MAA, MAOB, STS, VCX. In total 64 different microdeletion and 22 microduplication syndromes were detected, the majority (80%) found de novo in the patients. We were able to define the smallest region of overlap (SRO) for known syndromes, detect a novel microduplication (3q13.2q13.31) and confirm haploinsufficiency of KANSL1 in 14 patients as responsible for Koolen de Vries syndrome albeit with reduced penetrance in 3/14. aCGH is proving to be a powerful tool for the identification of novel chromosomal syndromes, allowing thus accurate prognosis and phenotype-genotype correlations.

2687F

Mortality in patients with 22q11 microdeletion syndrome. *G.M. Repetto[1,2], M.L. Guzman[1], M. Palomares[3,4], G. Lay-Son[1,2], C. Vial[1], K. Espinoza[1], H. Loyola[1].* 1) Center for Human Genetics, Clin Alemana Univ Desarrollo, Santiago, Chile; 2) Hospital Padre Hurtado, Santiago, Chile; 3) Hospital Dr. Luis Calvo Mackenna, Santiago, Chile; 4) Fundación Gantz, Santiago, Chile.

22q11 microdeletion syndrome (22q11DS) is one of the most common genomic alterations in humans, having an estimated incidence of 1/4000 live births. It accounts for approximately 10% of cases of congenital heart disease (CHD) and is also a recognizable cause of palatal anomalies, learning disabilities, immunodeficiency and psychiatric disease. Early mortality has been attributed to the presence of CHD, and diminished life expectancy has also been described in adults. The purpose of this study was to analyze demographic and clinical factors associated with mortality in a large cohort of 22q11DS patients from Chile. Through surveys of clinical cytogenetics laboratories, we found that 419 patients have had molecular confirmation of the deletion from years 1998 to 2012. Their current ages range from 6 months to 50 years. Of them, 58 (13.8%) are deceased, 25 males and 33 females. The average age at death was 1.15 y (1 sd =4.52), with a median of 4 months and ranging from newborns to 32 years. The primary cause of death was recorded in 43 cases, of which CHD and sepsis accounted for 21 cases each (48.9%). The main factors associated with mortality were the presence of CHD resulting in an OR of 5.11 (95% CI of 2.11-15.23) and airway malacia with an OR of 13.44 (95% CI 1.81-151.31). A non-significant trend for association was found with palate abnormalities (OR 0.37, 95% CI 0.05-2.84) or with the presence of gastrointestinal (OR 6.53, 95% CI 0.72-79.23) or renal (OR 8.57, 95% CI 0.16-95.44) anomalies. Patients with 22q11 are at an increased risk for early mortality, mostly due to CHD or infection. In addition to CHD, abnormalities of the airway are a relevant risk factor for early death. Funded by Fondecyt-Chile grants #1061059/1100131/1130392.

2688W

Array-CGH: known syndromes, private variants and new syndromes. *E. Biamino[1], E.F. Belligni[1], E. Di Gregorio[2], C. Molinatto[1], A. Calcia[2], A. Mussa[1], E. Grosso[2], A. Zonta[2], M.T. Ricci[2], L. Sorasio[1], G. Mandrile[3], G. Gai[2], V. Naretto[2], P. Pappi[2], F. Talarico[2], A. Guala[4], P. Vigliano[5], G. Restagno[1], E. Savin[2], N. Migone[2], G.B. Ferrero[1], M. Silengo[1], A. Brusco[2].* 1) Pediatrics Dept, University of Turin, Turin, Turin, Italy; 2) Dept of Genetics, Biology and Biochemistry, University of Turin, Italy; 3) Dept of Genetics, AO S. Luigi Gonzaga, Orbassano, Turin, Italy; 4) Dept of Pediatrics, Ospedale Castelli, Verbania, Italy; 5) Dept of Pediatric Neurology, Ospedale Martini ASL2, Turin, Italy.

Array comparative genomic hybridization (array CGH) is widely used as a first-tier clinical diagnostic test in unexplained developmental delay/intellectual disability (DD/ID) and/or congenital anomalies cases. We have tested by array-CGH (Agilent 60 K) a cohort of 513 patients affected by DD/ID associated with congenital malformations and/or dysmorphisms. The detected CNVs, unlisted in the Database of Genomic Variants, were assigned to one of the following four groups: 1) known microdeletion/microduplications syndromes or CNVs encompassing a disease-causing gene: 46/513 patients (8.9%), including del 22q11.21 (n=6), del/dup 15q13.3 (n= 2+2), del/dup 16p11.2 (n=3+4), del 17p21.31 (n=2), del 1p36 (n=2), del 1q21.1 (n=1), del 3q29 (n=1), del 7q11.23 (Williams-Beuren syndrome, n= 2), dup *NSD1* gene, del *CREBBP* gene (Rubinstein-Taybi syndrome), del *NF2* gene, del *PITX2* (Rieger syndrome), del *ANKRD11* gene (KBG syndrome); 2) deletion/duplications extending more than 7 Mb (range 7.5-29 Mb): 23/513 patients (4.5%), including 2 cases of mosaicism (>30%). The remaining CNVs involved chromosomal regions of challenging study by standard karyotype. Five of 23 cases derived from a parental balanced translocation; 3) 'likely pathogenic' deletion/duplications (150 Kb-1.4 Mb): 15/513 cases (2.9%), resulting de novo in 5/15 and inherited in 10/15. Interestingly we detected an atypical deletion of 1,3 Mb in 3q29 (encompassing 13 known genes), a deletion of 500 Kb in 5q12.3 (5 known genes) and a deletion of 1,4 Mb in 16p13.12p12.3 (14 known genes); 4) CNVs of unclear relevance (110-900 kb): 52/513 (10.1%). These results restate the effectiveness of array-CGH analysis as first-step test in patients presenting with complex phenotypes, not only to reach a clinical diagnosis but also to identify new candidate genes for DD/ID.

2689T

Heterozygous microdeletion of 16q covering *SALL1* and *RPGRIP1L* could be a novel contiguous gene syndrome with renal impairment. *N. Morisada[1], M. Taniguchi-Ikeda[1], S. Ishimori[2], T. Ninchoji[1], H. Kaito[1], K. Nozu[1], M. Adachi[2], Y. Takeshima[1], T. Sekine[3], K. Iijima[1].* 1) Pediatrics, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan; 2) Pediatrics, Kakogawa West City Hospital, Kakogawa, Hyogo, Japan; 3) Pediatrics, Toho University Ohashi Hospital, Tokyo, Japan.

Background. Microdeletion of 16q is a rare chromosomal abnormality. Nonsense mutation of *SALL1* at 16q12.1 causes Townes-Brocks syndrome (TBS, OMIM #107480) with branchiogenic anomalies, imperforate anus, and mild to severe renal impairment. The patients with *SALL1* heterozygous entire gene deletion develop milder clinical symptoms. Renal impairment in the patients with *SALL1* heterozygous entire deletion has not been reported. Homozygous mutation of *RPGRIP1L* at 16q12.2 cause Joubert syndrome (OMIM #611560) with nephronophthisis. There are few case reports with the deletion of both genes. Moreover, these several reported cases did not develop renal impairment. **Case reports.** Case 1: A 15-year-old Japanese boy had hearing loss, external ear malformation, mild developmental impairment and moderate renal impairment. Although his symptom was compatible with the category of branchio-oto-renal syndrome (OMIM #113650), no mutation was detected in *EYA1*. An array comparative genome hybridization (aCGH) analysis revealed a 5.2Mb deletion in 16q including *SALL1* and *RPGRIP1L*. Case 2: A 13-year-old Japanese boy had multiple anomalies such as bilateral renal hypoplasia leading to end stage renal failure, severe developmental delay, imperforate anus, congenital heart defect, and low set ears. He was diagnosed as TBS. Fluorescence *in situ* hybridization (FISH) analysis revealed entire gene deletion of *SALL1*. A subsequent aCGH analysis revealed 6Mb deletion in 16q including *SALL1* and *RPGRIP1L*. **Discussion.** Most of 16q heterozygous microdeletion cases with only one gene either *SALL1* or *RPGRIP1L* have no renal impairment. However, our cases with both heterozygous gene deletions had severe to moderate renal impairment. Therefore, we suggest that there may be some interactions between *SALL1* and *RPGRIP1L* gene, and our cases may be a new contiguous gene syndrome.

2690F

Fetal Alcohol Syndrome and Pitt-Hopkins Syndrome in four maternal-half siblings. *A. Asamoah[1], K.E. Jackson[1], Y. Senturias[2], K. Goodin[1], G.C. Gowans[1], K. Platky[1], J.H. Hersh[1].* 1) Dept Pediatrics, Univ Louisville, Louisville, KY; 2) Dept Pediatrics, Carolinas Medical Ctr, Charlotte, NC.

Fetal alcohol syndrome (FAS) is identified by the presence of growth impairment, characteristic facial features and central nervous system abnormalities in the background of prenatal alcohol exposure. Fetal Alcohol Spectrum Disorders (FASD) is an umbrella term for the group of conditions arising from prenatal alcohol exposure that results in physical, intellectual, developmental and behavioral disabilities in the affected individual. Disorders on the spectrum include Alcohol-Related Neurodevelopmental Disorder (ARND), Partial FAS (PFAS) and Alcohol-Related Birth Defects (ARBD). Pitt-Hopkins syndrome is caused by haploinsufficiency of TCF4 gene either due to a point mutation or a deletion of the chromosomal region 18q21.2. Affected individuals have distinctive facial features, developmental delay/intellectual disability, and hyperventilation and/or breath-holding episodes while awake. We present four maternal half siblings diagnosed with FAS based on maternal history of alcohol abuse during pregnancy and physical examination who on chromosomal microarray analysis were found to have 18q21.1 deletion that includes the TCF4 gene. The finding of TCF4 deletion in these 4 maternal half-siblings suggests their mother has a germ line deletion or is mosaic for the deletion herself. These findings suggest considering chromosomal microarray analysis in patients with FASD to determine if there are no concomitant chromosomal abnormalities contributing to the phenotype.

2691W

Normal appetite and BMI in a 9 year old girl with haploinsufficiency of SIM1 due to a 2.2 MB deletion at 6q16.2-q16.3. *G.A. Bellus, K. Zegar.* Pediatrics, University of Colorado, Aurora, CO., USA.

Many patients with interstitial deletions involving 6q16 are reported to have phenotypic features that resemble Prader-Willi syndrome with obesity, short hands and feet, hypotonia and developmental delays. Several genotype-phenotype studies have implicated haploinsufficiency of *SIM1* (homologue of Drosophila single minded /OMIM 603128) as a candidate gene for obesity and developmental delays in these individuals. In Drosophila, *sim* is a transcription factor that is considered to be a master regulator of neurogenesis. In the mouse, *Sim1* is required for the development of the paraventricular nucleus of the hypothalamus and *Sim1* heterozygous knock-out mice exhibit hyperphagia, early onset obesity and increased linear growth. On the other hand, individuals with obesity and developmental delays have been reported with deletions at 6q14.1-q15 that do not involve *SIM1*. This suggests that *SIM1* hapolinsufficiency may not be necessary to cause obesity and development delays in some individuals with deletions in this region. We report a 9 year old girl with a history of hypotonia and mild intellectual disabilities with a normal appetite and normal BMI who was found to have a 2.2 MB deletion at 6q16.2-q16.3 involving *SIM1*. Together these results suggest that haploinsufficieny of *SIM1* may be neither necessary nor sufficient to cause hyperphagia and obesity in individuals with 6q deletions.

2692T

Lack of Nablus mask-like facial syndrome phenotype in a patient with a de novo microdeletion of chromosome 8q21.2q22.1. *D. Cherukuri[1], B. Crandall[2], S. Kantarci[1].* 1) Department of Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 2) Departments of Pediatrics and Psychiatry, UCLA.

Nablus mask-like facial syndrome (NMLFS) is associated with deletions of 8q22.1 chromosomal region. The clinical manifestations of this rare microdeletion syndrome include tight appearing glistening facial skin, blepharophimosis, telecanthus, sparse arched eyebrows, flat and broad nose, long philtrum, distinctive ears, upswept frontal hairline, short and broad neck, and developmental delay. Happy disposition is a typical behavioral trait of NMLFS [MIM: 608156]. Here, we report a 2-year old girl with a history of developmental delay and dysmorphic features. This girl was born post-term to a healthy nonconsanguineous couple of European descent. The delivery was vaginal and uncomplicated. At birth, the birth weight was 7 pounds 8 ounces, and there were no particular concerns postnatally and she was discharged in 2 days. She had delay in sitting at 9 months of age and sat alone at about 13-14 months period. Head circumference was 10 to 25th percentile, weight was 75th to 90th percentile, and height was 25th percentile at 22 month. Physical examination revealed lower limb hypotonia, trigonocephaly and dysmorphic facial features, including hypertelorism, depressed broad nasal bridge, low lying columella, mild prognathism, deep philtrum, wide mouth, thin upper vermillion border with prominent cupid's bow, and broad and short neck. Her ears were normal. Currently, she had speech delays and distinctive happy and social personality. SNP-chromosomal microarray testing (SNP-CMA) with Affymetrix CytoScan HD array revealed a de novo 10.9 Mb deletion of chromosome 8q21.2q22.1 ranging from genomic position from 86,840,946 to 97,724,672 (GRCh37/hg19). This deletion interval includes 51 RefSeq genes and overlaps with the reported NMLFS critical region (93.98-96.22 Mb (hg19)). Our patient does not present with the characteristic facial features of this syndrome. However, her happy personality is a common behavioral trait observed among the NMLFS patients. Similar to our findings, there are a few reported patients, without the NMLFS clinical features, with overlapping microdeletions with the NMLFS critical region. In conclusion, our report supports that deletions of chromosome 8q22.1 region result in variable phenotypes with and without the NMFLS syndrome.

2693F

MECP2 deletion a in patient with Rett Syndrome. *S.S. Costa, D. Villela, C. Rosenberg.* Genetics and Evolutionary Biology Dpt, , University of Sao Paulo, Sao Paulo, SP, Brazil.

Rett syndrome (RTT) is a neurodevelopmental disorder caused by mutations in the X-chromosome MECP2 gene. Approximately 95% of the girls with classic RTT have a MECP2 mutation. Over 250 specific mutations are associated with RTT: missense mutations, nonsense mutations and few deletions. We report a case of a child who at 18 months initiated neuropsychomotor regression, and developed microcephaly, epilepsy, stereotypical hand movements, irritability and autistic behavior. Karyotype and sequencing of MECP2 did not reveal alterations. Array-CGH (60K - Oxford Gene Technology) disclosed a deletion of 23 Mb encompassing exons 3 and 4 of the MECP2 gene. Array-CGH was performed in the phenotypically normal mother, and exposed a duplication in Xp22.31 of ~ 1.6 Mb. X-chromosome inactivation (XCI) pattern in blood leukocytes was assessed both in the patient and in her mother by methylation-sensitive restriction digestion followed by PCR amplification across the androgen receptor (CAG) repeat region: the mother showed 19:81 proportion and her daughter 72:28. Mother and daughter were tested on a high-resolution genomic array platform (Illumina 850K array). The profile suggests that the MECP2 alteration in the patient is not a contiguos deletion but rather a deletion of exons 4 e 3 without deletion of the intronic region in between them. The possible connection between the X-chromosome alterations in mother and daugher is still unclear, but two similar cases have been reported. The mechanisms associated with the chromosomal rearrangements is discussed.

2694W

Case report of a 17q21.31 microdeletion associated with EFTUD2 mandibulofacial dysostosis with microcephaly identified by comparative genomic hybridization. *S.K. Gandomi[1], D.M. Reeves[3], M. Parra[1], C.L. Gau[1], V. Yap[2].* 1) Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656; 2) University of Arkansas for Medical Services, Arkansas Children's Hospital, 1 Children's Way, Slot 512-5, Little Rock, AR 72202; 3) University of Arkansas for Medical Services, Department of Neonatal-Perinatal Medicine, 4301 West Markham, Slot 512-5B, Little Rock, AR 72205.

Mandibulofacial dysostosis with microcephaly (MFDM) is a rare, sporadic malformation syndrome manifesting with severe craniofacial abnormalities, microcephaly, developmental delay, and additional dysmorphic features. Although most cases of clinically diagnosed MFDM remain genetically unexplained, recent sequencing studies have linked this condition to heterozygous EFTUD2 mutations in 15 probands in the literature. In this case report, we present a previously undescribed dizygotic female twin proband (Twin A) born at 36 weeks gestation with severe microcephaly, microcrania, cleft palate, severe retrognathia, oral and pharyngeal dysphagia, bilateral proximal radioulnar synostosis, 11 thoracic ribs, abnormal MRI findings, high-pitched cry due to unilateral vocal cord paralysis, and additional dysmorphic features. Newborn screening and a series of additional biochemical investigations were diagnostically negative. The proband's twin sister (Twin B) was born healthy and shows no phenotypic similarities. Family history is unremarkable for any known genetic syndromes, and the twins' parents are both reportedly in good health. Array comparative genomic hybridization (aCGH)+SNP analysis was performed on Twin A to assess for chromosome rearrangements and regions of homozygosity. Results of this assay identified a small de novo pathogenic deletion on chromosome 17q21.31, encompassing the EFTUD2 gene. The deleted region also included 13 additional genes considered unlikely to be responsible for the proband's phenotype. No regions of homozygosity were identified in the 400K array, which would also confirm a non-consanguineous family history. Of the total 15 reported MFDM-associated EFTUD2 mutations described to date, all alterations resulted in genetic haploinsufficiency, consistent with our proband's microdeletion pathomechanism. In addition, our proband's phenotypic features both overlap and expand on the clinical features of previously reported probands in the literature. As a result, we encourage continued genetic investigation and reporting of other individuals with EFTUD2 mutations and clinical MFDM to better delineate genotype-phenotype correlations for more accurate diagnosis of this complex condition.

2695T

Nablus mask-like facial syndrome: 3 additional cases add support that del 8q22.1 is necessary but not sufficient to cause the classic phenotype. *S.S. Jamuar[1], H. Duzkale[1], N. Duzkale[2], C. Zhang[1], F.A. High[1], L. Kaban[3], S. Bhattacharya[4], J.M. Stoler[5], A.E. Lin[6].* 1) Harvard Medical School Genetics Training Program, Boston, MA; 2) Dept of Medical Genetics, Osmangazi University School of Medicine, Eskisehir, Turkey; 3) Oral Maxillofacial Surgery, Massachusetts General Hospital, Boston, MA; 4) Dept of Anesthesia, Massachusetts General Hospital, Boston, MA; 5) Division of Genetics, Boston Children's Hospital, Boston, MA; 6) Genetics Unit, Massachusetts General Hospital, Boston, MA.

BACKGROUND: Deletion (del) 8q22.1 has been reported in patients (pts) with Nablus mask-like facial syndrome (NMLFS) with its striking craniofacial and dermatologic anomalies. 9 pts reported by Allanson et al. [AJMG 2012] with overlapping deletions, but without the complete NMLFS phenotype suggest that del 8q22.1 is necessary but not sufficient. We report 3 additional pts with del 8q22.1, 2 of whom had NMLFS; one is undergoing complex oral maxillofacial surgery (OMFS). CASES: Pt. 1: 14 yo white male with jaw ankylosis, cleft soft palate, blepharophimosis, small ears, conductive hearing loss and shiny tight facial skin. NMLFS was suspected clinically and a 2.7 Mb del 8q22.1 was confirmed by chromosome microarray (CMA). Procedures prior to diagnosis included distraction osteogenesis, tracheostomy and cleft palate repair. Surgery after diagnosis included curvilinear osteogenesis distraction. Pt. 2: 2 yo white male from Turkey with microcephaly, shiny tight facial skin, blepharophimosis, telecanthus, arched eyebrows, low set dysplastic ears, frontal upsweep, and a small atrial septal defect. CMA confirmed a 3.7 Mb del 8q21.3-q22.1. Pt. 3: 16 yo female with unilateral cleft lip and palate, bilateral moderate to severe hearing loss, scoliosis, cognitive and behavioral issues without NMLFS. CMA showed a 4.6 Mb del 8q22.1. RESULTS: We summarize 22 cases of del 8q22.1, 11 of whom had NMLFS. Craniofacial features reported in NMLFS include shiny tight facial skin, expressionless face, blepharophimosis, telecanthus , bulbous nasal tip (100% each), upswept frontal hair pattern (91%), high arched sparse eyebrows (82%), abnormal ear shape (82%) and micrognathia (82%). In contrast, only a minority of the individuals with del 8q22.1 but without NMFLS had tight facial skin (30%), blepharophimosis (10%), telecanthus (20%), bulbous nasal tip (27%), sparse eyebrows (30%), abnormal ear shape (10%) and micrognathia (40%), and no pts had expressionless face or frontal upsweep. Developmental delay was noted in most pts in both groups. In addition to Pt 1, 7 others had a single surgical procedure. CONCLUSION: These 3 new pts confirm that del 8q22.1 is necessary, but additional factors contribute to the NMLFS phenotype. Complex OMFS can be achieved with intensive multidisciplinary care, especially techniques for airway management (micrognathia, restricted neck movement, small mouth); tracheostomy is done as a last resort.

2696F

In the shadow of *MEF2C*: Genotype-phenotype correlation for 5q14.3q21 deletions. *J.A. Rosenfeld[1], K. Stoate[2], A. Asamoah[2], R.R. Lebel[3], S. Raskin[4], L. Russell[5], J.W. Ellison[1], L.A. Schimmenti[6].* 1) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA; 2) University of Louisville, Louisville, KY; 3) SUNY Upstate, Syracuse, NY; 4) Pontificia Universidade Catolica do Parana, Curitiba, PR, Brazil; 5) Department of Medical Genetics, McGill University Health Centre, Montreal, Quebec, Canada; 6) University of Minnesota, Minneapolis, MN.

High resolution microarray analysis for genomic deletions and duplications has continued to provide ongoing delineation of genomic syndromes and generation of novel hypotheses regarding gene function in human development. We report further delineation of the phenotypic consequences of genomic deletions within 5q14.3q21.1. Microarray-based comparative genomic hybridization was performed on a sample from a 5.5-year-old male with developmental delay, severe expressive speech delay, visual impairment, bilateral incomplete coloboma of the iris, bilateral sensory hearing deficits, and seizures. The results identified four small interstitial deletions within 5q14.3q21.1, distal to *MEF2C* and totaling 2.1 Mb. These findings suggest that genes distal to *MEF2C* may also cause epilepsy and neurodevelopmental impairment. A review of the literature showed that some of the specific features in our patient have been occasionally reported with deletions extending to 5q15q21, including iris colobomas, hearing loss, and urogenital anomalies. To further delineate phenotypes associated with deletions in this genomic region we conducted a survey of patients in our database of clinical microarray cases and identified additional incidences of these features, though they were not fully penetrant. A comparison of our patients to those previously reported supports possible roles for developmentally important genes in these phenotypes, specifically *NR2F1*, with roles in neural, optic, and otic development, and *RGMB*, with roles in BMP2/4 signaling. In particular, *NR2F1* haploinsufficiency is likely associated with hearing loss, optic nerve atrophy, and hypoplastic corpus callosum observed in our proband. This report represents the importance of ongoing delineation of genomic syndromes through case identification and comparison to past cases through a robust clinical database.

2697W

Overlapping Phenotype of Silver-Russell-like and 14q32 Microdeletion Syndromes in a Child with Submicroscopic 11p15.5 Duplication and 14q32 Deletion. *H.J. Mroczkowski[1], D.B. Lowenstein[2], H. Abdel-Hamid[2], D.N. Saller[1], A. Rajkovic[1,3,4], S.A. Yatsenko[1,3].* 1) Department of Obstetrics, Gynecology and Reproductive Sciences, Magee-Womens Research Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA; 2) Department of Pediatric Neurology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 4) Department of Human Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Paternally and maternally derived duplications of the 11p15.5 region are rare abnormalities associated with overgrowth and growth restriction syndromes: Beckwith-Wiedemann syndrome (BWS) and Silver-Russell syndrome (SRS), respectively. Chromosome 11p15.5 harbors a cluster of imprinted genes, paternally expressed *IGF2* and *KCNQ1OT1*, and the maternally expressed *CDKN1C*, *KCNQ1*, and *H19* genes. Duplications, deletions and translocations of 11p region have been shown to result in distinct syndromes depending on genetic alteration and breakpoint positions relatively to imprinted 11p15.5 domains.

Herein, we present a case of a 7 month old girl who initially exhibited core hypotonia, developmental delay, alternating exotropia, prominent frontal bossing, and dysmorphic facial features. She was hospitalized one month later for new onset seizures. EEG demonstrated infrequent left temporal sharp waves during sleep indicating focal cerebral hyperexcitability with a tendency to have seizures of partial onset from the left temporal region. Magnetic resonance imaging of the brain revealed a hypoplastic corpus callosum, gray-white blurring in the right posterior cingulate gyrus, and diffusion restriction in the right thalamus consistent with a postictal state. Array comparative genomic hybridization showed alterations in DNA copy number involving subtelomeric regions of chromosomes 11p and 14q, suggestive of a complex rearrangement, including a duplication of ~2.4 Mb segment of 11p15.5-pter, an ~0.16 Mb triplication of the proximal 11p15.5 region, and a 1.8 Mb terminal deletion of the 14q32.3 region. The 11p15.5 duplication encompasses the *IGF2* and *H19* genes, and exons 1-2 of the potassium voltage-gated channel (*KCNQ1*) gene, but does not involve maternally expressed *CDKN1C*. The complex chromosome abnormality in our patient is not associated with abnormal growth, but includes a clinical phenotype with SRS features. These features may therefore be partially attributed to a 14q microdeletion syndrome characterized by microcephaly, high forehead, lateral hypertrichosis, dysmorphic facial features, muscular hypotonia, intellectual disability, developmental delay, and seizures. Additional studies on patients with small duplications are required to further delineate the phenotype of imprinted genes in the 11p15.5 region.

2698T

SNP Arrays in Idiopathic Intellectual Disability. *E. Utine[1], G. Haliloglu[2], B. Salanci[1], A. Çetinkaya[3], P.O. Kiper[1], Y. Alanay[1], D. Aktaş[3], B. Anlar[2], M. Topçu[2], K. Boduroğlu[1], M. Alikaşifoğlu[3].* 1) Pediatric Genetics, Hacettepe Univ, Ankara, Turkey; 2) Pediatric Neurology, Hacettepe Univ, Ankara, Turkey; 3) Medical Genetics, Hacettepe Univ, Ankara, Turkey.

Intellectual disability (ID) has a prevalence of 3% and is classified according to its severity. An underlying etiology cannot be determined in 75-80% in mild ID, and in 20-50% of severe ID. After it has been shown that copy number variations involving short DNA segments may cause ID, genome-wide SNP microarrays are being used as a tool for detecting submicroscopic copy number changes and uniparental disomy. This project was performed to investigate the presence of copy number changes in patients with ID of unidentified etiology. Affymetrix® 6.0 SNP microarray platform was used for analysis of 100 patients and their parents, and data were evaluated using various databases and literature. Etiological diagnoses were made in 12 patients (12%). One novel finding of homozygous GRID2 deletion has been reported recently. Homozygous deletion in NRXN1 gene, duplication in IL1RAPL1 gene were detected for the first time. Two separate patients had deletions in FOXP2 and UBE2A genes, respectively, for which only several patients have recently been reported. Interstitial and subtelomeric copy number changes were described in 6 patients, in whom routine cytogenetic tools revealed normal results. In one patient uniparental disomy type of Angelman syndrome was diagnosed. SNP microarrays should be considered as a first-tier genetic test in the evaluation of patients with ID; being a screening test able to detect very small genomic changes, with a high etiological yield even in patients evaluated using traditional cytogenetic tools; offering analysis for uniparental disomy and homozygosity; and thereby being helpful to find novel disease-causing genes.

2699F

Bilateral cleft lip and bilateral thumb polydactyly with triphalangeal component in a patient carrying two de novo deletions on chromosome 4q32 and 4q34 involving *PDGFC, GRIA2* and *FBXO8* genes. *A. Brusco[1,2], A. Calcia[1], G. Gai[2], E. Di Gregorio[1,2], F. Talarico[2], V.G. Naretto[2], N. Migone[1,2], E. Pepe[3], E. Grosso[2].* 1) Dept. Medical Sciences, Univ Torino, Torino, TO, Italy; 2) A.O. Città della Salute e della Scienza, S.C.d.U. Medical Genetics, Torino, Italy; 3) A.O. Città della Salute e della Scienza, Dipartimento Chirurgia Generale e Specialistiche, Torino, Italy.

We report a newborn male with bilateral clefts of the primary palate, a duplicated bilateral triphalangeal thumb, and a patent foramen ovale. At 4 yr., he presented moderate psychomotor developmental delay with normal brain MRI. The association of clefts of the lip/palate (CL/P) and triphalangeal thumbs, per se an extremely rare finding, has never been reported so far. In our case, the array-CGH analysis revealed two de novo deletions (~1.2 Mb and ~400 Kb) on the long arm of chromosome 4, containing four genes: platelet-derived growth factor C (PDGFC), glycine receptor beta subunit (GLRB), glutamate receptor ionotropic AMPA 2 (GRIA2), and F-box protein 8 gene (FBXO8). PDGFC codes for a mesenchymal cell growth factor already associated with clefts of the lip. Pdgfc -/- mice have skeletal anomalies, and facial schisis resembling human cleft/lip palate; GRIA2 codes for a ligand-activated cation channel that mediates the fast component of postsynaptic excitatory currents in neurons, and may be linked to cognitive dysfunction. FBXO8, a gene of unknown function, is member of the F-box gene family, among which SHFM3 is mutated in human split-hand foot malformations type 3. The presence of overlapping deletions in patients who do not share the same phenotype of our case suggest an incomplete penetrance, with a possible effect of modifier genetic factors.

2700W

Normal intelligence and features of Bardet-Biedl syndrome in a family with a duplication of chromosome 20p13-p12.1. *S.M. Nikkel, A.G. Hunter.* Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada.

Pure duplications of chromosome 20p have been reported rarely in the medical literature. We presented a family with a duplication of chromosome 20p13-p12.1 of 9.2 Mb in size. The proband was initially referred in childhood for a submucous cleft palate with bifid uvula and unilateral high-grade myopia. She was later found to have a didelphic uterus, bilateral renal cysts, and cervical and thoracic vertebral anomalies. Her mother has a bicornuate uterus, bilateral high-grade myopia, and cervical anomalies. Physical features are shared with a maternal uncle, who has a VSD, and the maternal grandmother. The mouths are small, the columella are low hanging and the nasolabial creases are deep. There is brachydactyly and tapering of the digits. The voice quality is marked by hypernasality. Family members have pursued post-secondary education with success. The proband has significant anxiety, and this is seen to a lesser extent in her mother. Duplications of 20p are reported to be associated with cognitive impairments, but the regions involved are larger than what we describe. Vertebral anomalies are common in trisomies of this region and may reflect dosage effects of the *JAG1* gene. The *MKKS* gene is known to cause a recessive phenotype due to loss of function mutations. However, this family demonstrates features of Bardet-Biedl syndrome (uterine anomalies, high palates, renal cysts, and anxiety), despite having an increase in copy number. They do not have polydactyly or retinitis pigmentosa, although ocular problems are present with extremely high myopia. In summary, large 20p duplications are associated with a number of anomalies, but a normal cognitive outcome is possible.

2701T

Inherited Yq12ter Deletion Associated with Congenital Cataracts, Microphthalmia and Autistic Spectrum Disorder in 3 Brothers. *P. Bitoun[1], A. Delahaye[2,3], B. Benzacken[2,3], E. Pipiras[2,3].* 1) Gen Med, CHU Paris-Nord, Hopital Jean Verdier, BONDY, France; 2) Embryo-Cytogenetics, CHU Paris-Nord, Hopital Jean Verdier, Bondy, France; 3) INSERM 676, Robert Debre Hospital, Paris, France.

Purpose: We screened a cohort of 65 patients with syndromal ocular disorder for CNV by microarray analysis after informed consent. We identified a family of male twins and an older brother with congenital cataract and possibly secondary microphthalmia associated with autistic spectrum disorder with cognitive delay, absent speech, introverted affect and severe communication disorder respectively aged 29 and 30 years old and a novel CNV. Methods We used an Illumina microarray platform with Human Hap 300, Human Cyto SNP 12, with 300k whole genome markers upon DNA extracted from lymphoblastoid cell lines. Illumina Software with Genome Studio 2010.3 , CNV partition 3.1.6 was used for analysis and NCBI build 36 (hg18) as reference. Patients were screened with brain MRI, and screened for mutation of SOX2, OTX2, RAX and PAX6 using Sanger sequencing. DNA was extracted from lymphoblastoid cell lines form all patients and cultured cells were used for FISH confirmation of CNV. Results A 296.8 Kb deletion in the Yq12ter region was identified and confirmed by FISH on metaphase lymphoblastoid cells using the amplitech subtelomeric Yq/Xq probe in all 3 boys and healthy father. Three brothers had normal brain MRI as well as negative SOX2, OTX2, RAX and PAX6 mutation analysis by Sanger sequencing. This Yq12ter deletion is within the Pseudo-Autosomal Region PAR2 identical to the Xq28 region. The possible role of the 3 deleted genes from the region in the autistic spectrum disorder will be discussed and the fact that the deletion was inherited from father cannot clearly exclude its pathogenic role. Discussion: The fact that there is a history of cataract in father's as well as mother's family without either parent being affected raises the possibility that the cataract is a separate trait unrelated to the autistic spectrum disorder and inherited from carrier parents in a recessive or even possibly dominant fashion. The deletion CNV identified contains 3 genes and it is difficult without further functional analysis to ascertain causality of this novel CNV in the pathogenesis of the autistic spectrum disorder in the 3 brothers.

2702F

MALFORMATION VARIABILITY ASSOCIATED TO CHROMOSOME TRISOMIES. CLINICAL AND PHENOTIPICAL IMPLICATIONS IN SEVERAL PATIENTS AT A PEDIATRIC HOSPITAL IN MEXICO. *M. Barrientos[1], J.M. Aparicio-Rodriguez[2,6], M.L. Hurtado-Hernandez[3], M.A. Cubillo-Leon[4], S. Chatelain-Mercado[5].* 1) Endrocrinology; 2) Genetics; 3) Cytogenetics; 4) Rehabilitation Therapy, Hospital para el Nino Poblano, Puebla, Puebla; 5) Biotechnology; 6) Estomatology.

Chromosome trisomies are considered alterations in the chromosome number or structure. A trisomy is therefore a type of polysomy in which there are three chromosome copies, instead of the normal two. A trisomy is considered an aneuploidy or abnormal number of chromosomes. There are two different trisomy types; Full trisomy" where an entire extra chromosome has been copied. "Partial trisomy" means that there is an extra copy of part of a chromosome. Depending on the chromosome, a trisomy is named as 'Autosomal trisomies" (trisomies of the non-sex chromosomes) and "Sex-chromosome trisomies." In this study both Autosomal and Sex-chromosome trisomies are described in differents patients, depending on the affected chromosome. Among 4617 chromosomal studies performed during 19 years (from 1992 to 2011), at Hospital Para el Niño Poblano in México, 34.6% (1596 patients) had chromosomal alterations. Among these studies population, a male and female pediatric patients are described, with different chromosome trisomies, were chromosome changes are classified as structural or numeric alterations. All trisomies patients were described in this study analyzing their phonotypical and clinical features, medical treatments and prognosis.

2703W

Congenital primary microcephaly and type B-like brachydactyly , a new syndrome? *A. Lavillaureix[1], J. Masliah-Planchon[1,2], S. Passemard[1,2], S. Drunat[1,2], A. Verloes[1,2].* 1) Department of genetics, Robert Debré Hospital, Paris, France; 2) INSERM U676, Robert Debre Hospital, PARIS, France.

We report a 10 year-old girl with primary microcephaly and brachydactyly . Microcephaly was suspected during the pregnancy by echography. At birth, at term, head circumference was 29cm (-5 SD), weight 2770g, length 50cm . Pregnancy was normal (no drug nor alcohol) and there were no perinatal problems and no feeding difficulties. She is the first child of unrelated parents none of the members of this family have brachydactyly or microcephaly. She started walking at 8 month-old, first words were around 1 year. She followed mainstream schooling. In 2nd grade, problems with motricity and writing fatigability were observed. Clinical examination at 8 year-old: showed a child of medium stature with generalised amyotrophy and microcephaly 41,5cm (-8 SD). Neurological examination was normal. She had facial dysmorphism with upslanted palpebral fissures and microdontia. She has bilateral brachydactyly ressembling type B of hands and feet, respecting thumbs but involving halluces, discovered at birth. On X rays, the terminal phalanges were missing on most digits, intermediate phalanges were hypoplastic or aplastic, with anonychia and cone-shaped epiphyses. MRI revealed a brain of reduced volume but normal structure and gyration. CGH array (180 k), caryotype and mitomicyn-induced chromosome breakage test were normal. To our knowledge it is the first described case of severe congenital microcephaly with normal IQ associated with brachydactyly and anonychia. This disorder is clinically distinct from Jawad syndrome (due to mutations in RBBP8/CTIP - currently under sequencing), which shows mental retardation and ungual hypoplasia, and Teebi anonychia-microcephaly syndrome, which has much milder anomalies.

2704T

Microrearrangements in individuals within the Holoprosencephaly spectrum. *L.A. Ribeiro-Bicudo[1], B.F Gamba[1], C. Rosenberg[2], A.L.B. da Rocha[1], A.L.C. Gaspar[1], R.M.C.S. Sandri[1], A. Richieri[1].* 1) Genetics Department, Hospital for Rehabilitation of Craniofacial Anomalies, Bauru, Sao Paulo, Brazil; 2) Human Genome Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, São Paulo, SP, Brazil.

Holoprosencephaly (HPE) is a malformation sequence where the cerebral hemispheres fail to separate into distinct left and right halves. It can be associated with midline structural anomalies of the central nervous system and/or face. The etiology of HPE is complex, with both environmental and genetic factors being implicated. Numerous different heterozygous mutations have been identified in HPE patients and include missense, nonsense, deletion, and frameshift mutations that are located throughout the gene. Chromosomal abnormalities have been attributed as the main commonly identified cause and high frequency of rearrangements have been reported in studies with array CGH. In the present work we found rearrangements trough Multiplex Ligation-dependent Probe Amplification (MLPA) and arrayCGH analysis in eight non related individuals who presented within the holoprosencephaly (HPE) spectrum previously screened for mutations in some HPE determinant genes such as SHH, GLI2, SIX3, TGIF, and PTCH. Molecular findings showed microdeletions and a microduplication involving SHH, TGIF and ZIC2 genes in five non related individuals, and three individuals presented chromosomal microrearrangements consisting in two duplications, one in a critical region harboring the SIX3 gene (2p21) and other in a region not related with HPE genes (13q14); and a third individual presented a microdeletion (8p23), which is also not associated to HPE phenotype. Because the clinical and genetic heterogeneity existing in patients with HPE and the increasing survival of these patients we concluded that these analysis are indicated in those cases where the search for mutations in main causative genes were negative providing a better tool for genetic counseling.

2705F

Maternal consumption of clay during pregnancy: an unexpected cause of recurrent congenital microcephaly with intracranial calcifications in babies from French Guyana (pseudo-Aicardi-Goutière syndrome). *A. Verloes[1], S. Passemard[1], V. Lambert[2], G. Carles[2], J. Goullé[3], A. Laquerrière[4]*. 1) Departmen,t of Genetics, Robert DEBRE University Hospital and INSERM UMR 676, Paris, France; 2) Department of Obstetrics , Regional Hospital, Saint-Laurent du Maroni, French Guyana, Fance; 3) Department of Toxicology, Rouen University Hospital, Rouen, France; 4) Department of Pathology, Rouen University Hospital, Rouen, France.

Eating clay ("pemba") during pregnancy is a traditional behavior in the Bushinengue population living on the border of the Maroni river, in the French department of Guiana. Clay consist in aluminium silicate. It is a powerful chelator of iron, and this practice (linked to traditional medicine), is responsible for a high incidence of severe anemia of pregnancy in this area. We report on two sibs born to a pemba-eater mother. The first child was born at term with IUGR severe microcephaly. Intracranial "calcifications" were observed by ultrasound screening during the second trimester. CT scan confirmed massive radio-opaque deposits in the brain basis. The clinical diagnosis of TORCH or Aicardi-Goutières syndromes were suggested initially. The child survived with major developmental delay. At age 7y, she has an OFC of 39 cm (-10 SD) and a height of 10 cm (-4SD). CSF interferon and TORCH screening were negative. Recurrence of microcephaly during the second pregnancy lead to TOP, after diagnosis of a similar microcephaly. Neuropathological examination confirmed severe microcephaly, with extensive microcalcifications dispersed throughout the brain. Electron microscopy made it possible to visualize intraneuronal aluminium silicate deposits, resembling aluminium deposition observed in post-vaccinal myofasciitis. The most likely mechanism to explain this recurrence is an association of IUGR secondary to severe maternal anemia combined with accumulation of exogenous silicates in the neural cells. This appears to be the first description of fetal brain disruption secondary to ingested clay. The syndrome superficially mimics Aicardi-Goutières syndrome, and convey a high risk of recurrence.

2706W

Cystic lymphangioma in a 9-year-old boy with Sotos syndrome: review of the tumoral risk in this overgrowth syndrome. *O. Cracco[1], G. Jedraszak[1], T. Dery[1], B. Devauchelle[2], V. Strunski[3], L. Burglen[4], .JF. Ikoli[5], B. Demeer[1], M. Mathieu[1], A. Leke[6], G. Morin[1]*. 1) Genetic department, Amiens University Hospital, Amiens, France; 2) Maxillofacial surgery, Amiens University Hospital, Amiens, France; 3) Otolaryngology, Amiens University Hospital, Amiens, France; 4) Genetic department, Trousseau Hospital, Paris, France; 5) Pathology service, Amiens University Hospital, Amiens, France; 6) Pediatric Reanimation, Amiens University Hospital, Amiens, France.

Background Overgrowth syndromes (OGS) form a heterogeneous group of disorders in which the main characteristic is a weight, height or head circumference over +2 standard deviations. Benign tumours and neoplasms are not exceptional in OGS. Sotos syndrome (SS) (OMIM 117550), one of these OGS, is characterized by cerebral gigantism, distinctive craniofacial appearance, and variable learning disabilities. Haploinsufficiency of the NSD1 gene was identified as the cause of the disorder. From its first description by Sotos in 1964, more than 25 malignancies and 10 benign tumours were reported in the approximately 500 known patients worldwide. Material and method The case of a patient with SS and cystic lymphangioma is retrospectively described. A Pubmed research crossing 'Sotos syndrome or cerebral gigantism' and 'tumour or tumor or cancer or neoplasm or malignancy' was realized. Results are reported. Case report This male patient was the third child of related healthy parents. Pregnancy was unremarkable. Delivery occurred by caesarian section at 37 weeks of amenorrhea in a context of acute foetal suffuring and umbilical cord procidence. Neonatal period was difficult with prolonged tracheal ventilation, E Coli sepsis and ulcero-necrotizing enterocolitis. In the first years of life he developed psychomotor retardation, hyperactivity, advance stature (+3SD), weight (+4SD), and macrocephaly (+2.5SD). Diagnosis of SS was made at the age of 4, and a heterozygous mutation of NSD1 gene was identified (c.3659_3660delAG / p.Glu1220AlafsX5 - Burglen L). At the age of 9 years, a subcutaneous tumour of the neck was discovered. Ultrasound examination and TDM confirmed the presence of a well delimited mixed tumour. This tumour was surgically removed and pathological examination confirmed the diagnosis of cystic lymphangioma. Discussion To our knowledge, this is the first reported case with SS and cystic lymphangioma. This association could be coincidental. However, with an approximate tumoral risk of 7% (at least 35 cases of tumours for above 500 reported cases), SS is really a tumour associating disease. Tumours are often embryonal, but various other kinds of tumours have been reported. Age at time of tumour detection varied greatly, from birth to adulthood. The tumoral predisposition in SS probably requires an attentive clinical follow-up, but the variety of tumour types and localisations makes targeted investigations difficult.

2707T

Cromosome 4 Deletions and Translocations Among 4617 Cariotype Studies at a Third Level Pediatric Mexican Hospital. 4p-, 4q-, T (1; 4), T (3; 4), Six Cases Report. *R. Zamudio[1], J.M. Aparicio-Rodriguez[2,5], M.L. Hurtado-Hernandez[3], F. Cuellar-Lopez[4], H. Chavez-Ozeki[5], S. Chatelain-Mercado[6]*. 1) Cardiology; 2) Genetics; 3) Cytogenetics; 4) Urology, Hospital para el Nino Poblano, Puebla; 5) Estomatology, Benemérita Universidad Autónoma de Puebla; 6) Biothecnology, Universidad Autónoma Metropolitana, Mexico.

Chromosome aberrations are considered changes in the chromosome number or structure. The etiology factor is due to gametogenesis inborn error (meiosis) or during the zygote first cellular divisions. It might occurs during metaphase from the cellular cycle, where DNA loses are seen (clastogenic processes) due to DNA repair processes deficiency o total absence, among others. Six genetic patients associated to chromosome 4 aberration were analyzed; three Wolf-Hirschhorn syndrome patients, a deletion of long arm 4 chromosome and two 1;4 and 3;4 chromosome translocations among 4617 Karyotype studies performed during 19 years period of time (from 1992 to 2011) at a Pediatric Hospital in Mexico. These chromosome changes are classified as structural alterations where these six patients from different families were chosen to evaluate their clinical characteristics, medical or surgical treatments according to their different genetic aberration.

2708F

Bench to bedside... The role of clinical genetics in the age of genomic medicine: The shifting paradigm. *T. Bardakjian[1], A. Slavotinek[2], A. Schneider[1]*. 1) Dept Peds/Genetics, Albert Einstein Med Ctr, Philadelphia, PA; 2) Dept. Pediatrics, Univ California, San Francisco, CA.

Advances in molecular diagnostics are catapulting genetics into all areas of medicine. The classic paradigm of clinical geneticist as dysmorphologist collecting phenotypic information to identify a syndrome or differential diagnosis and then ordering diagnostic testing is changing. New technology enables identification of gene mutations which are not associated with well-described phenotypes. While new diagnostic technologies offer great potential to provide a genetic diagnosis, when mutations are not accompanied by clinical prognosis or syndromic information the utility of the diagnosis for the patient is in question. The molecular diagnosis is not the end of the involvement of the clinical geneticist but rather a pivotal moment when genetics really becomes the key to best practices of medical care by a knowledgeable team with regular follow-up and monitoring. This paradigm shift is exemplified in the case of gene identification in the rare birth defect anophthalmia/microphthalmia (A/M). Molecular diagnosis is possible in up to 40% of individuals. Numerous eye development genes can now be tested for, but the natural history of these mutations have not been established. SOX2 as an example, was identified as a causative gene in 2005 and was termed SOX2 anophthalmia syndrome. Collection of clinical data has identified a wide variety of ocular findings and systemic issues not initially described. Testing is routinely ordered by a variety of physicians. However, there is limited knowledge of long term issues and it is not possible for most physicians to offer accurate anticipatory guidance for these patients. The A/M Registry at Einstein Medical Center Philadelphia has collected 30 cases of individuals with SOX2 syndrome. Combining these cases with others reported in the literature, the Registry has been able to develop a natural history and phenotypic spectrum to provide some guidance. It is import to note that given the infancy of this diagnosis, ongoing follow-up is essential. As more eye development genes are identified and clinical findings broaden the spectrum for each gene, the collection of clinical data over time will be critical to the provision of optimal medical care. Long term clinical and outcome data collection and analysis by clinicians is essential to accurately describe the phenotypes which correspond to identified gene mutations. This will enable more accurate genetic counseling and support for families with rare disorders.

2709W

Targeted next-generation sequencing for the molecular genetic diagnostics of mandibulofacial dysostosis. *Y. Kuroda[1], I. Ohashi[1], T. Saito[2], J. Nagai[2], K. Ida[1], T. Naruto[1], M. Masuno[3], K. Kurosawa[1].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Department of Clinical Laboratory, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Genetic Counseling Program, Kawasaki University of Medical Welfare, Kurashiki, Japan.

Mandibulofacial dysostosis (MFD) is a clinically and etiologically heterogeneous group of conditions characterized by significant malar and mandibular hypoplasia. Conductive hearing loss, lower eyelid anomalies, dysplastic ears and cleft palate are frequent associated features. Although many distinct MFD have been described clinically, phenotypic overlap makes it difficult to distinguish syndromic MFD and other craniofacial conditions. Causative mutations have only been identified for syndromic MFD, including TCOF1, POLR1D, and POLR1C in Treacher Collins syndrome, DHODH in Miller syndrome, SF3B4 in Nager syndrome, and EFTUD2 in Mandibulofacial dysostosis with microcephaly (MFDM). We performed targeted next-generation sequencing for MFD. Five patient samples were sequenced by MiSeq (Illumina). DNA libraries were enriched for sequences by capture-based approach. (HaloPlex, Agilent Technologies) Amplicons were designed for mixed panel of six MFD genes (TCOF1, POLR1D, POLR1C, DHODH, SF3B4, EFTUD2) and 69 ciliopathy genes, covering all coding regions and UTRs, in total, 398,456bp. Mean depth of coverage over all samples was 336× and bases covered by at least 15 reads were 97.5% of CDS. Data were analyzed by BWA ver.6 + GATK pipeline. Calling CNV was based on log ratio and z-score of read depth on each exon. CNVs were found in two of five patients and confirmed by array CGH. No mutation considered likely to pathogenic was found in all patients in MFD genes. Patient 1 was 2-year-old boy, the first child of nonconsanguineous and healthy parents. He had esophageal atresia, bronchomalacia, external ear atresia, cleft palate, microcephaly, and mild developmental delay. He had also microtia, mandibular hypoplasia, and slightly hypoplastic zygomatic arches. NGS analysis revealed deletion of exon 16 to end in EFTUD2 gene. Array CGH (Agilent 400k) revealed 37kb deletion at 17q21.31 encompassing EFTUD2 (5' part) and qPCR confirmed the deletion. Parents had no deletion of EFTUD2 gene. The final diagnosis was MFDM. Patient 2 was older brother of siblings with mandibulofacial dysostosis. Their parents were nonconsanguineous and healthy. NGS analysis revealed deletion of all exons of POLR1D. Array CGH (Agilent 400k) revealed 470kb deletion at 13q12.1 encompassing POLR1D gene, causative gene of Treacher Collins syndrome. Targeted next-generation sequencing is an efficient approach for detecting CNVs or intragenic deletions as well as pathogenic mutations.

2710T

FATCO syndrome: Report Nine Cases in Peru and Case Review. *H. ABARCA-BARRIGA[1, 2], B. GALLARDO[1], M. TRUBNYKOVA[1].* 1) Genetic, Instituto Nacional de Salud del Niño, IIMA, LIMA, Peru; 2) Centro de Investigación de Genética Humana Wiñay.

FATCO syndrome is characterized by the presence of Fibular Aplasia, Tibial Camptomelia (tibial arched), Oligosyndactyly (finger/toes absent and union of one or more) (MIM 246570). It's describe for Hecht and Scott at 1981; but Courtens at 1985 coins the term FATCO syndrome. We present nine children evaluated at the Department of Genetics, Instituto Nacional de Salud del Niño of Peru in the last four years with clinical features of the syndrome FATCO. Seven are males and two females; none familiar background, non-consanguinity, all patients do not come from any specific region of Perú. None of patients had development delay. They have normal anthropometry. The most affected leg is the right, two patients showed upper limb malformations only. We believe that there are some genetic and environmental factors in Peru own predispose to FATCO syndrome in more frequently than in other regions. To date, eleven cases have been published worldwide, excluding the nine cases reported in this edition.

2711F

A Novel Mutation in the MASP1 Gene Causes Autosomal Recessive Multiple Congenital Anomaly Syndrome. *H. Boulos[1], Y. Bejaoui[1], N. Khattab[1], Y. Al-Sarraj[1], M. Kambouris[1,2], H. El-Shanti[1,3].* 1) Shafallah Medical Genetics Center, Doha, Qatar; 2) Yale University School of Medicine, Genetics, New Haven CT, USA; 3) University of Iowa, Pediatrics, Iowa City, IA, USA.

Three patients from two related and consanguineous sibships of Pakistani ethnic origin are affected by a recognizable pattern of multiple congenital anomalies. The clinical picture includes increased inner canthal distance, hypoplastic upper lid with ptosis, blepharophemosis, maxillary hypoplasia, facial asymmetry, cleft lip/palate, high arched palate, irregular dentition, low set ears and low posterior hairline, mild scoliosis and decreased carrying angle of elbow. The apparent clinical characteristics overlap, but do not identify solely, with the individual Malpuech, Michels, Mingarelli or Carnevale syndromes, or what has been collectively referred to as the 3MC syndrome; hence, the referral to the phenotype as Multiple Congenital Anomaly syndrome. The family was studied by homozygosity mapping, and Whole Exome Sequencing of a single affected individual performed on ABI SOLiD4. A novel homozygous mutation [c.G542A] affecting the evolutionary conserved residue p.C181Y was identified at 3q27 in the MASP1 encoding a mannose-associated serine protease 1. The variant was confirmed by Sanger sequencing, segregates with the phenotype in the family and is predicted to be damaging by PolyPhen and SIFT. MASP1 functions as a component of the lectin pathway of complement activation. Mutations in the MASP1 gene and another gene (COLEC11) involved in the same pathway have been associated with human craniofacial malformation indicating an impending role for complement pathway elements in vital developmental processes during embryogenesis. The identified autosomal recessive variant extends further support to this hypothesis.

2712W

Diagnostic Criteria in Gomez-Lopez-Hernandez Syndrome: Contribution of brazilian patients. *C.H.P. GRANGEIRO[1], L.B. MESQUITA[1], J.A. JOSAHKIAN[1], C.M. LEVEPROST[1], M.L.M. CASTRO[1], N.R. QUARE-SEMIN[1], L.A.F. LAUREANO[2], A.C. SANTOS[4], J.M. PINA-NETO[1,3].* 1) 1 Serviço de Genética Médica - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 2) Laboratório de Citogenética - Hospital Das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 3) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 4) Divisão de Radiologia do Departamento de Clínica Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Centro de Ciências das Imagens e Física Médica.

Gomez-Lopez-Hernandez syndrome (GLHS) or cerebello-trigeminal-dermal dysplasia is a rare neurocutaneous disorder (so far, 34 sporadic cases have been reported), whose etiology is unknown. It is characterized by the triad of rhombencephalosynasis (RES), trigeminal anesthesia (TA) and bilateral scalp alopecia (SA). Based on the description of new patients and other clinical findings, subsequent reported cases have expanded the spectrum of craniofacial, neurobehavioral and cognitive phenotype. According to data, rhombencephalosynasis and bilateral scalp alopecia constitute obligate criteria. Brachycephaly/turribrachycephaly and midface retrusion are major craniofacial criteria. Strabismus, widely spaced eyes, plagiocephaly and lambdoid craniosynostosis are minor craniofacial ones. Motor delay, ataxia, hypotonia, intellectual disability and head shaking or other stereotypic movement are the neurological criteria. After identifying a new patient with GLHS in our service, totaling seven patients (20% of all cases worldwide), we decided to characterize our patients according to the proposed standard criteria and propose some adjustments. Our sample consists of 7 sporadic cases, 4 boys and 3 girls, one of them born from consanguineous parents (second cousins). The G-banding karyotypes were normal. We believe that trigeminal anesthesia must be considered a mandatory criteria since, not only all patients have this finding, they also had corneal opacity (a complication). All our patients presented broad forehead, upslanting palpebral fissures and low-set ears so major craniofacial criteria must include these dysmorphic features. Prognathism and post natal short stature should also be considered minor criteria. Neurological findings in our sample are consistent with the literature, including intellectual disability in variable degrees. We did not evaluate the presence of stereotyped head movements. Diagnostic criteria are essential for clinical reasoning construction, search of new patients and discovery of etiologic factors, especially in disorders which underlying mechanisms are not yet known.

2713T

New syndrome: brain malformations, Peters anomaly and multiple intestinal atresias. *J.G. Pappas[1], A.L. Shanske[2].* 1) Dept Pediatrics, Clin Genetic Scvs, New York Univ, Sch Med, New York, NY; 2) Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY (retired).

We present a newborn girl with brain malformations, bilateral Peters anomaly and multiple intestinal atresias and we compare this case with the only other similarly affected case reported in the medical literature (Shanske AL et al, 2002). Our case was born at 31.3 weeks vaginally without complications to a 34yo primigravida mother who was prenatally followed for mild fetal ventriculomegaly. The family history was not significant. Our clinical examination revealed weight, length and head circumference at about the 50% for gestational age. She had short palpebral fissures, two skin tags in place of the left tragus, a skin tag on the left cheek, corneal opacities and normal tone and activity. Feeding intolerance prompted an abdominal x-ray that suggested proximal jejunal atresia. Exploratory laparotomy revealed proximal jejunal atresia 5 cm distal to the ligament of Treitz, multiple discrete atretic segments of jejunum, a 5 cm segment of atretic distal ileum, and colonic atresia in the mid transverse colon. The atretic segments were resected, a mucous fistula constructed, and a gastrostomy tube placed. Brain MRI demonstrated bilateral schizencephaly communicating with large biparietal extra-axial cysts, partial absence of corpus callosum, ventriculomegaly, small areas of polymicrogyria and generally immature sulcation and myelination pattern for age. Ophthalmology concluded that the corneal opacities were part of Peters anomaly. Whole genome chromosome SNP microarray using the Affymetrix Cytoscan HD platform was reported normal. DNA sequence test of the exons, flanking regions and exon-intron boundaries of the *B3GALTL* gene to evaluate for Peters Plus syndrome revealed no mutations. Seizures started on day 12, and episodes of respiratory depression and bradycardia on day 19. She expired on day 20. Shanske AL et al, 2002 described a male newborn with extensive neuronal migration defect, bilateral Peters anomaly, multiple stenotic and atretic lesions in the jejunum and normal karyotype. At 21 months, he was severely microcephalic and developmentally delayed. The authors concluded that the abnormalities were due to a vascular disruption sequence. Both our case and the case reported in the medical literature present with migration defects, bilateral Peters anomaly and multiple intestinal atresias. This phenotype does not fit a recognizable pattern of malformation and we suggest that it may represent a new syndrome.

2714F

Richieri-Costa and Pereira Syndrome: severe phenotype. *S. RASKIN[1,2], M. SOUZA[3], M.C. MEDEIROS[3], M. MANFRON[2], DC. CHONG E SILVA[2].* 1) Group for Advanced Molecular Investigation, Graduate Program in Health Science, Health and Biosciences School, Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Parana, Brazil; 2) Hospital Pequeno Príncipe, Curitiba, Parana, Brazil; 3) Health and Biosciences School, Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Parana, Brazil.

Richieri-Costa/Pereira syndrome (RCPS, OMIM 268305) is a rare autosomal recessive disorder characterized by short stature, Robin sequence (micrognathia, glossoptosis, and cleft palate), cleft mandible and limb malformations. We report on a new case with a very severe phenotype. The patient was the first born of a 1st cousin consanguineous marriage, with no familial history of genetic disorders. Prenatal care and ultrasound exams showed no fetal morphological alterations. Delivery was at 38.5 weeks of gestation, the infant weighed 2,805 g, was 44 cm long, with a OFC of 31 cm and an Apgar score of 5 at 1 and 5 minutes, respectively. Malformations included: radial dysgenesis and finger anomalies, club feet, short limbs, toe anomalies, micrognathia, thoraccic deformity with severe sterno-clavicular chondral bilateral dysfunction, fatty hyperplasia of the anterior and posterior surface of the neck, webbed neck, low-set ears, ear deformities, and cranio-facial pansynostosis evolving to craniosynostosis. He had prolapse of base of tongue over the larynx and agenesis of the epiglottis; facial CT scan showed mandibular agenesis and deviated nasal septum. Limb radiographs showed radial agenesis, hypoplastic ulna, malformations of hands and fingers, hyperplasia of the tibia, agenesis of fibulae and toe deformities. Tube feeding was necessary from the time of birth, due to anomalies of the air ways. Three-dimensional CAT scan showed agenesis of the anterior mandible arch and epiglottis. Cranial CAT scan showed prominence of extra-axial space of the frontal convexity. Alterations were observed in skull formation with overriding of lambdoid sutures. To the present, 32 Brazilian and one non-Brazilian case have been described, with a great variability in the expression, but none reported the same degree of severity as the present case. Absence of Robin sequence and cleft mandible have been previously reported in several cases, but their absence particularly in a patient with a severe phenotype is unexpected. The recent finding of a causative genetic alteration at the 5̀ UTR of the EIF4AE gene as the mutation leading to RCPS may generate information on the variability of expression and further elucidate the basis of severe phenotypes such as that reported in the present study.

2715W

Minor facial malformations in relatives of patients with Goldenhar syndrome. *P. Santos[1], S. Oliveira[2], H. Saffatle[3], M. Cordoba[3], I. Ferrari[2], J. Mazzeu[4].* 1) Programa de Pos-graduaçao em Ciencias da Saude, Universidade de Brasília, Brasília, Distrito Federal, Brazil; 2) Laboratorio de Genetica, Departamento de Genetica e Morfologia, Instituto de Ciencias Biologicas. Universidade de Brasilia, Brasilia, Brazil; 3) Ambulatorio de Genetica, Hospital Universitario de Brasilia, Brasilia, DF, Brazil; 4) Programa de Pos-Graduaçao em Ciencias Genomicas e Biotecnologia, Universidade Catolica de Brasilia, Brasilia, DF, Brazil.

Goldenhar syndrome (GS), also known as oculo-auriculo-vertebral syndrome, is a congenital defect from anomalous development of first and second branchial arches. Its etiology includes unknown genetic factors, different chromosome aberrations and environmental factors, like maternal vasoactive medication, thalidomide and maternal diabetes. Several candidate genes have been proposed but none have been confirmed as causative of the phenotype. It's a disease with genetic heterogeneity and variable expressivity, commonly underdiagnosed. The majority of cases are sporadic so an environmental contribution to the phenotype cannot be excluded. Only a few familial cases have been reported so far. We studied a large family with GS, with three patients presenting classic signs of Goldenhar syndrome such as facial asymmetry, hemifacial microssomia, microtia or anotia and pre-auricular tags. Other family members exhibited mild malformations: mild facial asymmetry, hyperfolded ear helix and pre-auricular pits. These signs though common in the normal population when observed in family members of Goldenhar syndrome patients may be indicative of an inherited form of the syndrome. Financial support: CAPES, FAPDF/PPSUS.

2716T

Wiedemann-Beckwith syndrome associated with pre- and postnatal supraventricular tachycardia. *M. Willems[1], F. Brioude[2], S. Guillaumont[3], P. Amedro[3], M. Vincenti[3], O. Pidoux[5], N. Fries[4], L. Begue[4], C. Dumont[4], P. Sarda[1], P. Blanchet[1], L. Pinson[1], E. Haquet[1], J. Puechberty[1], G. Lefort[1], C. Coubes[1], I. Netchine[2], D. Genevieve[1].* 1) Genetics Dept, INSERM U844, Hopital Arnaud de Villeneuve, Montpellier, cedex 5, France; 2) Explorations Fonctionnelles Endocriniennes, Hopital Armand Trousseau, Paris, cedex 12, France; 3) Service de Cardiopédiatrie, hopital Arnaud de Villeneuve, Montpllier, cedex 5, France; 4) Service de Gynécologie Obstetrique, hopital Arnaud de Villeneuve, Montpellier, France; 5) Service de Pédiatrie et Néonatalogie, hopital Arnaud de Villeneuve, Montpellier, cedex 5, France.

We describe a baby who has Wiedemann Beckwith syndrome due to ICR2 (CDKN1C/KCNQ1OT) loss of methylation, associated with severe fetal and postnatal supraventricular tachycardia. The patient is the third child born from unrelated parents after in vitro fertilization with donor sperm. It was a bichorial biamniotic pregnancy with an early vanishing twin. Prenatal ultrasound screening at 14 WG revealed an omphalocele, leading to perform an amniocentesis. Karyotype analysis was normal. At 29WG, the mother was referred to the Genetic Department for the association of fetal macrosomia, macrostomia, nephromegalia, hepatomegalia and hydramnios. Wiedemann-Beckwith syndrome was suspected. At 30 WG, fetal supraventricular tachycardia was identified, requiring a maternal treatment with flecainide during late pregnancy. Diagnosis of Wiedemann Beckwith was confirmed at birth by clinical features and identification of an isolated ICR2 loss of methylation. Echocardiography was normal but supraventricular tachycardia recurred, requiring a treatment with digoxin and amiodaron which is ongoing at four months of age. Ectopic atrial tachycardia was reported only in two patients with Wiedemann Beckwith syndrome in 1985, associated with a focal cardiomyopathy. Conversely, a few patients with Wiedemann-Beckwith syndrome have congenital long QT associated with KCNQ1 expression defect but pathophysiology of these rhythmic troubles are unrelated. Fetal supraventricular tachycardia occurs in less than 1/2000 pregnancies. It can be associated with structural abnormalities of the heart or isolated. Two thirds of all patients with prenatal tachycardia will develop postnatal arrhythmia. We think that infraclinical focal cardiomyopathy could explain this rare but recurrent complication of Wiedemann-Beckwith syndrome.

2717F

Amyoplasia with congenital eye malformations and wrinkled skin: a new syndrome.
D.F.G.J. Wolthuis[1], E.V. van Asbeck[1], H.C. Andersson[2], E. Morava-Kozicz, MD, PhD[1,2]. 1) Pediatrics, Radboud University Nijmegen Medical Centre, Nijmegen, Gelderland, Netherlands; 2) Human Genetics, Tulane School of Medicine, New Orleans, LA.

Amyoplasia is characterized by the congenital absence of muscle tissue, and the most common underlying condition in arthrogryposis multiplex congenita. Patients present with typical symmetrical contractures, including internal rotation of the shoulder, extension of the elbow, flexion of wrists and hands, equinovarus feet and variable contractures of knees and hips with additional supra-articular dimples. Midline hemangiomas are also a common. All published amyoplasia cases are sporadic. Compared to the genetically heterogeneous distal arthrogryposis (DA) syndromes, no associated organ malformations have been described in amyoplasia patients. DA syndromes describe conditions of contractures in 2 or more body parts in the absence of a primary neurologic or muscular disease. We evaluated 2 patients diagnosed with muscle biopsy and imaging-proven amyoplasia in combination with unusual symptoms. Both patients showed eye abnormalities in addition to absent muscle: patient 1 had macular abnormalities and patient 2 had cataract. Feeding problems were present in both patients as well as spontaneous bone fractures, excessive skin in the nuchal region and typical facial features. Patient 1 also had diaphragmatic paresis. Amyoplasia has never been described in combination with eye abnormalities in the medical literature. However, retinal abnormalities are a feature that is commonly seen in DA, especially in type 5 and 5D. In DA5D a genetic defect was identified in *ECEL1*, encoding endothelin-converting enzyme-like 1. Mutation in this endopeptidase leads to reduced terminal branching of motor neurons resulting in decreased development of neuromuscular junctions, which was ruled out in our patients. The underlying pathomechanism in amyoplasia has not yet been discovered. Infectious and autoimmune causes, muscle dystrophy, spinal motor neuropathy and congenital disorders of glycosylation, affecting both muscle and eye development, were ruled out in our patients. Immune histology of extracellular matrix components, however, was abnormal. We suspect the role of genetic factors, like genes involved in collagen synthesis, affecting the extracellular matrix, and development of both the muscle tissue and eye structures, underlying the unique phenotype in our patients.

2718W

The contribution of discrepant DNA variations in discordant monozygotic twins with Esophageal Atresia/ Tracheo-esophageal Fistula.
E. Brosens[1,2], R.W.W. Brouwer[3], D. Veenma[1,2], D. Tibboel[1], W.F. van IJcken[3], A. de Klein[2]. 1) Pediatric Surgery, Erasmus MC-Sophia, Rotterdam, Zuid-Holland, Netherlands; 2) Clinical Genetics, Erasmus MC-Sophia, Rotterdam, Zuid-Holland, Netherlands; 3) Centre for Biomics, Department of Cell Biology, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands.

Esophageal Atresia with/or without Trachea-esophageal Fistula (EA/TEF) is a congenital anomaly that can either be present as isolated anomaly or in association with other birth defects. EA/TEF is believed to have a multifactor etiology, is associated with known (genetic) syndromes and can occur in combination with specific chromosomal aberrations, Copy Number Variations (CNV) or mutations. Although suspected, there is not much known about the genetic factors underlying EA/TEF. Twin studies can help to find potential causal or predisposing genetic factors of a disease. Monozygotic (MZ) twins typically share their genetic content and environment during development. We hypothesize that de novo mutations arisen early in embryonic development could explain the discordant phenotypes found in MZ twins. In order to detect these, complete or mosaic, DNA discrepancies six Monozygotic Twins (MZ) discordant for EA/TEF were characterized with SNP array and Exome-NGS. Hundreds discrepant SNPs and InDels were revealed using SNP-array genotyping and exome-NGS. Visual inspection of these events with Illumina's Genomestudio and the integrated genome viewer (IGV) indicated that most discrepancies were actually false positive differences caused by technical limitations, analysis settings (thresholding) and/or limitations of the variant calling. We compared the results obtained with multiple exome capturing techniques (Agilent Human 50 Mb All Exome v2, v4 and HaloPlex Exome), and analysis pipelines (GATK, CLC-bio read-mapper, CLC-bio®, SAMtools mpileup) and down-stream analysis strategies (CLC-bio® reference filter, an in-house developed comparison tool, Cartagenia®, Ingenuity® Variant Analysis) to determine if we could distinguish these false positive differences from actual discordant variants. Currently, we are evaluating whether these remaining discrepancies are true differences and could contribute to the disconcordant phenotype.

2719T

A case with single deletion of 17q21.31 involving KANSL1 gene and phenotype of CHARGE association.
ym. chan[1], kw. choy[1], ty. leung[1], ca. bacino[2]. 1) Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong , Hong Kong; 2) Department of Molecular and Human Genetics Baylor College of Medicine.

We describe a single case with prenatal ultrasound showing increased fetal nuchal translucency and features suggestive for CHARGE association with a small deletion in 17q21.31 involving the KANSL1 gene. On routine first trimester screening ultrasonography performed at 13 weeks of gestation, nuchal translucency was increased to 4.53mm. Chorionic villi sampling was performed. Conventional karyotype was normal. Array CGH detected 573kb single deletion at 17q21.31 involving the KANSL1 gene region, but not included other genes like MAPT and CRHR1. The clinical significance of the deletion at this region was not well known at that juncture. Anomaly scan was unremarkable. Multiple abnormalities were noted after delivery including bilateral membranous choanal atresia, bilateral coloboma involving fovea and optic nerve, bilateral malformed pinna with bilateral severe-profound hearing loss, patent ductus arteriosis compatible with CHARGE association. CHD7 sequencing analysis was normal with no CHD7 mutation identified. Patients with KANSL1 related intellectual disability may present with dysmorphgic features resemble those seen in CHARGE association. However, no reports of choanal atresia or colobomas were previously reported so this is the first report for these malformations in a patient with a KANSL1-related intellectual disability deletion.

2720F

Femoral-facial syndrome: long term follow-up and associated array CGH abnormalities.
A. Jacquinet[1], H. Valdes-Socin[2], C. Libioulle[1], J.H. Caberg[1], A. Verloes[1,3]. 1) Center for Human Genetics, CHU & University of Liège, Liège, Belgium; 2) Department of Endocrinology, CHU & University of Liège, Liège, Belgium; 3) Department of Medical Genetics and INSERM U676, APHP-Robert Debré University Hospital, Paris, France.

The femoral-facial syndrome is usually sporadic and its aetiology remains unknown. Non-genetic factors as maternal diabetes mellitus have been associated. Reports of familial cases have otherwise suggested autosomal dominant inheritance. We report the 20 years clinical follow-up of a girl with femoral-facial syndrome diagnosed at birth. Recently, array CGH investigation identified a 1400 kb duplication at 9q31.1, including the gene SMC2, and a 343 kb deletion at 12q24.33 including the genes CHFR, ZNF26, ZNF140, ZNF10 and ZNF268. Moreover, the patient presents a Mayer-Rokitansky-Kuster-Hauser syndrome diagnosed at puberty. Femoral-facial syndrome and Mullerian agenesis may reflect different defects in the primary axial mesodermal development, being the consequences of same environmental or/and genetic factors during blastogenesis. Among these genetic factors, we suggest the possible involvement of the two copy number variants reported here.

2721W

Identification of Mosaic Activating Mutations in Overgrowth Syndromes Using a Customized Next Generation Sequencing Panel on both Prenatal and Postnatal Samples. *L. Liu[1], F. Chang[1], E. Fang[1], G. Zhang[1], M.M. Li[1,2].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dan Duncan Cancer Center, Baylor College of Medicine, Houston, TX.

Overgrowth Syndromes are genetically heterogeneous diseases caused by both germline and somatic mutations of different genes. Recent studies have showed that a group of overgrowth syndromes, such as CLOVES and Proteus syndromes, are caused by postzygotic activating mutations in the genes involved in the PI3K-AKT signaling pathway, such as PIK3CA and AKT1. In addition, both germline and somatic mutations of PI3K-AKT pathway genes have been reported in two overlapping disorders, megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) and megalencephaly-capillary malformation (MCAP). Non-syndromic overgrowth features, such as isolated macrodactyly, have also been reported to be associated with PIK3CA mutations. Due to the low-abundance nature of these mutations, routine Sanger sequencing often yields negative results. We developed a next generation sequencing (NGS) test that targets all known mutations in multiple genes involved in the PI3K-AKT pathway. For differential diagnosis purposes, we also included a few other genes associated with overgrowth syndromes, such as PTEN and GNAS. Four patients including two prenatal cases, an amniotic fluid and a POC samples, and two postnatal cases suspected of CLOVES or Proteus syndromes were tested in our laboratory using the NGS panel. All four cases are positive for PIK3CA somatic mutations including one G542K, one H1047L, and two H1047R. These mutations are only present in the affected tissues. In the amniotic fluid case, the G542K mutation is positive in the DNA extracted from cultured amniocytes but negative in uncultured fluid of the same specimen. These results suggest that the mutant allele is present in the direct amniotic fluid sample at a very low frequency beyond the detection limit of the test. The activating mutation may render growth advantages to the cells carrying the mutation in culture, resulting in the enrichment of mutant allele. To the best of our knowledge, this is the first case of prenatal diagnosis of CLOVES syndrome. Our experience demonstrates that cultured amniocytes can be used for prenatal diagnosis of these overgrowth syndromes. In summary, the custom-designed NGS panel shows high accuracy and sensitivity for the detection of causal mutations in the overgrowth syndromes and facilitates clinical diagnosis both prenatally and postnatally.

2722T

Novel GATA4 promoter polymorphism associated with congenital heart disease in south Indian patients. *S. Mattapally[1], KS. Murthy[2], S. Nizamuddin[3], K. Thangaraj[3], SK. Banerjee[1].* 1) Division of Pharmacology, Indian Institute of Chemical Technology (IICT), Hyderabad,500 007, India; 2) Innova Children's Heart Hospital, Tarnaka, Hyderabad, India; 3) Centre for Cellular and Molecular Biology, Habsiguda, Uppal Rd, Hyderabad 500 007, India.

Background: Congenital heart diseases (CHDs) usually refer to abnormalities in the heart's structure or function that arise before birth. Although the exact mechanism behind this cardiac abnormality is not known, transcription factors play an important role in embryonic heart development. GATA4 is one of the candidate transcription factors and GATA4 mutation may leads to different types of CHD such as ASD, VSD, TOF and Single ventricle (SV). The aim of this study is to find the genetic association of CHD with GATA4 mutations from south Indian CHD patients. Method: GATA4 gene was genotyped in 100 CHD patients (ASD, VSD, TOF and SV) and 200 control samples in a case control study using sangers di-deoxy chain terminator cycle sequencing. Genotyping and mutational analysis was carried out with DNA star software. Functional significance of the GATA4 mutation was analyzed by in-silico software like polyphen, Sift, Pmut and other bioinformatics tools. Results: In this study we identified one 5' UTR (ptomoter region -490 to 100 bp) mutation i.e., 620 C>T(rs61277615 ,p>0.0007) and two coding region mutations i.e., c.1734 C>A (Pro394Thr) and c.1827 G>A (Asp425Asn). We also found five 3'UTR mutations i.e., 2400 T>C(rs884662), 2415 T>C(rs904018), 2470 A>G(rs804291), 2446 C>G (rs12825), 3139 A>T (rs12458). All mutation present in the CHD patients are absent in 200 healthy volunteers. Our 'In silico' data also provide evidence that all mutations reported above are pathological or alter the gene expression through micro RNA binding. Conclusion: The present study found that GATA4 genetic variations are associated with CHD in South Indian patients. Our bioinformatics study provide further evidence that those GATA4 mutations observed in Indian patients mutations may alter the function of the transcription factor binding and micro RNA binding , may leads to disease.

2723F

Trying to unravel the etiology of multiple midline congenital anomalies misdiagnosed as VACTERL. *E.M. Pereira[1], M.P. Ramos[2], J.M. Greally[2], R.W. Marion[1].* 1) Division of Medical Genetics, Department of Pediatrics, Children's Hospital at Montefiore, Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY; 2) Division of Computational Genetics, Department of Genetics, Albert Einstein College of Medicine, Bronx, NY.

A 13 year old girl presented to genetics clinic with a diagnosis of VACTERL. She was the product of a Mexican union with a nonsignificant family history. The patient was noted to have esophageal atresia at birth. In childhood, her medical condition was complicated by right kidney reflux which required nephrectomy. She developed scoliosis though there were no vertebral anomlies; surgical correction was necessary to correct the spinal curvature. The patient was never found to have cardiac or limb abnormalities. She was in stable condition until a few months prior to presentation when she had several days of hematuria. A pelvic MRI showed a malformed, dilated uterus, an atretic/absent vagina, and a cystic structure that may represent a portion of the cervix; her ovaries were structurally normal. The patient has no cognitive defects and does exceptionally well in school.

Though the patient was originally given a diagnosis of VACTERL, her multiple congenital anomalies do not fit clinical criteria for this association. Mayer-Pokitansky-Küster-Hauser syndrome can lead to absence of the vagina and uterus in the most extreme cases, but it would not explain the rest of her abnormalities. Interestingly, the patient has many midline defects from various embryological origins.

The patient's multiple congenital anomalies warranted a microarray. The results revealed a 0.91 Mb duplication at chromosome 6q11.1 (61,971,892 - 62,877,253) of unknown significance. Two genes in the area, KHDRBS2 and MTRNR2L9, have not been associated with any of the abnormalities seen in the patient. Given her unique combination of congenital anomalies, our division selected this patient for whole exome sequencing. Results are currently pending.

2724W

Associated malformations among patients with urinary congenital anomalies. *C. Stoll, B. Dott, Y. Alembik, MP. Roth.* Faculte de Medecine, Strasbourg, France.

Infants with urinary congenital anomalies (UCA) often have other associated malformations. The purpose of this investigation was to assess the prevalence and the types of associated malformations in children with UCA in a defined population. The associated malformations in infants with UCA were collected in all livebirths, stillbirths and terminations of pregnancy during 26 years in 347,810 consecutive births in the area covered by our population based registry of congenital malformations. Of the 1703 infants with UCA born during this period, 563 (33 %) had associated malformations. There were 119 (7%) patients with chromosomal abnormalities including 33 trisomies 18, and 168 (10%) nonchromosomal recognized dysmorphic conditions. There were no predominant recognised dysmorphic conditions, but VA(C)TER(L) association, Meckel-Gruber syndrome, and prune-belly syndrome . Two hundred seventy six (16 %) of the patients had multiple congenital anomalies, non syndromic, non chromosomal (MCA). Malformations in the musculoskeletal, the digestive, and the cardiovascular, and the central nervous systems were the most common other malformations. Prenatal diagnosis was performed in 70 % of the fetuses with UCA. The molecular basis of UCA remains unknown. However, mutations of certain genes have been associated with UCA including HNF1 beta, GDNF, RET, TCF2, PAX2, EYA1, SIX1, SIX2, SIX5, AGT, REN, ACE, AGTR1, BMP4, UPIIIA, FRAS1, FREM1, ROBO2, SOX17, HPSE2, CHRM3, and SALL1.The overall prevalence of associated malformations, which was one in three infants, emphasizes the need for a thorough investigation of infants with UCA. A routine screening for other malformations may be considered i infants and in fetuses with UCA.

2725T

A new case of Crane-Heise syndrome with comparative review of literature. *A. Handel[1], K. Fay[1], M. Costaldi[2], R. Lebel[1].* 1) SUNY Upstate Medical University, Syracuse, NY; 2) Pathology Associates of Syracuse, Crouse Hospital, Syracuse, NY.

Crane-Heise syndrome is a lethal constellation of dysmorphic features described in only 9 cases (first reported in 1981). The syndrome is notable for 1) poorly mineralized calvarium, 2) facial anomalies including cleft lip, cleft palate, low-set ears, hypertelorism, and 3) skeletal abnormalities including dysplastic clavicles, vertebral anomalies, and talipes equinovarus. We present the tenth case of this syndrome. The G4P2012>2022 Caucasian 35-year-old woman had one ectopic pregnancy and two surviving daughters; one said to be affected by autism and seizures, the other by ADHD and Asperger syndrome. Ultrasound at 7 weeks provided gestational dating. Repeat ultrasound at 13 weeks revealed gastroschisis, scoliosis, dangling choroid plexus, and left clubbed foot. Ultrasound at 16 weeks confirmed those observations and added cleft lip and palate, and bilateral ventriculomegaly. Amniocentesis revealed a 46,XX karyotype and normal oligo-microarray. Alpha-fetoprotein was elevated with a faintly positive acetylcholinesterase band. Ultrasound at 21 weeks showed fetal demise, apparently three weeks earlier by measurements, and delivery was induced. Autopsy revealed many features seen in Crane-Heise syndrome: markedly under-mineralized calvarium, cleft lip, cleft palate, hypertelorism, soft tissue syndactyly of one hand, low-set ears, and left clubbed foot. Our patient also had a large gastroschisis with extracorporeal viscera; abdominal wall defects have been described in 2 of the 9 previous cases. The patient had renal hypoplasia, not described in previous cases. There was also scoliosis, a narrow thorax, and markedly hypoplastic lungs (each described in only one previous case). Features commonly reported in the previous cases, but absent in the present case are: IUGR, agenesis of vertebral bodies, micrognathia, absent corpus callosum, absent pubic bone, cardiac abnormalities, and hypoplasia of phalanges. This fetus had findings that did not fit together unless one considers very rare syndromes. London Morphology Database offered Crane-Heise syndrome as a strong candidate. We propose that Crane-Heise syndrome should be considered in the differential diagnosis for prenatally discovered gastroschisis. There is strong evidence that this syndrome has autosomal recessive etiology, and our case does not contradict that hypothesis.

2726F

Pathogenic CNVs and causative gene analysis by SNP arrays as the third screening for 646 patients with intellectual disability and multiple congenital anomalies of unknown etiology. *D.T. Uehara, S. Hayashi, J. Inazawa.* Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan.

In order to identify genetic factors responsible for congenital disorders of unknown etiology in the Japanese population, we have investigated 646 subjects presenting with clinically uncharacterized multiple congenital anomalies and intellectual disability for eight years. We first performed a two-stage screening using two types of in-house bacterial artificial chromosome (BAC)-based arrays, which have allowed the identification of pathogenic copy number variants (CNVs) in 130 cases (20%). Next, we performed a third screening in 432 negative cases using a single nucleotide polymorphism (SNP) arrays platform (Illumina HumanOmniExpress BeadChips) to identify smaller causative CNVs undetected in the previous screenings. Besides the high resolution, this type of array has an additional advantage of detecting uniparental disomy (UPD) through the analysis of copy-neutral loss of heterozygosity (LOH). A preliminary analysis of the results revealed around 10% of the cases with CNVs comprising known pathogenic genes, for example, a deletion in the NRXN1 gene. So far, we have identified five cases with possible pathogenic CNVs that might reveal novel candidate genes for intellectual disability, and one case with uniparental disomy affecting chromosome 20. We are currently focusing on these cases by carrying out functional analyses, and we also plan to further elucidate the mechanisms that originated these rearrangements. Our goal is to finally determine the causative genes delineating these pathologies through a reliable correlation between genotype and phenotype.

2727W

The oculoauriculovertebral spectrum: refining the estimate of birth prevalence. *M.T. Gabbett.* Genetic Health Queensland, Royal Brisbane & Women's Hospital, Herston, Queensland, Australia.

The oculoauriculovertebral spectrum (OAVS) is a pattern of congenital malformations characterized by hemifacial microsomia and/or auricular dysplasia. However, the birth prevalence of OAVS is poorly characterized. Figures ranging from 1 in 150,000 through to 1 in 5,600 can be found in the literature. This study aims to evaluate the reasons behind such discrepant figures and to refine the estimated birth prevalence of OAVS. Published reports on the incidence and prevalence of OAVS were systematically sought after. This evidence was critically reviewed. Data from appropriate studies was amalgamated to refine the estimate of the birth prevalence for OAVS. Two main reasons were identified why birth prevalence figures for OAVS are so highly discrepant: differing methods of case ascertainment and the lack of a formal definition for OAVS. This study refines the estimate of birth prevalence for OAVS to between 1 in 40,000 and 1 in 30,000. This number needs to be confirmed in a prospective study using a formally agreed-upon definition for OAVS.

2728T

Initial data for benign CNVs distribution in Bulgarian patients. *S.P. Hadzhidekova[1], D.M. Avdjieva-Tzavella[2], B.B. Rukova[1], D.V. Nesheva[1], R.S. Tincheva[2], D.I. Toncheva[1].* 1) Medical Genetics, Medical University - Sofia, Sofia, Sofia, Bulgaria; 2) Section of Clinical Genetics, State University Pediatrics Hospital 'Queen Evdokia', Medical Faculty, Medical University-Sofia, bul. 'Iv. Geshov' 11, Sofia 1660, Bulgaria.

Introduction: Molecular karyotyping is coming up as an extremely suitable method for genetic diagnosis of patients with unclear dysmorphic syndromes and intellectual disability. In this study we present our results from microarray analysis of 52 patients with developmental delay and congenital malformations. Methods: Oligo array-CGH (BlueGnome CytoChip oligo 2× 10^5K, v1.1, 35kbp backbone resolution) was applied in 52 patients with developmental delay and multiple congenital anomalies. Results: A total of 247 CNVs were detected, of which 15 pathogenic, 108 with unknown clinical significance and 124 benign - mean number of CNVs per patient - 4,5. All pathological findings were validated by FISH analysis. In addition, the majority of the patients tested (41 patients) showed normal variations in the number of copies and variations of unknown clinical significance (34 patients). Analyses of the type and distribution of the different variations was performed and the clinical significance of variants of unknown nature was discussed. Conclusion: Our results show the advantages of high resolution microarrays for clinical diagnosis of patients with intellectual disability and congenital malformations, but also highlight the need for extensive population studies revealing the molecular nature and clinical significance of different CNVs and the necessitate for creation of detailed maps of variations in the Bulgarian population. This would facilitate the interpretation of unknown genomic imbalances in clinical aspect. Besides, it would help the widespread introduction of CGH microarray in diagnostic practices - postnatal and prenatal genetic diagnosis. Acknowledgements: Grant 02/76-21.12.2009, National Science Fund, Bulgaria.

2729F

Novel autosomal-recessive syndrome with short stature, distinct facial appearance, myopia, retinitis pigmentosa, bilateral hearing loss, and mild intellectual disability. *E. Schrock[1], T.M. Neuhann[2], I. Neuhann[3], A. Bier[4], B. Novotna[5], N. Di Donato[1].* 1) Institut fuer Klinische Genetik, Medizinische Fakultaet Carl Gustav Carus, Technische Universitaet Dresden, Germany; 2) Medizinisch Genetisches Zentrum, Munich, Germany; 3) MVZ Prof. Neuhann, Munich, Germany; 4) Gemeinschaftspraxis fuer Humangenetik, Dresden, Germany; 5) Sozialpaediatrisches Zentrum am Universitaetsklinikum Carl Gustav Carus, Dresden, Germany.

Retinal anomalies in combination with progressive hearing loss accompany a variety of genetic syndromes. We report here on three patients with retinitis pigmentosa, bilateral hearing loss, peculiar facial phenotype and an unusual combination of other clinical findings incompatible with any of the known genetic conditions. Patient 1 (3 years) presented with short stature (-2, 5 SDS), distinct combination of minor facial anomalies (prominent forehead, short palpebral fissures, deep-set eyes, bulbous nasal tip with broad columella, thin upper lip with accentuated cupid's bow), brachydactyly, high myopia (-3dp) and progressive bilateral hearing loss. Her motor development is mildly delayed. Patient 2 (40 years) is the paternal aunt of patient 1. Her brother, the father of patient 1, is unaffected. Like her niece patient 2 shows a short stature (-3, 5 SDS), brachydactyly and strikingly similar facial anomalies. She has progressive bilateral hearing loss with deafness at the age of 35 years. The complex eye involvement includes progressive myopia, retinitis pigmentosa (onset in the 2nd decade), glaucoma and corneal dystrophy (both started in the 4th decade). Patient 2 attended a regular school but needs support in her daily life. Both patients have normal findings in their brain MRI. Patient 3 (28 years) is a male patient, not related to patients 1 and 2. His family history is unremarkable. He presented with short stature, brachydactyly, progressive bilateral hearing loss, high myopia and retinitis pigmentosa (onset in the 2nd decade). His facial minor anomalies are remarkably identical to patients 1 and 2. The only additional clinical finding not present in the first family is alopecia areata of the scalp. Patient 3 shows mild intellectual disability. Comprehensive genetic tests (conventional and molecular karyotyping in all patients; sequencing of LTBP2, ADAMSTS17 and ADAMTS10 in patient 2, and ERCC8 and ERCC6 in patient 3) were all normal. Taken together we report on an apparently new genetic syndrome associated with short stature, distinct facial anomalies, myopia, retinitis pigmentosa, progressive hearing loss and mild intellectual disability. Based on the pedigrees we suggest an autosomal recessive inheritance. This condition is distinctive from Usher syndrome and from Weill-Marchesani syndrome. Exome sequencing is ongoing to clarify the molecular cause of the disease.

2730W

The use of exome sequencing to disentangle complex phenotypes. *H.J. Williams[1], C. Baccheli[1], J. Hurst[2], F. Lescai[1], L. Ocaka[1], C. James[1], C. Pao[3], E. Rosser[4], P. Beales[1].* 1) The Centre for Translational Genomics - GOSgene, Institute of Child Health, UCL, London, UK; 2) Royal Free Campus, UCL Medical School. London, UK; 3) Respiratory and General Paediatrics, Barts and The London Children's Hospital, London, UK; 4) Clinical Genetic, Great Ormond Street Hospital, London, UK.

The success of whole-exome sequencing (WES) to identify mutations causing single gene disorders has been well documented. However, WES has had limited success in the identification of more complex phenotypes resulting from the disruption of multiple genes. We describe a family where two offspring from healthy consanguineous parents present a complex congenital nonsyndromic phenotype consisting of peripheral neuropathy and bronchiectasis that has not been described previously. Through the use of WES we were able to simplify this complex phenotype and identified a causative mutation (R1070X) in the gene PRX, a gene previously shown to cause Charcot-Marie-Tooth Syndrome 4F and Dejerine-Sottas syndrome when this mutation is present in a homozygous state. For the bronchiectasis phenotype there was no single mutation or compound heterozygote identified which was deemed to be causal, reflecting the heterogeneous nature of this phenotype. This study highlights the potential utility of WES to disentangle complex phenotypes where multiple contributing loci in combination with environmental factors make it difficult to assign a clinical diagnosis; this then has further implications regarding the clinical management and use of therapeutics for such patients. In conclusion we show that WES has the power to improve patient diagnosis and therapy by disentangling complex phenotypes through the identification of causative genetic mutations for distinct clinical disorders that were previously masked.

2731T

Association study of genetic polymorphisms in DNA repair genes APE1/Ref-1 and DNA oxidative damage with the risk of neural tube defects. *J. Wang[1], X. Han[1], J. Guo[2], X. Wang[1], F. Wang[2], C. Ji[1], Z. Guan[1], Q. Xie[1], Z. Zhu[1], B. Niu[1], T. Zhang[2].* 1) Department of biotechnology, Capital Institute of Pediatrics, Beijing, China; 2) Department of molecular immunology, Capital Institute of Pediatrics, Beijing, China.

Neural tube defects (NTDs) are one of the most common human birth defects. Folate deficiency is closely related to NTDs. However, the mechanism remains unclear. Folate can prevent the effect of homocysteine on oxidative stress. Apurinic/apyrimidinicendonucleasee 1/redox-factor 1 (APE1/Ref-1) plays critical roles in DNA oxidative damage repair, oxidative stress. Previous study found the gene polymorphisms were associated with NTDs in a Californian population. In order to investigate the association of APE1/Ref-1 gene polymorphisms and oxidative stress with NTDs among the folate deficiency population in the high-risk area, a case-control study of 335 NTDs fetuses and 336 normal fetuses was conducted in Lvliang areas of Shanxi province with a high prevalence of NTDs. Total 36 single nucleotide polymorphisms (SNPs) in APE1/Ref-1 gene were genotyped by Sequenom MassARRAY Genotyping. The Superoxide dismutase (SOD) activities of fetal liver and maternal plasma were detected by the Cu-Zn/Mn-SOD assay kit and the numbers of apurinic/apyrimidinic sites (AP sites) in DNA of fetal brain tissues were measured by the DNA Damage Quantification Kit. The allele and genotype frequencies of 36 polymorphisms showed no statistically associated with NTDs. After stratifying subjects by NTD phenotype, we observed 4 polymorphisms statistically associated with NTDs subtypes. As to rs1130409, compared with TT genotype, GG genotype of was associated with the decreased risk for spina bifida, in female TG genotype was associated with higher risk of encephalocele compared with those harboring the TT genotype. Allele C of rs3136817 was associated with an increased risk for single encephalocele compared with those harboring allele T. In male, allele T of rs77794916 had a 1.547-fold increased risk for NTDs compared with allele C. GG genotype of rs1760944 was associated with an increased risk for cephalic malformations compared with those harboring the TT genotype. The SOD activity of fetal liver tissue in the control group was significantly higher than that of the case group. In male, the numbers of DNA AP sites in the NTD group were significantly larger than control group and statistically different among the three rs1760944 genotypes in female spina bifida. The results of this study imply that the APE1/Ref-1 gene polymorphisms were correlated with susceptibility to NTDs, oxidative stress and DNA oxidative damage was associated with NTDs in a high-risk area of China.

2732F

Whole exome sequencing of a girl with Rubinstein-Taybi syndrome. *H. Yoo[1,2], K. Kim[3], I. Kim[1], S. Rho[4], J. Park[1], S. Kim[5], N. Kim[3].* 1) Department of Psychiatry, Seoul National University Hospital, Seongnam, Gyeonggi, Korea; 2) Seoul National University College of Medicine, Seoul, Korea; 3) Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea; 4) Softmatter Research Center, Goyang, Gyeonggi, Korea; 5) Department of Pharmacology, Eulji University College of Medicine, Daejon, Korea.

Objectives: The Rubinstein-Taybi syndrome(RSTS) is a rare condition with a prevalence of 1 in 125,000~300,000, with dysmorphic features of face, hands & feet(Rubinstein et al., 1963). The genetic mutations of RSTS are not confirmed, though *CREBBP* or *EP300* mutations have been reported (Tsai et al., 2011). The purpose of this study is to evaluate the genetic causes of RSTS with Whole Exome Sequencing(WES). **Methods**: A 6-year-old Korean girl with RSTS was clinically phenotyped and behaviorally assessed with WPPSI, VABS, Leiter-R, ADOS, ADI-R, CBCL & Pittsburgh Sleep Quality Index(Park et al., 1996; Sparrow et al., 1994; Gale et al., 1997; Lord et al., 1994; LeCouteur et al., 2003; Achenback et al., 1991; Buysse et al., 1989). Blood samples are drawn from the proband and both biological parents. Read mapping, duplicate removal, local re-alignment, SNP and short indel genotyping have been performed by BWA, Picard, and GATK. Gene annotation and variant filtering is done by in-house bioinformatics pipeline and control database comprised of 54 normal Korean individuals. **Results**:(1)Phenotype: The proband is the only child of a healthy non-consanguineous couple. Her features include; 1) Typical facial dysmorphism: highly arched eyebrows, long eyelashes, down-slanting palpebral fissures, beaked nose with the nasal septum extending below the alae, dental crowding & micrognathia; 2) Broad & angulated thumbs & halluces, thickening of the soft tissue of the phalanges & persistent fetal pads; 3) Microcephaly, growth retardation(head circumference, Wt & Ht<3 percentile); 4) Congenital cataract; 5) Skin problems; 6) Swallowing difficulty; 7) EEG abnormality; 8) Mental retardation(IQ=37, PIQ=45) with severe language delay, motor delay & hyperactivity; 9) Autistic disorder fulfilling diagnostic criteria in all domains. (2) WES: In depth quality statistics, we obtained mean read depth coverage of 120x over the targeted genomic regions. We had found candidate mutations; 3 for de novo, 1 for homozygous, 9 for compound heterozygous models. Sanger sequencing validated 3 de novo variants; A frameshift mutation of *CREBBP*(c.2199delG) and missense mutations of *TNC*(c.323G>A) & *IGFALS*(c.1415C>T). **Conclusion**: We observed de novo mutations of *CREBB-BP, TNC* & *IGFALS* in WES of a girl with RSTS. It is the first report of *TNC* & *IGFFALS* as possible causative genes for RSTS. We presume that the mutation of *TNC* might be related to its neurological phenotypes, and *IGFALS* to growth retardation.

2733W

A *de novo* deletion at 16q24.3 involving *ANKRD11* in a Japanese patient with KBG syndrome. *S. Miyatake[1], A. Murakami[1], N. Okamoto[2], M. Sakamoto[3], H. Saitsu[1], N. Miyake[1], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University, Yokohama, Japan; 2) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 3) Takarazuka Municipal Center for Handicapped Children, Takarazuka, Japan.

KBG syndrome is a rare autosomal dominant congenital syndrome comprising developmental delay with various neurological involvements, macrodontia of the upper central incisors, characteristic facial dysmorphism, and skeletal anomalies. *ANKRD11* was recently identified as the gene responsible for this syndrome. To date, there have been only several KBG syndrome families described, each carrying a single base substitution or a small deletion of this gene. Here we present a Japanese patient with clinically confirmed KBG syndrome carrying a de novo deletion at 16q24.3 disrupting *ANKRD11*. He had characteristic facial appearance, macrodontia of the upper central incisors, hand anomalies, delayed bone age and intellectual impairment without autistic features. Hypoplasia of cerebellar vermis was pointed out by brain Magnetic Resonance Imaging. Copy number analysis using an Affymetrix Cytogenetics Whole-Genome 2.7M Array and Quantitative real-time PCR on the patient and his parents revealed that he had a *de novo* 690-kb deletion at 16q24.3 involving part of *ANKRD11*. Interestingly, the deleted region overlaps with the critical region for 16q24.3 microdeletion syndrome, featuring autosomal dominant intellectual impairment and autism spectrum disorder. We compared the clinical aspects of KBG syndrome to those of 16q24.3 microdeletion syndrome. Although dental information is scarce among the patients with 16q24.3 microdeletion syndrome, the other clinical features, such as facial dysmorphism or skeletal abnormalities, are similar. However neurological involvement seemed to be different. In 16q24.3 microdeletion syndromes, intellectual impairment is relatively mild and autistic spectrum disorders are frequently observed , while in KBG syndrome, intellectual disability range from mild to severe, and common behavioral disturbances are hyperactivity, attention deficit or easy frustration rather than autistic features. Our patient was very significant in that he had 16q24.3 microdeletion, but his neurological symptoms are more close to those of KBG syndrome. The other specific finding of our patient was hypoplasia of the cerebellar vermis, which had been reported as a rare complication of KBG syndrome, and not of 16q24.3 microdeletion syndrome. Our patient is significant for considering whether these two syndromes are different. It is necessary to study further patients with these two syndromes to clarify this issue.

2734T

Molecular investigations of Polish patients with Beckwith-Wiedemann syndrome. *D. Jurkiewicz[1], M. Kugaudo[1, 2], A. Tańska[1], E. Ciara[1], D. Piekutowska-Abramczuk[1], M. Pelc[1], S. Łuczak[1], J. Trubicka[1], M. Borucka-Mankiewicz[1], P. Kowalski[1], A. Jezela-Stanek[1], A. Cieślikowska[1], K. Chrzanowska[1], M. Krajewska-Walasek[1].* 1) Department of Medical Genetics, Children's Memorial Health Institute, Warsaw, Poland; 2) Department of Child and Adolescent Psychiatry, Medical University of Warsaw, Warsaw, Poland.

Beckwith-Wiedemann syndrome (BWS) is characterized by overgrowth, macroglossia, abdominal wall defects and a high risk of childhood tumors. BWS is caused by various 11p15 genetic or epigenetic defects leading to defective expression of imprinted genes. The genes in 11p15 region are organized into two imprinted domains controlled by two Imprinting Centers: IC1 (H19DMR) and IC2 (KvDMR). Molecular defects of the region resulting in BWS phenotype include loss of methylation at IC2 (50%), paternal UPD of 11p15 (20%), gain of methylation at IC1 (5%), mutations in *CDKN1C* gene (5%), and chromosomal rearrangements (2%). In ~20% of the patients a molecular alteration is unknown. Forty-nine patients with clinical symptoms of Beckwith-Wiedemann syndrome were investigated. Molecular analysis was performed by methylation sensitive multiplex ligation-dependent probe amplification (MS-MLPA) on leukocyte DNA. Hypomethylation in IC1 region was found in 23 patients including 2 monozygotic twins. Two patients had hypermethylation in IC2 region. Paternal UPD was identified in 8 patients and was subsequently verified by microsatellite analysis. Two patients carried duplications involving both IC1 and IC2 regions, that were confirmed by arrayCGH. Overall the study revealed the presence of mutations in 35 patients (71%). The molecular background of ~30% of studied cases is still unknown. In this group of patients the presence of mutations in the *CDKN1C* gene is possible (the analysis is under way). Moreover, failure to detect UPD in leukocytes due to somatic mosaicism associated with this etiology can not be excluded. There is also a possibility of the presence of molecular alterations in other genomic loci. The presented study is a first complex molecular characterization of a significant group of Polish BWS patients. The pattern of identified genetic defects is comparable with other western populations. The study was financed by National Science Centre, project no. 1149/B/P01/2011/40 (NN407114940) and EU Structural Funds, project POIG.02.01.00-14-059/09.

2735F

Gynecologic Issues in patients with Smith-Lemli-Opitz Syndrome. *M.A. Merideth[1], S.K. Conley[2], F.D. Porter[2].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Program in Developmental Endocrinology and Genetics, NICHD, NIH, Bethesda, MD.

Objective: To evaluate the gynecologic issues in females with Smith-Lemli-Opitz Syndrome Background: Smith-Lemli-Opitz Syndrome (SLOS) is a rare autosomal recessive multiple anomaly syndrome of abnormal cholesterol metabolism that is caused by mutations in 7-dehydrocholesterol reductase (DHCR7). Though the clinical spectrum is varied, manifestations include microcephaly, growth retardation, moderate to severe intellectual disability, postaxial polydactyly, 2-3 toe syndactyly and characteristic facial features. Current treatments include cholesterol supplementation, early intervention referral and physical/occupational/speech therapy. Despite an expanded phenotypic description of patients with SLOS, there is little information available about the gynecologic issues in this patient population. Methods: Seven females with SLOS were evaluated at the National Institutes of Health Clinical Center. Testing included gynecologic history, laboratory testing, review of outside records, physical examination and pelvic imaging if indicated. Results: The patients ranged in age from 3-36 years. Six patients had normal external genitalia; 1 had redundant perineal tissue. One patient had premature pubic hair development at age 3.5 years. Median age at menarche was 12.5 years (range 10-14 years). Four patients had irregular menstrual cycles and 5 had significant premenstrual mood swings/behavioral problems. Four patients were on hormones to regulate their cycles or help with premenstrual symptoms. One patient had to discontinue hormone use due to elevated liver transaminase levels. Two patients underwent hysterectomy due to difficulties managing menstrual periods. Dysmenorrhea was reported in 3 patients and suspected by parents in 2 patients. No patients had recurrent vaginal infections. No pregnancies were reported. Four patients underwent pelvic imaging with normal findings in all patients. Conclusions: There is a paucity of information in the literature about gynecologic issues in females with SLOS; families and physicians often have questions about puberty, management of menstrual periods and premenstrual behavioral issues. Questions also arise regarding vulnerability in this patient population. Gaining a better understanding of the range of gynecologic problems in SLOS will expand the phenotype and is the first step toward developing gynecologic therapeutic strategies for this patient population.

2736W

Maternal Uniparental Disomy 16 in an infant with intrauterine growth retardation, dysmorphic features, multiple congenital anomalies and dermatoglyphics features suggestive of chromosomal abnormalities: A neglected consideration. *Y. Lacassie[1], M. Narayanan[2].* 1) Dept Ped/ Div Clin Gen, LSU Hlth Sci Ctr and Children's Hospital, New Orleans, LA; 2) LSU School of Medicine in New Orleans.

At the 47th Annual Meeting of the ASHG in 1997, the first author reported that the absence of 'd' triradius predicted a 50% chance of a chromosomal abnormality. At the 59th meeting in Hawaii in 2009, he presented how dermatoglyphics, usually ignored in the examination, provide a clue to suspect cryptic chromosomal disorders, including microdeletions and duplications. We present a 3-month-old male with history of intrauterine growth restriction born to a 38yo mother at 366/7 WGA weighting 1,342g. On day 1 of life, the patient experienced severe respiratory distress. After appropriate intervention and stabilization, a complex congenital heart defect (CHD) and bilateral pulmonary hypoplasia were diagnosed and various dysmorphic features were noticed. Abnormally high levels of plasma amino acids, urine organic acids, and carnitine were found; however, the newborn screen was negative. Initial chromosomal analysis (450-500 bands resolution level) and aCGH (180,000 probes, 300kb resolution) were normal. The patient was transferred to Children's Hospital New Orleans at 53 days of life. Further work up showed multiple cardiovascular anomalies, including malaligned VSD, PFO, left pulmonary artery stenosis, non-compacted LV, hypoplastic aorta, and left SVC. Also, posteriorly rotated ears, anteverted nostrils, first degree hypospadias, cryptorchidism, right inguinal hernia, severe growth retardation, the presence of vertical crease in the soles secondary to gap between the hallux and second toe, single palmar flexion creases (type II at right), minor mesobrachydactyly of 5th fingers and absence of the right axial triradius with a radial arch on the hypothenar area were detected. Suspecting a cryptic chromosomal microdeletion, SNP microarray was requested. This showed normal dosage but two extended contiguous regions of allelic homozygosity restricted to 16p13.3p13.12 (13.4 Mb) and 16q21q23.3 (16.5 Mb), suggestive of UPD 16. Parental study (LabCorp) demonstrated the maternal origin of UPD. A literature review of 54 cases reveals that the major clinical features of patients with mat UPD 16 are highly variable. They may include moderate to severe IUGR, CHD, pulmonary hypoplasia, inguinal hernia, TEF, anal atresia, and dysmorphic facies. This patient shows that UPD 16 should be considered in the differential diagnosis of apparent chromosomal microdeletions or duplications as they may present similar phenotypic findings, including dermatoglyphics.

2737T

Audiovestibular findings in Myotonic dystrophy type 1 (DM1) patients from the National Rehabilitation Institute (INR). *M. Arenas-Sordo[1], B. Rivera-Mercado[2], D. Gutierrez-Tinajero[2], A. Martínez-García-Ramos[2], M. Trujillo-Bracamontes[2], O. Hernandez-Hernandez[1], J. Magaña-Aguirre[1].* 1) Genetics, INR, Mexico DF, Distrito Federal, Mexico; 2) Audiology, INR, Mexico DF, Distrito Federal Mexico.

Introduction.Myotonic dystrophy type 1(DM1)is a progressive and multisystemic degenerative neuromuscular disease, characterized by myotony and muscular weakness,secondary to an unstable mutation of CTG repeats in the DMPK gen locus 19q13.3. Precocious presbyacusis has been reported in DM1 patients,indicating a compromise of the auditory system. Purpose.The present work shows the results from audiological and vestibular evaluation in Mexican patients with DM1 from the National Rehabilitation Institute (INR).Methods.42 patients with molecular diagnosis for DM1 were analyzed. An extensive audiological evaluation was performed. Audiograms classified the range, nature and degree of hearing loss and speech, detected the sensorial or conductive nature of hearing loss. Tympanometry, acoustic reflex threshold, and transient otoacoustic emissions were part of the evaluation. The impact of the vestibular disease in daily life was evaluated by the Dizziness Handicap Inventory (DHI)test.Falling in patients with balance deficiency and vestibular disorders was measured by the dynamic gait index-(DGI).Results.From 84 ears analyzed, 26 showed normal hearing (30.95%),58 presented several degree of affectation(69.05%), being the most prominent effect the mild-sensorial hearing loss on low frequencies (22.62%)and mild sensorial hearing loss on high frequencies (19.05%).Most of the patients showed auditory alteration in both ears.High frequency audiometry revealed falls from 40 to 50 dB in 57% of the ears. The most prevalent timpanometry curves were that of type 'A' and 'As' with absence of acoustic reflex thresolds in the 60% of them. Transient otoacoustic emission revealed an inadequate response in 60% of ears. The emotional scale of vestibular evaluation indicated no disability in the 78.26% of the patients and moderate disability in 21.73%. Functional scale showed 21.73% and 8.69% of moderate and severe disability respectively. Physycal scale evaluation showed a 13.04% of moderate and severe disability respectively. Finally the DGI revealed a high risk of falling in 34.78% of the evaluated patients.Conclusions. Hearing loss in low and high frequencies indicate an affection of the base and apex of the cochlea where the outer hair cells play a key role in sound perception through a contractile-like movement. Future research efforts focused on the cells in the context of CTG repeats should be done to understand the pathophysiology of the disease at this level.

2738F

A case of Nonsyndromic hearing-impairment whit S199F GJB2 Homozygous mutation. *G. Giraldo, LP. Barragan Osorio, J.C. Prieto.* Instituto de Genètica Humana, Pontificia Universidad Javeriana , Bogotà , Bogotà D.C, Colombia.

Hearing impairment (HI) affects 1 in 650 newborns which makes it the most frequent congenital sensory impairment. Nonsyndromic HI is very heterogeneous and 37 different genes with a wide variety of functions have been identified. Mutations in GJB2 are the most common genetic etiology of prelingual non-syndromic sensorineural hearing loss. Rikkert et al reported in 2005 a total of 83 different mutations in GJB2; they found that 35delG mutation was the most common with an allelic frequency of 72.44%; they also found a reduced allelic frequency of 0,3% for the S199F mutation. In the study of Olarte M et al in Colombian population they reported that the most common mutations were 35delG and S199F; the allelic frequency for S199F mutation was 42.1%. In this paper we report a case of a Colombian 8 years old female child with a history of bilateral sensorineural hearing loss diagnosed at 5 years old. Her parents were non-consanguineous and her family history was unremarkable. She was born at term after a normal pregnancy. Mother denied exposure to teratogens or infections during pregnancy. In order to establish the etiology of her sensorineural hearing loss a sequence analysis of the gene GJB2 and GJB6 was performed, the report showed S199F homozygous mutation in GJB2 and a normal sequence in GJB6. Most patients with mutations in GJB2 present a prelingual hearing loss but in this case our patient has a S199F mutuation in GJB2 with a non classical postlingual bilateral sensorineural hearing loss and an atypical audiometry pattern.

2739W

Somatic CTNNB1 mutation in hepatoblastoma from a patient with Simpson-Golabi-Behmel syndrome and germline GPC3 mutation. *R. Kosaki[1], T. Takenouchi[2], N. Takeda[3,4], M. Kagami[5], K. Nakabayashi[6], K. Kosaki[2,7].* 1) Div Med Genet, Natl Ctr Child Hlth & Dev, Tokyo, Japan; 2) Dept of Pediatrics, Keio Univ Sch Med, Tokyo, Japan; 3) Dept of Surgery ,Natl Ctr Child Hlth & Dev, Tokyo, Japan; 4) Dept of Surgery ,Kitasato Univ,Kanagawa,Japan; 5) Dept of Molecular Endocrinology, Natl Research Institute of Child Hlth & Dev, Tokyo, Japan; 6) Dept of Maternal-Fetal Biology, Natl Research Institute of Child Hlth & Dev, Tokyo, Japan; 7) Ctr Med Genet, Keio Univ Sch Med, Tokyo, Japan.

Simpson-Golabi-Behmel syndrome is a rare overgrowth syndrome caused by the GPC3 mutation at Xq26 and is clinically characterized by multiple congenital abnormalities, mental retardation, pre/postnatal overgrowth, distinctive craniofacial features, macrocephaly, and organomegaly. Although this syndrome is known to be associated with a risk for embryonal tumors, similar to other overgrowth syndromes, the pathogenetic basis of this mode of tumorigenesis remains largely unknown. Here, we report a boy with Simpson-Golabi-Behmel syndrome who had a germline loss-of function mutation in GPC3. At 9 months of age, he developed hepatoblastoma. A comparison of exome analysis results for the germline genome and for the tumor genome revealed a somatic mutation, p.Ile35Ser, within the degradation targeting box of beta-catenin. The same somatic mutation in CTNNB1 has been repeatedly reported in hepatoblastoma and other cancers. This finding suggested that the CTNNB1 mutation in the tumor tissue represents a driver mutation and that both the GPC3 and the CTNNB1 mutations contribute to tumorigenesis in a clearly defined sequential manner in the propositus. The current observation of a somatic CTNNB1 mutation in a hepatoblastoma from a patient with a germline GPC3 mutation supports the notion that the mutation in GPC3 may influence one of the initial steps in tumorigenesis and the progression to hepatoblastoma.

2740T

Establishment and validation of iPS cells and knockout mice for dermatan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos syndrome (DDEDS). *T. Kosho[1,2], F. Yue[3], S. Saka[2,4], N. Tsumita[2,4], Y. Kasahara[2], T. Okada[2], S. Mizumoto[5], M. Kobayashi[6], J. Nakayama[7], N. Miyake[8], Y. Nomura[4], T. Era[9], A. Hatamochi[10], F. Fukushima[1], N. Matsumoto[8], K. Sugahara[5], K. Sasaki[3], S. Takeda[2].* 1) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 2) Department of Molecular Therapy, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan; 3) Department of Histology and Embryology, Shinshu University School of Medicine, Matsumoto, Japan; 4) Scleroprotein and Leather Research Institute, Tokyo University of Agriculture and Technology, Faculty of Agriculture, Fuchu, Japan; 5) Laboratory of Proteoglycan Signaling and Therapeutics, Hokkaido University Graduate School of Life Science, Sapporo, Japan; 6) Miya Kobayashi: Department of Food and Nutritional Environment, College of Human Life and Environment, Kinjo Gakuin University, Nagoya, Japan; 7) Department of Molecular Pathology, Shinshu University Graduate School of Medicine, Matsumoto, Japan; 8) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 9) Department of Cell Modulation, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto, Japan; 10) Department of Dermatology, Dokkyo Medical University, School of Medicine, Mibu, Japan.

Dermatan 4-O-sulfotransferase 1 (D4ST1)-deficient Ehlers-Danlos syndrome (DDEDS), caused by recessive loss-of-function mutations in CHST14, is a recently delineated form of EDS [Kosho et al., 2011], characterized by a unique set of clinical features consisting of progressive multisystem fragility-related manifestations (skin hyperextensibilty and fragility, progressive spinal and foot deformities, large subcutaneous hematoma) and various malformations (facial features, congenital multiple contractures). Multisystem connective tissue fragility is caused by impaired assembly of collagen fibrils through loss of dermatan sulfate (DS) replaced by chondroitin sulfate in the decorin glycosaminoglycan sidechains. Complete loss of DS in patients' urine was also detected. Because the patients suffer from progressive multisystem fragility-related complications, appropriate disease modeling is indispensable in view of developing etiology-based therapy. In this study, we report establishment and validation of induced-pluripotent stem cells (iPSCs) and knockout (Chst14-/-) mice. Cultured skin fibroblasts were obtained from a patient and transduced into iPSCs. Morphologically, patient-derived iPSCs (P-iPSCs) had a smaller size and more vacuoles than control iPSCs (C-iPSCs) derived from a healthy individual. Apoptosis, undifferentiation status, or pluripotency were not different between these iPSCs. Decorin staining on teratoma samples derived from P-iPSCs using human decorin antibody was weaker than that from C-iPSCs, similar to the results from patients' skin specimens. Neural progenitor cells and neurons were differentiated, and neural differentiation potency was weaker in P-iPSCs than that in P-iPSCs. Frozen sperm from Chst14+/- male mice were obtained from the Mutant Mouse Regional Resource Center. After reproducing Chst14+/- mice, Chst14-/- mice were generated and preliminary phenotypic studies were undertaken. Complete loss of DS in urine from Chst14-/- mice was demonstrated, suggesting these mice to reflect glycobiological abnormalities in DDEDS. Although postnatal weight gain was smaller in Chst14-/- mice than in Chst14+/- or WT mice, no apparent congenital multiple contractures or multisystem fragility-related manifestations were noted. Further studies to delineate pathophysiological features of cells from other systems differentiated from iPSCs and longitudinal physical features including pathological abnormalities of Chst14-/- mice are necessary.

2741F

Report of a Colombian case of Werner Mesomelic Syndrome with eight fingers in both hands and feet. *J.C Prieto[1,2], G. Giraldo[1], T. Pineda[1].* 1) Inst de Genetica Humana, Univ Javeriana, Bogota, Colombia; 2) Hospital La Victoria. Secretaria Distrital de Salud Bogota, Colombia.

Werner mesomelic syndrome (WMS) is an autosomal dominant disorder first described by Werner in 1912. It is characterized by hypo or aplasia of the tibiae with remarkable short stature associated with preaxial polydactyly of the hands and feet and/or five-fingered hand with absence of thumbs. Wieczorek et al identified the molecular basis of the syndrome in 2010; they established that this condition is caused by a specific point mutation in the ZRS (the sonic hedgehog regulatory region) at position 404. In this paper we report the case of a Colombian 1 year old infant patient with Werner mesomelic syndrome. She was the first daughter of non-consanguineous young parents and her family history was unremarkable; she presented with rizomelic and mesomelic shortening of the lower limbs with bilateral hypoplasia of the tibiae, right femur hypoplasia and bilateral absence of the hallux; she also presented bilateral triphalangeal thumb and preaxial polydactyly with 8 fingers in both hands and feet. This is the first case of Werner mesomelic syndrome reported in Colombia and we describe new phenotypic findings.

2742W

A Novel WNT7A Mutation Causes Autosomal Recessive Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome. *Y. Sahin[1], PO. Simsek-Kiper[2], A. Cetinkaya[1], GE. Utine[2], K. Boduroglu[2].* 1) Department of Medical Genetics, Hacettepe University, Ankara, Turkey; 2) Department of Pediatrics, Department of Pediatric Genetics, Hacettepe University, Ankara, Turkey.

Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome is a very rare autosomal recessive disease presenting with pelvic hypoplasia, thoracic dystrophy and severe limb defects, especially affecting ulnae and fibulae. A 2-year old girl with several features of this syndrome including bilateral ulnar aplasia, hand defects, fibular aplasia, pelvic hypoplasia and rib fusion anomalies sought medical advice in our genetics clinic. Her limb abnormalities were noticed in the prenatal ultrasound study. She had normal psychomotor development, but she had growth delay. Her abdominopelvic ultrasonography revealed left kidney agenesis which was also confirmed by scintigraphy. She had no other systemic abnormalities. Her parents are from Southeastern part of Turkey and are first cousins. In the light of these clinical findings, Sanger sequencing of patient blood DNA was performed for all 4 exons of WNT7A, which revealed c.713T>G mutation in homozygous state while her parents and two sibs carried this mutation in heterozygous state and were healthy. This mutation is not reported in dbSNP database and is not one of the 6 mutations previously linked to Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome. This mutation is predicted to cause a missense mutation substituting valine in the 238th position to glycine [p.Val238Gly]. This site is conserved among vertebrates and mutation prediction programs, Mutation Taster and Polyphen-2, predict that this mutation is disease causing. WNT7A is an effector of wnt/β-catenin signalling pathway. It is suggested that this specific type of wnt protein is involved in limb and genitourinary development. The phenotypic features of the patient presented here are in compliance with reported effects of WNT7A deficiency in human. To our knowledge, this patient represents the second case of unilateral renal agenesis reported in this syndrome, emphasizing the role of WNT7A in renal development. In addition, the case of this patient adds c.713T>G mutation to the spectrum of WNT7A mutations causing Al-Awadi/Raas-Rothschild/Schinzel Phocomelia Syndrome.

2743T

Three generation family with kyphomelic dysplasia suggests autosomal dominant inheritance. *O. Vanakker[1], U. Fränkel[2], B. Callewaert[1].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Department of Paediatrics, Zorgsaam Hospital, Terneuzen, The Netherlands.

Kyphomelic dysplasia (KD) is a rare skeletal dysplasia belonging to the group of bent bone dysplasias. It represents a heterogenous group of disorders, with at least three distinct entities identified: Schwartz-Jampel syndrome (SJS), Cartilage Hair Hypoplasia (CHH) and 'true' KD. For the latter, the genetic background is unknown and autosomal recessive inheritance is presumed based on several sporadic cases with parental consanguinity. We report a three-generation family presenting with a true KD phenotype. The proband was born at term after an uneventful pregnancy. Her length and weight at birth were 37cm and 1578g respectively. She had severe rhizomelic shortening of upper and lower limbs and radiographs showed stubby femora which were extremely curved. Iliac wings were short and widened; the humeri were also short and dumbbell-shaped. The mother of the proband has a similar disproportionate short stature phenotype. Radiographs taken at birth revealed severe femoral bowing, though less prominent compared to her daughter. Follow-up radiographies in the mother demonstrated a gradual decrease of the bowing over several years. The maternal grandfather of the proband had an identical phenotype. All had a normal neuromotor development. There were no signs of myotonia or ocular problems (SJS) nor evidence for immune deficiencies or anaemia (CHH). This family, suffering from true KD, suggests that at least a subtype of KD has an autosomal dominant mode of inheritance. This observation has important consequences for genetic counselling of patients with KD, and identification of other families may allow clarification of the genetic background.

2744F

Further evidence of brain anomalies related to ALX4 mutations: possible genotype-phenotype correlation. *T. Almeida[1], M. Valente[1], G.L. Yamamoto[2], C.A. Kim[1], M.R. Passos-Bueno[2], D.R. Bertola[1,2].* 1) Genetic Unity, Children's Institute, São Paulo, São Paulo, Brazil; 2) Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, São Paulo, Brazil.

Brain anomalies has been found in patients with loss-of-function mutations in ALX4, either in the homozygous state, with the frontonasal phenotype, showing abnormalities in corpus callosum and cerebellar hypoplasia, or in the heterozygous state, with parietal foramina, presenting polymicrogyria, high insertion of the tentorium cerebri, bilateral choroid plexus cysts, corpus callosum agenesis and mild cerebellar hypoplasia. Here we describe the cranial MRI of four cases with ALX4 mutations. Patient 1, with a homozygous missense mutation in the homeodomain (p.R215W) and a frontonasal phenotype, showed occipital and cerebellar hypoplasia, high insertion of the tentorium cerebri, agenesis of the septum pelucidum and occipital simplified gyra, with no abnormalities in corpus callosum, previously described for all patients with the frontonasal phenotype. Patient 2, heterozygous for the same mutation, showed no brain anomalies. Patients 3 and 4 (mother and son) with a heterozygous frameshift mutation after the homeodomain (p.D326fs*21) and a frontonasal phenotype, showed no structural anomalies. Contrary to what is observed in the literature for the frontonasal patients, our study describes two individuals with this phenotype without brain anomalies. We believe that this could be related to the preservation of the homeodomain since this is the only family with a mutation that does not encompass that region. The investigation of brain anomalies in the heterozygotes for the ALX4 mutations with parietal foramina were carried out just for a small number of patients. Brain malformations were described in a family with a missense mutation at the homeodomain and in some patients with a microdeletion of this region. On the other hand, a patient harboring a frameshift mutation in ALX4 before the homeodomain, associated with microdeletion in the region 22q11, showed no brain anomalies. This could lead to the conclusion that the brain phenotype has an incomplete penetrance in the heterozygous state and cannot be related to the presence of the parietal foramina, or the position and type of the mutations. This study is a preliminary investigation about the association of the ALX4 gene with brain malformations and requires further studies to establish the real extension and genotype-phenotype correlation of the brain anomalies.

2745W

Further occurrence of "TMCO1 defect syndrome` in a non-Amish population: the first South American patient presenting the same p.R87X mutation recently found in a Turkish individual. *W.A.R. Baratela[1], G.L. Yamamoto[1], M.R. Passos-Bueno[2], D.R. Bertola[1,2].* 1) Unidade de Genética, Instituto da Criança - Hospital das Clínicas - Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Centro de Estudos do Genoma Humano - Departamento de Genética e Biologia Evolutiva - Instituto de Biociências - Universidade de São Paulo.

Autosomal recessive syndrome of craniofacial dysmorphisms (brachycephaly, flat face, arched bushy eyebrows, synophrys, orbital hypertelorism, wide nose bridge, high-arched palate, cleft lip and palate, microdontia of primary teeth, gingival hyperplasia and low-set ears), skeletal anomalies (Sprengel deformity, pectus excavatum, rib anomalies, vertebrae fusions and club foot), and developmental delay/mental retardation associated with TMCO1 defects was first diagnosed in 11 individuals, all from the Old Order Amish of northeastern Ohio. TMCO1 homozygous nonsense mutation in the coiled coil region of the protein was found in all 9 tested patients, with a carrier frequency in the community of 0,7%. Recently a similar case from Turkey was published, with a different mutation in TMCO1, but in the same protein region, with a rather more profound developmental impairment. We describe a 28 yo lady, daughter of a first cousin Brazilian couple, with normal birth measurements. She presented with tall stature, bushy and arched eyebrows, synophrys, ocular hypertelorism, divergent strabismus, cleft lip and palate and gingival hypertrophy. She also had polydactyly, presented in other members of the family in an autosomal dominant pattern. She walked at 2y4mo, first words with 1y6m, attended special education from the beginning. At 3yo developed seizures and started with antiepileptic drugs. Skeletal survey showed rib fusions and scoliosis. Ecocardiogram, and abdominal ultrasound were normal. Molecular studies with exome sequencing found a homozygous mutation in TMCO1, c.C259T (p.R87X), the same reported in Turkey. We describe a further occurrence of 'TMCO1 defect syndrome' in a non-Amish population: the first South American patient, indicating that this disorder is probably more widespread. The family of our patient is from a small city in the Northeast of Brazil, in which there was Portuguese colonization. It is therefore plausible to consider that this mutation has European origin, shared by the Turkish patient. This far, the clinical findings seem quite homogenous, but as this is a new disorder, its full spectrum remains to be depicted.

2746T

Sibs with Hydrops fetalis, arthrogryposis multiplex congenital, neuronal migration disorder, adrenal and pulmonary hypoplasia and renal abnormalities associated with a mutation in the FAT4 gene, encoding a giant cadherin. *D. Chitayat[1,2], T. Uster[1], P. Shannon[3], M. Srour[4], S. Robertson[5], J. Michaud[4].* 1) Prenatal Diag & Med Gen, Mount Sinai Hosp, Toronto, ON, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 3) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Ste-Justine Hospital Research Center, Montreal, Quebec, Canada; 5) Department of Pediatrics and Child Health, Dunedin School of Medicine, University of Otago, New Zeland.

The human FAT gene family consists of the FAT1, FAT2, FAT3 and FAT4 genes. Human FAT family genes encode large proteins with extracellular cadherin repeats, EGF like domains, and laminin G like domain(s). FAT4 is involved in the maintenance of planar cell polarity and inhibition of cell proliferation. FAT4 gene has been reported to be mutated in several types of human cancers, such as melanoma, pancreatic cancer, gastric cancer and hepatocellular carcinoma. However, so far no human condition has been reported to be caused by germline mutations in this gene. We report a male and female sibs, born to a consanguineous Turkish couple, with hydrops fetalis (HF), neuronal migration disorder, joint contractures, pulmonary and adrenal hypoplasia and renal abnormalities. To our best knowledge this is a hitherto new, autosomal recessive condition. The parents were healthy and first cousins of Turkish descent. Patient A, a male, was born at 31.8 weeks gestation following a pregnancy complicated with hydrops fetalis, mild ventriculomegaly, misalignment of the skull bones and lumbar lordosis. He died shortly after delivery and autopsy showed HF, cleft palate, ambiguous genitalia, bilateral talipes equinovarus, camptodactyly of the fingers and clenched fists. There was a butterfly T5 vertebra and hypoplastic lungs, kidneys and adrenal glands. The brain showed megalencephaly, abnormal gyration and an unusual pattern of cytoarchitectural disturbance in the cerebral cortex. The karyotype was 46, XY. Patient B, a female, the product of the couple's second pregnancy, was found to have an increased nuchal translucency and body edema on 12 weeks ultrasound. Ultrasound at 20 weeks gestation showed bilateral talipes equinovarus, flexed wrists, echogenic kidneys, pericardial effusion and an abnormal cerebral configuration. The pregnancy was terminated at 21 weeks gestation and the autopsy showed the same findings as in her late brother apart from her female internal and external genitalia. Microarray analysis was normal and female. Extensive investigation done on the two failed to identify etiology. Exome sequencing showed a homozygous mutation in the FAT4 gene.

2747F

Genetic heterogeneity in Mabry syndrome: a novel phosphatidylinositol glycan (GPI) anchor deficiency disorder. *D. Cole[1], T. Roscioli[2], M. Nezarati[3], E. Sweeney[4], P.N. Krawitz[5], H. van Bokhoven[6], C. Marcelis[6], B. DeVries[6], D. Andrade[7], W.M. Burnham[8], A. Munnich[9], M. Thompson[1].* 1) Dept Laboratory Medicine, Univ Toronto, Toronto, ON, Canada; 2) Dept of Women's and Children's Health, Sydney Childrens' Hospital and University of New South Wales; 3) Dept of Genetics, North York General Hospital, Toronto, Canada; 4) Royal Liverpool Children's Hospital, Liverpool, UK; 5) Inst für Medizinische Genetik, Charité Universitäts Medizin, Berlin, Germany; 6) Dept of Human Genetics, University Medical Centre St. Radboud, Nijmegen; 7) Division of Neurology, Toronto Western Hospital, Toronto, Canada; 8) Department of Pharmacology, University of Toronto; 9) INSERM U781-Université Paris Descartes-Hôpital Necker-Enfants Malades, Paris.

Mabry syndrome (hyperphosphatasia with developmental disability) was first described in 1970 (OMIM#239300). At first considered rare, improved phenotyping has led to recruitment of more than twenty families world-wide. Salient features of the disorder include: characteristic facial dysmorphology (hypertelorism, a broad nasal bridge and a tented mouth); variable shortening of middle and distal phalanges; and neural abnormalities on biopsy (plaques disrupting Schwann cells). Like many infantile metabolic storage disorders, seizures associated with Mabry syndrome have an onset in the first year of life. Persistently elevated serum alkaline phosphatase (ALP) levels are now known to be associated with abnormalities of the phosphatidylinositol glycan (GPI) anchor due to mutations in the PIGV, PIGO and PGAP2 genes. Lysosomal storage material detected in some patients with Mabry syndrome has been putatively identified as glycolipid in nature - possibly resulting from changes in lipid raft functions that result from GPI anchor disruption. Not all patients with a clinical diagnosis of hyperphosphatasia with neurologic deficit result from known GPI disruptions. We present data on patients that suggest there are further subtypes of the disorder.

2748W

Exome sequencing of the X-chromosome in Aicardi syndrome. *C. Lund, H.S. Sorte, Y. Sheng, M.D. Vigeland, O. Røsby, K.K. Selmer.* Oslo university Hospital, Oslo, Norway.

Exome sequencing of the X-chromosome in Aicardi syndrome C. Lund 1, H. S. Sorte2, Y. Sheng2, M. D. Vigeland2, O. Rosby 2, K. K. Selmer2. 1) National Centre for Rare Epilepsy-related Disorders, Oslo University Hospital, Norway, 2) Department of Medical Genetics, Oslo University Hospital, Norway. Background: Aicardi syndrome is a rare neurodevelopmental disorder, a well-defined syndrome since its first description in 1965. The main diagnostic features are agenesis of corpus callosum, chorioretinal lacunae and infantile spasms. The etiology is unknown. However, it is only seen in females and 47, XXY males in healthy families. Therefore, it is assumed to be caused by de novo mutations on the X-chromosome, inherited in a dominant manner with hemizygous lethality in males. Methods: To search for disease-causing mutations, we performed whole exome sequencing on seven patients with Aicardi syndrome. The exome was enriched by using the Agilent SureSelect Human All Exon Capture kit (50 Mb, v3). Sequencing was performed with 100 bp PE reads on an Illumina HiSeq2000, yielding an average coverage of 25x-135x. We identified 17 000-22 000 genetic variants per patient. The filtering process was focused on identifying rare, protein-altering and heterozygous variants, shared by multiple patients. Results: After the filtering process, none of the remaining variants appeared in more than one patient. Conclusions: With the method of exome sequencing of the X-chromosome of seven patients, we were not able to reveal the etiology of Aicardi syndrome. Assuming that the hypothesis of X-linked dominant inheritance is correct, the results of this study suggest that the causal locus lies in a non-coding region or in a gene poorly covered by the exome sequencing kit. However, further studies of these patients will involve trio testing and examination of the autosomal chromosomes.

2749T

Mutations in extracellular matrix genes NID1 and LAMC1 cause autosomal dominant Dandy-Walker malformation and occipital cephaloceles. *V. Mahajan[1], B. Darbro[2], L. Gakhar[3], J. Skeie[1], E. Campbell[2], S. Wu[2], X. Bing[2], K. Millen[4], W. Dobvns[4], J. Kessler[5], A. Jalali[5], J. Cremer[6], A. Segre[6], J. Manak[2,7], K. Aldinger[4], S. Suzuki[8], N. Natsume[8], M. Ono[8], H. Dai Hai[9], L. Thi Viet[9], S. Loddo[10,11], E. Valente[10,12], L. Bernardini[10], N. Ghonge[13], P. Ferguson[2], A. Bassuk[2].* 1) Department of Ophthalmology, University of Iowa, Iowa City, IA; 2) Department of Pediatrics, University of Iowa, Iowa City, IA; 3) X-ray Crystallography Facility, University of Iowa, Iowa City, IA; 4) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 5) Department of Neurology, Northwestern University Evanston, IL; 6) Department of Computer Science, University of Iowa, Iowa City, IA; 7) Department of Biology, University of Iowa Iowa City, IA; 8) Aichi-Gakuin University, Japan; 9) Odonto Maxillo Facial Hospital, Vietnam; 10) IRCCS Casa Sollievo della Sofferenza, Italy; 11) Department of Experimental Medicine, Sapienza University, Rome, Italy; 12) Department of Medicine and Surgery, University of Salerno Salerno, Italy; 13) Diwanchand Satyapal Aggarwal Imaging Research Centre, India.

Background The Dandy-Walker spectrum of disorders including autosomal dominant Dandy-Walker malformation and occipital cephaloceles (ADDWOC) are characterized by variable cerebellar hypoplasia, meningeal anomalies, and occipital skull defects. We have previously reported deletions that encompass FOXC1 or ZIC genes in rare familial and sporadic cases of Dandy-Walker spectrum disorders, but ADDWOC families do not have mutations in these genes. Thus, we aimed to identify the genetic etiology underlying ADDWOC. Methods We previously characterized the phenotype of a family with ADDWOC of Vietnamese origin, and a second family with a similar phenotype from India. Whole exome sequencing was performed to identify candidate mutations in the larger Vietnamese pedigree, and whole exome sequencing followed by protein interaction network analysis was performed to identify a mutation in the Indian pedigree. A known X-ray crystallography structure of the interacting proteins was used to elucidate the effect of the mutations on the interacting proteins. Findings We identified a mutation in the extracellular matrix protein encoding gene NID1 in the original Vietnamese pedigree. In the Indian family, protein interaction network analysis identified a mutation in LAMC1, which encodes a NID1 binding partner. Structural modeling the NID1-LAMC1 complex demonstrated that each mutation disrupts the interaction. Interpretation These findings implicate the extracellular matrix in the pathogenesis of Dandy-Walker spectrum disorders.

2750F

Mapping a new locus for autosomal dominant nonsyndromic comitant strabismus.
M. Patel[1,3], X. Ye[1,2,3], C. Shyr[2,3], Z. Zong[1,3], M. Thomas[1], P. Power[1], N. Roslin[4], S. Narasimhan[5], D. Giaschi[5], W. Wasserman[1,2,3]. 1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics; 3) Child and Family Research Institute; 4) FORGE Canada Consortium; 5) Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC, Canada.

Strabismus, defined as a lack of parallel gaze, is a common problem found in 1-4% of the general population yet little is known about its genetic basis. Strabismus is found at an elevated frequency in a large number of different syndromes, precluding the use of syndromology to unravel the genetic basis of this common disorder. The lack of common molecular or cell biologic features between all the syndromes featuring strabismus may also suggest that there are many possible disruptions that can result in an interruption of parallel gaze. One way to address this complexity is to study families with Mendelian inheritance of the trait of interest. We report here a 7 generation pedigree with many affected family members in an autosomal dominant pattern. Affected members had isolated esotropic or hypertropic comitant strabismus that was nearly always congenital and of varying severity. No other ocular or neurological abnormalities segregated in the pedigree and there were no minor anomalies that segregated with the trait. A high density genome-wide mapping study was undertaken in parallel with exome sequencing. Results will be reported at the meeting. Study of this rare family provides an opportunity to uncover a molecular entry point into the biology of nonsyndromic strabismus.

2751W

Exome sequencing identifies mutations in a gene not previously related to skeletal dysplasias in patients with spondylometaphyseal dysplasia.
G.L. Yamamoto[1,2], W.A.R. Baratela[1], T.F. de Almeida[1], C.A. Kim[1], M.R.S. Passos-Bueno[2], D.R. Bertola[1,2]. 1) Genetica Medica ICR-FMUSP, Universidade de Sao Paulo, Sao Paulo SP, Brazil; 2) Centro de Estudos do Genoma Humano IB, Universidade de Sao Paulo, Sao Paulo, SP, Brazil.

We report on four cases, two cousins, 12yo and 15yo and two siblings, 2yo and 17yo, from two distinct Brazilian consanguineous families. The families are from two different, small cities in the Northeast of Brazil. The proband in the first family was born at term, with normal birth measurements. She developed progressive bowing of the lower limbs at 1yo. Skeletal survey showed platyspondyly and marked metaphyseal flaring since early infancy, with normal to mildly enlarged epiphysis. No other abnormalities were observed, including normal ophthalmologic evaluation. Skeletal findings in the second family were strikingly similar to the previous family, however, the siblings also presented progressive visual impairment with macular pigmentary changes. Exome sequencing identified a region of loss of heterozigozity in chromossome 3q29 in both families. Two new homozygous mutations, a frameshift in family one and a missense in family two, never before described in control populations, were identified in a single gene within 3q29 that we believe to be associated with the skeletal phenotype. We further hypothesized that a second gene in region 3q29 could be responsible for the visual impairment in family 2, since a homozygous, possibly pathogenic, missense mutation was also identified in a vitamin A metabolism related gene.

2752T

A novel missense mutation in PRPS1 leads to PRS-I deficiency in females displaying retinitis pigmentosa and variable expression of a neurologic phenotype.
B. Almoguera[1], J. Liang[2], P. Fernandez[3], M. Corton[3], Y. Guo[1], B. Keating[1,4,5], J. Zhang[2], H. Hakonarson[1,4,5], X. Xu[2], C. Ayuso[3]. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Department of Genetics and Genomics. IIS-Fundacion Jimenez Diaz Hospital. 28040, Madrid, Spain; 4) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

Phosphoribosyl pyrophosphate synthetase I (PRS-I) is a key enzyme in nucleotide metabolism. Missense mutations in PRPS1 may lead to PRS-I deficiency and cause three different phenotypes: Arts Syndrome (MIM 301835), X-linked Charcot-Marie-Tooth (CMTX5, MIM 311070) or X-linked non-syndromic sensorineural deafness (DFN2, MIM 304500). All three have in common to be rare, X-linked recessively inherited and to display variable levels of central and peripheral neuropathy in males affected. In the current study, we have identified a novel missense mutation in PRPS1 leading to PRS-I deficiency in 4 affected females across 3 generations of a family with a phenotype consisting of retinitis pigmentosa (RP) and variable expression of optic atrophy, ataxia, peripheral neuropathy and hearing loss. Four members from the family were subjected to whole exome sequencing. After variant filtering analysis assuming dominant inheritance, a novel missense variant in PRPS1 was selected among 107 nonsynonymous variants segregating in the family. Sanger sequencing in 12 members of the family showed complete segregation of the mutation with the disease. PRS-I enzymatic activity in erythrocytes from three affecteds evidenced variable levels of deficiency, with severity correlating with the age of onset of ophthalmologic manifestations. Further assessment of X inactivation skewing confirmed different patterns of X inactivation in the four affecteds, which may be also responsible for the variable phenotypic expression in the females. However, since no male was found affected, a lethal effect of the mutation in males cannot be ruled out. In addition to the observation of PRS-I deficiency in females, this is the first time that RP is part of the myriad of symptoms these syndromes display, being in this family the first and only common manifestation to all affecteds. Visual cycle is especially sensitive to GTP levels, which are expected to be low in the family, and that may lead to an eventual death of photoreceptor cells and would explain the RP phenotype. However, further functional studies are needed in order to elucidate the mechanism underlying RP. This study also highlights the enormous phenotypic variability observed in PRS-I related syndromes inter and intra family.

2753F

Mutation analysis of androgen receptor gene: A complementary molecular approach in management of androgen insensitivity syndrome.
F. Baghbani-arani[1], A. Shojaei[2], F. Behjati[3], R. Ebrahimzadeh-Vesal[2], J. Tavakkoly-Bazzaz[2]. 1) Genetics Dept, Varamin-Pishva Branch,Islamic Azad University, Varamin, Iran, varamin, Iran; 2) Department of Medical Genetics, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; 3) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Androgen receptor gene mutations are the leading cause of sex development disorders exhibited by sexual ambiguity or sex reversal. In this study, seven families containing patients clinically diagnosed as androgen insensitivity syndrome (AIS) were genetically evaluated through cytogenetic and molecular analysis including karyotype, polymerase chain reaction (PCR) and sequencing of SRY and AR genes. Two brothers and their mother were hemizygous and heterozygous respectively for c.2522G>A variant, while one of their healthy brother was completely normal hemizygous. Family 2 assessment demonstrated the c.639G>A (rs6152) variant in two siblings who were reared as girls. The SRY gene was intact in all of the study's participants. We could not find any mutation or polymorphism in neither AR nor SRY genes of the other five patients from five families. Our findings in family 1 could be a further proof for the pathogenicity of the c.2522G>A variant. Given the importance of AR mutations in development of morbidities such as disabling sex life and in particular prostate cancer in AIS patients, definitive diagnosis based on molecular genetic approaches accompanied by comprehensive genetic counseling could have promising impacts in clinical management and also in prenatal diagnosis of prospect offspring.

2754W

Ocular Findings in the Marfan Syndrome. *I.H. Maumenee[1], S. Wehrli[2], WW. Xu[2], S. Rahmani[2], S. Kurup[2], I. Kassem[1], M.K. Durbin[2], N. Azar[1], A.A. Fawzi[2], A. Lyon[2], M.B. Mets[2], Chicago Marfan Eye Consortium.* 1) Ophthalmology, University of Illinois Eye and Ear Infirmary, Chicago, IL; 2) Lurie Children's Hospital of Chicago, Chicago, IL.

Purpose: To identify diagnostic criteria and risk factors for vision loss in Marfan patients based upon clinical examination, biometry and OCT imaging of the posterior segment. Material and Methods: 73 patients were seen at Children's Hospital Ophthalmology Clinic in Chicago during the 2012 National Marfan Foundation Conference. Fifty-six patients fit diagnostic criteria for Marfan syndrome and underwent the following examinations: slit-lamp, indirect ophthalmoscopy, biometry, OCT measurements using the Macular Cube 512x128 and Optic Disc Cube 200x200 of the Cirrus HD-OCT (Carl Zeiss Meditec, Dublin, CA). Criterion for OCT inclusion was signal strength ≥ 6 out of 10 without artifacts; the data were analyzed using Cirrus software. Retinal examinations were performed by a vitreoretinal specialist. Color ultra-widefield retinal images were obtained using the Optos 200Tx on 54 patients (108 eyes). The ultra-widefield images were evaluated by a masked vitreoretinal specialist. The color composite images were analyzed with Optos imaging review software. Results were analyzed using Pearson's correlation, 2-tailed tests or one sample t-test. Results: Mean age of patients was 20.7 y (range 2 to 56). Keratometry obtained on 54 eyes showed flattening of the cornea (mean K=41.58D; SD=2.09), ↑corneal thickness, (mean=542.2 µm; SD=42.25); deepened anterior chamber, (mean=3.6mm; SD=0.61), reduced lens diameter (mean=3.84mm; SD=0.48); and increased axial length, (mean=25.39mm; SD=1.91). A dislocated lens was observed in 32% of patients. OCT data showed that 27% of eyes had RNFL thickness <5% of population, suggestive of RNFL loss. There were less significant macular OCT changes. RNFL thickness and macular volume were negatively correlated with axial length but not correlated to age, IOP, or C:D ratio. Peripheral retinal pathology (e.g.: lattice, scars, RD, holes, white without pressure) was identified in 20% of eyes on clinical examination compared with 24% of eyes analyzed using Optos 200 TX. Corneal curvature, AC depth, lens thickness and axial length of the globe (p<.05) can contribute significantly to the diagnosis of Marfan syndrome. 27% of eyes had RNFL thickness <5% suggestive of RNFL loss; 24% had peripheral retinal pathology, both indicative of need for close patient observation.

2755T

Peace Sign Craniosynostosis: A Novel Presentation in Association with Saethre-Chotzen Syndrome. *D.M. McDonald-McGinn[1, 2], N. Bastidas[3], A. Santani[2, 4], C. Stolle[2, 4], C. Bergfield[5], S. Bartlett[2, 3], E.H. Zackai[1, 2].* 1) Div Human Gen, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA; 3) Div Plastic & Reconstructive Surgery, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Dept Path, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Div Plastic Surgery, Seattle Children's Hospital Seattle, WA.

The combination of metopic, bicoronal and sagittal sutural synostosis produces a severe turribrachycephaly, hypotelorism, bitemporal and biparietal constriction which we have coined 'Peace Sign Craniosynostosis' given the resemblance to the iconic symbol. Saethre-Chotzen syndrome (SCS) is an autosomal dominant condition typically characterized by bicoronal craniosynostosis, ptosis, downslanting palpebral fissures, prominent ear crus, hearing loss, syndactyly, broad laterally deviated halluces, and +/- intellectual disability. Mutations, deletions or duplications of the TWIST1 gene, a basic helix-loop-helix transcription factor located on chromosome 7p21.2, are thought to be responsible for osteodifferentiation and thus premature suture fusion and limb anomalies associated with SCS. To date, over 100 abnormalities involving the TWIST1 gene have been causally related to SCS and although insertions, deletions, nonsense and missense mutations have been described, no genotype-phenotype correlation has been established, suggesting that sequence alterations lead to a loss of functional TWIST1 protein irrespective of the mutation type. Precise penetrance data is not available, however wide phenotypic variability and incomplete penetrance are well described. Here we report the novel association of Peace Sign Craniosynostosis in five patients with SCS. A retrospective review of all patients with bicoronal, metopic and at least partial sagittal synostoses at The Children's Hospital of Philadelphia and Seattle Children's Hospital was performed. Five patients with Peace Sign Craniosynostosis and SCS were identified. 4/5 were found to have TWIST1 abnormalities including a frame shift duplication, two deletions, and a novel point mutation. Results are pending on the 5th child. In terms of management, Patient I underwent three intracranial procedures and required a VP shunt. All other patients had posterior cranial vault distraction as the initial procedure and are awaiting anterior cranial vault remodeling. A second child also required a VP shunt. Thus, synostosis of the metopic, bicoronal and sagittal sutures appears to be associated with SCS and produces a unique yet characteristic skull morphology which we are terming Peace Sign Craniosynostosis. Early long-term data suggests a high rate of reoperation and increased necessity for a VP shunt. Most notably, neither this distinctive presentation nor the etiology has been reported previously.

2756F

Characterization of three families with provisional diagnosis of autosomal dominant retinitis pigmentosa using whole exome sequencing. *L. Vazquez[1], J. Liang[2], B. Almoguera[1], X. Liu[2], P. Fernandez[3], Y. Guo[1], M. Corton[3], B. Keating[1, 4, 5], X. Xu[2], C. Ayuso[3], H. Hakonarson[1, 4, 5], J. Wang[2, 6, 7, 8].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Department of Genetics and Genomics. IIS-Fundacion Jimenez Diaz Hospital. Madrid, Spain; 4) Division of Human Genetics, The Children's Hospital of Philadelphia, 19104 Philadelphia, PA, USA; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 6) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 7) King Abdulaziz University, Jeddah, Saudi Arabia; 8) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark.

Introduction: Retinitis pigmentosa (RP) is a clinically and genetically heterogeneous disease with a prevalence of approximately 1 in 4,000. Many modes of inheritance have been described, and more than 60 gene variants have been identified to date. These loci explain ~50% of all cases, and genetic characterization and diagnosis of this condition is complex. Methods: Six families (A-F) with a provisional diagnosis of autosomal dominant RP (adRP) were included in the study. Common adRP variants were previously discarded. A total of 38 individuals (10 affected, 28 unaffected members across families) participated in whole exome sequencing (WES) using the SureSelect Human All Exon kit (Agilent Technologies) and HiSeq 2000 (Illumina). Alignment and variant-calling was performed using standard methods and custom scripts. We filtered variants: 1) in coding regions, 2) segregating in the family under an adRP model, 3) at non-synonymous, splice sitedisrupting, and frameshift sites, and 4) with a MAF<1% in 1000 Genomes Project; NHLBI GO Exome Sequencing Project, and our in-house whole-exome repository. Information on genes associated with RP was retrieved from Retinal Information Network Database and Sanger sequencing was used for variant validation. Results: In three out of the six families, we identified a previously described mutation: two missense mutations in PRPH2 (NP_000313.2, p.D186N, family A) and C1QTNF5 (NP_056460.1, p.S163R, family E), and one frameshift deletion in RPGR (NM_000328.2: c.485_486delTT, family D). Sanger sequencing confirmed the complete segregation of variants in the three families. Discussion: Notably, the only adRP locus identified was PRPH2. This gene had previously been screened in the family but the mutation was missed. In the case of C1QTNF5, p.S163R is a rare mutation responsible for late-onset retinal degeneration (LORD, MIM# 605670), which in the latest phases can mimic RP phenotype. Similarly, RPGR is associated with X-linked RP instead of adRP - an inheritance model not initially considered due to the complete expression of RP in 3 females in the family. Conclusions: WES enabled the characterization of 3 families that had been misclassified in terms of inheritance model and/or phenotype, which is critical for calculating familial risk, and providing adequate genetic counseling.

2757W

Unexplained persistently elevated liver transaminase levels - a presenting feature for Duchenne/Becker muscular dystrophy.
J.A. Wilkinson, M. Martin, B. Hay. Pediatrics - Genetics, UMass Memorial Medical Center, Worcester, MA.

Duchenne/Becker muscular dystrophy (DMD/BMD) occurs due to mutations in the DMD gene. Patients with the more severe DMD phenotype generally present with progressive muscle weakness in early childhood, causing them to be wheelchair bound by adolescence, and an associated cardiomyopathy. Patients with the milder BMD phenotype generally present later in childhood with less severe or even subclinical muscle weakness and longer preservation of ambulation. The majority of affected males will be found to have elevated serum creatinine kinase (CK) levels. Dystrophin staining on muscle biopsy is usually absent in DMD and low to normal in BMD. Due to the invasive nature of muscle biopsy as well as the possible false negative results of dystrophin staining, molecular genetic testing of the DMD gene is the preferred method confirming a suspected clinical diagnosis. We describe four families identified in our clinic as having DMD/ BMD. All four probands presented with persistently elevated liver transaminases of unknown etiology and were referred for possible metabolic/genetic liver disease. In all four patients, findings of elevated liver transaminase levels were incidental. Two patients presented with rash/fevers, one patient presented with new onset seizures, and one patient had undergone routine laboratory surveillance while on stimulant medication. Liver transaminase levels remained elevated despite resolution of the acute illness in three of the cases and discontinuation of the stimulant medication in the fourth case. Two of the patients underwent liver biopsy, which revealed increased hepatic glycogen stores. Upon presentation to the Genetics clinic, CK levels were found to be >10x normal in all four patients, prompting molecular testing of the DMD gene. Three patients were found to have in-frame deletions and one patient was identified as having a complex out-of-frame duplication. At the time of diagnosis, none of the patients had obvious symptoms of muscular dystrophy, although mild findings were noted following the diagnosis in at least two of the probands. Three of the four families underwent targeted DNA analysis and maternal inheritance was confirmed in all three. Based on our clinic's experience with these cases, we suggest that CK levels should be considered in patients with unexplained persistent elevated liver transaminase levels, and molecular genetic testing of the DMD gene be offered in those cases in which the CK is found to be elevated.

2758T

Phenotypic Evolution in Börjeson-Forssman-Lehmann Syndrome: Serial Photos, Endocrine Findings and Dermatologic Observations.
G. Graham[1,3], N. Carson[1,3], D. Picketts[2,3]. 1) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 2) Ottawa Hospital Research Institute, Ottawa, ON, Canada; 3) University of Ottawa, Ottawa, ON, Canada.

Borjeson-Forssman-Lehmann syndrome (BFLS) is an uncommon X-linked recessive condition caused by mutations in the PHF6 gene. First described in 1962, there are only 51 related publications, a paltry average of one per year. The exact number of reported individuals who have bona fide BFLS is difficult to ascertain but likely between 50 and 75. Many were described prior to the attribution of the gene in 2002, so their diagnosis is not molecularly proven. Some reports do not include photographs; some include photographs that are not convincing and some lack essential clinical details. BFLS is characterized by intellectual disability, epilepsy, central obesity, gynaecomastia, hypogonadism, distal extremity findings and facial dysmorphism, including deep-set eyes and large ears. We present the evolution of this phenotype in half-brothers with a PHF6 mutation, now 13 and 18 years of age, originally felt to have mild BFLS as a result of a splicing mutation that generates an internal deletion of 34 amino acids. Although neither has epilepsy and their intellectual outcome remains more favorable than is often described, one has developed marked facial coarsening and both have developed the obesity, gynecomastia and hypogonadism that is typical of BFLS. We present serial photographs of these brothers as a diagnostic aid. We also present their endocrine profiles, the latter being relatively neglected in the literature. Finally, we highlight what we believe to be unrecognized dermatologic features of BFLS. Both brothers developed multiple pigmented cutaneous nevi over the lateral face and neck and to a lesser extent over the torso and extremities. These are also present in their mother and maternal grandmother, who are manifesting carriers. Both brothers have also developed keloid scars; in one of the two brothers these were severe and disfiguring. We reviewed all published photos of individuals with BFLS. Where photos allow this determination to be made, we noted that the majority of patients have multiple pigmented nevi. We also note keloid scars in at least one of two published patients whose photographs included surgical incision sites. In light of what is understood about PHF6 gene function, we discuss the possibility that these dermatologic findings are a feature of this syndrome and caution against cosmetic surgery in individuals with this diagnosis.

2759F

Identification of a Missense Mutation in the MBTPS2 Gene as the Causeof X-Linked Form of Olmsted Syndrome.
A. Haghighi[1,2], C. Scott[3], D. Poon[3], R. Yaghoobi[4], N. Saleh-Gohari[5], V. Plagnol[6], D. Kelsell[3]. 1) UNiversity of Oxford, Oxford, UK; 2) Department of Genetics, Harvard Medical School, Boston, MA, USA; 3) Centre for Cutaneous Research, The Blizard Institute, Barts & The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 4) Department of Dermatology, Jundishapur University of Medical Sciences, Ahvaz, Iran; 5) Genetic Department, Kerman University of Medical Sciences, Kerman, Iran; 6) Darwin Building, University College London Genetics Institute, University College London, London, UK.

Olmsted syndrome is a very rare severe keratinization disorder characterized by a combination of bilateral mutilating palmoplantar keratoderma and periorificial keratotic plaques. Other reported associated features include corneal lesions, diffused alopecia, digital constriction, diffused alopecia, nail dystrophy, high-tone loss of hearing, infections and squamous cell carcinomas. Treatment of Olmsted syndrome includes topical keratolytics, surgical removal of the keratotic palmoplantar masses followed by autograft, and use of systemic retinoids. However, the outcome is often disappointing and recurrence rate is very high. We investigated the genetic basis of Olmsted syndrome in a consanguineous extended kindred with Olmsted syndrome using whole exome sequencing. Exome sequencing revealed a mutation, which to our knowledge is previously unreported: c.1391T>C (p.F464S) in exon 11 of the X-chromosome gene membrane-bound transcription factor protease, site 2 (MBTPS2) (NM_015884.3) in both male probands. Immunohistochemical staining indicated that MBTPS2 is mainly expressed in the upper granular layer in normal skin, as previously shown; however, in OS skin, MBTPS2 was expressed throughout the epidermis. There was no apparent difference in MBTPS2 localization in the skin of a KFSD patient with the p.N508S mutation. This expands our knowledge about the genetic basis and pathology of Olmsted Syndrome and has applications in genetic counseling.

2760W

Exudative Retinopathy, Cerebral Calcifications, Duodenal Atresia, Preaxial Polydactyly, Micropenis, Microcephaly and Short Stature: A New Syndrome?
B. Isidor[1], A. David[1], Y. Crow[2]. 1) CHU Nantes, Nantes, France; 2) Manchester Academic Health Science Centre, University of Manchester, Genetic Medicine, Manchester, UK.

The association of Coats disease with intrauterine growth retardation, intracranial calcification, leukodystrophy, brain cysts, osteopenia, and gastrointestinal bleeding defines Coats plus syndrome caused by mutations in the CTC1 gene, encoding conserved telomere maintenance component Here, we report on a child with exudative retinopathy, cerebral calcifications, duodenal atresia, preaxial polydactyly, micropenis, microcephaly, and short stature, in whom no mutations in CTC1 were found. Our patient shares some features seen in other diseases associated with telomere shortening including Hoyeraal-Hreidarsson and Revesz syndromes. We therefore measured telomere length by Flow-Fish which was normal. The association of duodenal atresia and microcephaly also suggested a diagnosis of Feingold syndrome. However, direct sequencing ofMYCN was normal, and we did not detect any hemizygous deletion of the miR-1792 polycistronic miRNAcluster.To our knowledge, the phenotype we report on has not been described previously, leading us to speculate that this conditionmay represent a new syndrome.

2761T

Psychological and cognitive profile in four new patients with MOMO syndrome. *C. Passalacqua, M. Garcia, L. Dueñas.* Hospital Carlos Van Buren, Valparaiso, valparaiso, Chile.

MOMO is an acronym for macrosomia, obesity, macrocephaly and ocular abnormalities (OMIM 157980). The first patients were presented in 1993 by Moretti-Ferreira et al. with a total of seven patients published thus far. All the cases presented mental disabillity and in one case austism was described. We present four new cases with a neuropsycologycal profile, there were referal to geneticist and then to psychologist and child psychiatry. All parent patients signed a consent form to participate in the study. All patients were seen by child psychologist who performed WISC III (adapted for Chilean population) and also with child psychiatrist who did the MINI-KID test, used to assess the presence of 24 DSM-IV child and adolescent psychiatric disorders as well as the risk of suicide. The patients are three male (11, 12 and 17 years old) and a female of 13 years old. The four patients fulfill the criterias of macrocephaly, obesity and ophtalmological abnormalities; macrosomia at birth was seen in one patient (male, birth weight: 4985 g >>P99) and another had tall stature to date (male, 11 years, height 159 cm +1.81SD). The karyotype and the Prader Willi methylation test were normal in three patients, in the remain patient are still in process. The four patients had an history of developmental delay, after the psychological evaluation two had ranked as intellectual disability (IQ 51 and 58) and the others as borderline intellectual functioning (IQ 77 and 72). The subscales showed in two patiens asymmetry between verbal skills and execution functions, in three patients the lower skills were assimilate information and cubes assembly. The MINI-KID demostrate that three had psychiatric comorbilities (ansiety, phobia and behavior conduct disorder) the other patient (17 years old) had borderline intellectual functioning, without psychiatric comorbilities and he is more independent in the living day activities. Patient 3 and 4 are siblings from a healthy non consanguineous couple, Moretti-Ferreira suggested an autosomal dominant inherence, while Di Donato theorized an autosomal recessive trait. We think that are still few patients to propose an inherence trait, but the existence of brothers with healthy parents is more suggestive to an autosomal recessive inherence. This is the first neurocognitive evaluation of MOMO patients and we agree with Di Donato who proposed change the M from Macrosomia to Mental (intellectual) disability.

2762F

Pregnancy in Autosomal Recessive Polycystic Kidney Disease/Congenital Hepatic Fibrosis (ARPKD/CHF). *N. Banks, J. Bryant, R. Fischer, M. Huizing, W. Gahl, M. Gunay-Aygun.* Section on Human Biochemical Genetics, Medical Genetics Branch, NHGRI, NIH, Bethesda, MD.

ARPKD/CHF is characterized by dilated renal collecting ducts resulting in renal insufficiency and CHF often complicated by portal hypertension. It is caused by mutations in *PKHD1*, which encodes fibrocystin, a protein on the primary cilia-basal body/centriole. The consensus clinical diagnostic criteria for ARPKD/CHF require characteristic kidney and liver involvement, family history consistent with autosomal recessive inheritance, and absence of congenital anomalies. Little is known about pregnancy outcomes for patients with ARPKD/CHF. Patients with other causes of chronic kidney disease are at risk of pregnancy-related decline in renal function, preeclampsia, and impaired fetal outcomes, including higher rates of intrauterine growth retardation and pre-term delivery. Pregnancy with portal hypertension poses similar risks for the mother and baby.

In the ongoing NIH study on ARPKD/CHF and other ciliopathies (www.clinicaltrials.gov, trial NCT00068224), we evaluated 78 patients who fulfilled the clinical diagnostic criteria including 12 females over 18 years of age. Six had children; 4 carried pregnancies and delivered, 1 used assisted reproductive technology with a surrogate, and 1 adopted. Patient 1 is a 61 year old who received kidney-liver double transplantation at age 59. She was unaware of her condition until age 46 when she was diagnosed with Caroli's syndrome (CS) and subsequently with ARPKD/CHF; she had an uneventful pregnancy and delivery at age 33 via cesarean section. Patient 2 was diagnosed with ARPKD/CHF at age 6 months. She became pregnant at age 22 and was initially normotensive with a creatinine of 1.15 mg/dL and an eGFR of 63.9 mL/min/1.73m^2. Her pregnancy and vaginal delivery were clinically uncomplicated, but by delivery creatinine had increased to 1.78 mg/dL and eGFR was 36 mL/min/1.73m^2. Patient 3 was diagnosed with CHF at age 7, underwent splenic embolization at age 13, and delivered vaginally at age 23 following an uneventful pregnancy. Patient 4 was diagnosed with ARPKD/CS at age 24, 3 months after an uncomplicated pregnancy and delivery.

Patients with ARPKD/CHF are at risk of pregnancy-related worsening of kidney and liver disease. Outcomes are likely influenced by pre-pregnancy renal function and severity of portal hypertension. Appropriate preconception counseling is critical to inform patients of pregnancy risks and alternative reproductive options.

2763W

X-linked Joubert Syndrome: neuroimaging and clinical features associated with a novel mutation in OFD1. *B. Hashemi[1], H.M. Branson[2], M. Moharir[3], G. Yoon[1, 3].* 1) Division of Clinical Genetics, Hospital for sick children, Toronto, ontario, Canada; 2) Division of Pediatric Neuroradiology,Hospital for sick children, Toronto, ontario, Canada; 3) Division of Neurology,Hospital for sick children, Toronto, ontario, Canada.

Oral-facial-digital syndrome type1 (OFD1; OMIM#311200) is a rare developmental disorder characterized by craniofacial, oral, skeletal abnormalities, cystic kidneys and neurological involvement including mental retardation. The syndrome is caused by mutations in the OFD1 gene and results in male lethality in the first or second trimester. The spectrum of phenotypes associated with mutations in OFD1 continues to expand, and genotype-phenotype correlations are beginning to be recognized. The spectrum includes Simpson-Golabi-Behmel syndrome type 2 (SGB2, OMIM#300209), and X-linked Joubert syndrome (JBS type10) phenotype in addition to the classical OFD1 phenotype with early male lethality. We describe the clinical and neuroimaging findings in a 17 year old male with a clinical diagnosis of Joubert syndrome associated with a novel mutation in the OFD1 gene (JS-10). Our case confirms the reported association of OFD1 gene mutation with X-linked recessive Joubert syndrome. This family is the fifth reported with X linked Joubert syndrome, associated with a novel mutation in exon 17 of the OFD1 gene. We describe unique neuroimaging and clinical features associated with OFD1 mutations and highlight the utility of advanced neuroimaging techniques in characterizing the phenotype.

2764T

Xq11.1-11.2 deletion involving *ARHGEF9* in a girl with autism spectrum disorder. *G. Bhat[1, 2], D. LaGrave[3], A.N. Lamb[4], R. Matalon[2].* 1) Pediatrics, New York Medical College, Valhalla, NY; 2) Medical genetics, Pediatrics, University of Texas Medical Branch, Galveston, TX; 3) Department of Cytogenetics, ARUP laboratories, Salt Lake City, UT; 4) Department of Pathology, University of Utah, Salt Lake City, UT.

We report a de novo *ARHGEF9* deletion in a 6 year old girl presenting with autism spectrum disorder, ADHD, severe speech delay, developmental delays and cognitive impairment. Our patient was born at term without any known complications. She walked at 16 months and started speaking at 2 years of age; she is still unable to make two word sentences. Her hearing is normal and she receives speech therapy at school. There is no history of seizures or sensory hyperarousal and she is not dysmorphic. She was diagnosed with ASD and ADHD by a behavior specialist and is currently receiving therapy. Cytogenomic microarray, using the CytoScan HD platform, demonstrated a de novo 82 Kb deletion involving the X chromosome from Xq11.1 to Xq11.2. This deletion includes a single exon and the 5' end of each of three isoforms of the gene *ARHGEF9*, as well as the micro-inhibitory RNA 1468. X-inactivation studies on peripheral blood did not indicate the presence of significant skewed X-inactivation. *ARHGEF9* codes for collybistin, a brain-specific GEF which interacts with gephyrin, a key protein of the scaffolding system of inhibitory synapses, that is essential for postsynaptic clustering of both GABA and glycine receptors. Gephyrin has been implicated in genetic risk for autism, schizophrenia and epilepsy. Point mutations, disruptions, and deletions of *ARHGEF9* have been reported previously in five patients and are associated with developmental delay, intellectual disability, epilepsy, hyperekplexia and mild dysmorphic features. In our patient, the given clinical features are consistent with the deletion of *ARHGEF9* associated with haploinsufficiency in the presence of non-skewed X inactivation. Our patient is the first reported case of *ARHGEF9* deletion with autism spectrum disorder. This case strengthens the role of *ARHGEF9* in cognitive development and its involvement in a subset of patients with neurodevelopmental disorders. The single gene deletion affecting collybistin may lend itself to benefit by treatment with GABA analogues in these patients. In conclusion, *ARHGEF9* deletion should be considered in patients with autism spectrum disorder, further investigative studies are warranted to explore the treatment options.

2765F

A four generation family with a novel *HRAS* mutation and predominantly ectodermal findings of Costello syndrome.
D. Earl[1], S. Wallace[1, 2]. 1) Genetic Medicine, Seattle Children's Hospital, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, WA.

Costello syndrome is characterized by feeding issues and failure to thrive in infancy, short stature, developmental delay, coarse facial features, sparse curly hair, and loose and/or thickened skin. Cardiac findings include congenital heart defects, cardiac hypertrophy, and arrhythmia. Individuals have a 15% lifetime risk for malignant tumors. *HRAS* is the only gene currently known to be associated with Costello syndrome. Germline mutations leading to amino acid substitutions of the glycine residue at positions 12 or 13 are typical. Most probands have a *de novo* mutation. We present a four generation family with a novel missense *HRAS* mutation and predominantly ectodermal findings. The patient presented at 2 ½ years of age with short stature (-2 SD) and failure to thrive (-2.7 SD). Her head circumference was normal (0.3 SD). Birth parameters were normal. Video swallow study showed oral phase dysphagia, microaspiration, and gastroesophageal reflux. She had mild obstructive apnea and mild central apnea. Bone age was mildly delayed. Her IGF-1 was borderline low. IGF BP-3 and thyroid function studies were normal. She walked and used her first words at 1 year of age. At 5 years of age, she continues to meet normal developmental milestones. Physical examination revealed proportionate short stature, woolly hair, sparse and kinky eyebrows, ptosis with downward eyelashes, epicanthal folds, a wide nasal root, hypoplastic nails, and an eczematous rash involving cheeks, eyebrows and extremities. On ophthalmologic evaluation she had mild anisometropia and possible left amblyopia. Echocardiogram showed a patent foramen ovale and no evidence of hypertrophic cardiomyopathy at 4 years of age. Abdominal imaging has been normal. Platelet function assay was normal despite a report of easy bruising. The patient's mother, maternal grandfather, maternal great-grandfather, maternal uncle and his son have woolly hair, inflammatory skin findings with hyperkeratosis, nail abnormalities, and normal stature. The patient's mother was diagnosed with unilateral breast cancer at 30 years of age. There were no other reported cancers in the family. A novel heterozygous *HRAS* mutation in exon 3, c.175G>A (p.Ala59Thr), was found in the proband, her mother, and maternal grandfather. This family may represent a mild Costello phenotype or a new autosomal dominant disorder characterized by woolly hair and cutaneous findings of ulerythema ophryogenes, eczema, hyperkeratosis, and dysplastic nails.

2766W

Muenke syndrome: phenotype variability of P250R mutation in FGFR3 gene.
T. Felix[1], T.W. Kowalski[1], J. Ferrari[1,2], L.T. Souza[1], M.T.V. Sanseverino[1], M.V. Collares[1], V.S. Mattevi[2]. 1) Serviço de Gen Medica, Hosp Clinicas Porto Alegre, Porto Alegre, RS, Brazil; 2) Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brazil.

Craniosynostosis is the premature fusion of cranial sutures. It has a prevalence of 1 in 2000 newborns. The FGFR3 gene mutation P250R is responsible for Muenke syndrome. It is an autosomal dominant mutation and it has been shown that accounts for 8% of all craniosynostosis cases. The aim of this study was to analyze P250R mutation in FGFR3 gene in subjects with apparently isolated craniosynostosis. Twenty five subjects with isolated craniosynostosis were included in the study. Of the twelve cases with coronal synostosis, four cases (33%) were positive for P250R mutation. Of the 4 cases with sagittal synostosis, one (25%) was positive for the mutation. None case presenting metopic or lambdoid suture synostosis was diagnosed with the mutation. Radiological evaluation of hand and feet did not showed carpal or tarsal synostosis in all positive cases. Analysis of parent of origin of the mutation showed maternal inheritance in 2 cases and a de novo mutation in one case. Both mothers with P250R mutation also had coronal craniosynostosis. In one of these families, a brother had macrocephaly and ADHD and P250R mutation was diagnosed. The mother of the case with sagittal synostosis had normal head shape but a history of Klippel-Fiel anomaly requiring surgery in infancy. Klippel-Fiel was described previously in the literature in one single family with the P250R mutation, suggesting this anomaly should be part of the phenotype. In conclusion, our study showed the high prevalence of P250R mutation in apparently isolated craniosynostosis and also the variable clinical expressivity of this disorder.

2767T

Investigations on the molecular genetic etiology of Mayer-Rokitansky-Kuster-Hauser syndrome in two cousins using exome sequencing.
M. Herlin[1], S.P. Jonstrup[1], A.T. Højland[1], I.S. Pedersen[2], P.H. Madsen[2], A. Ernst[2], H. Okkels[2], V.Q. Le[2], H. Krarup[2], M.B. Petersen[1]. 1) Department of Clinical Genetics, Aalborg University Hospital, Aalborg, Denmark; 2) Section of Molecular Diagnostics, Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark.

Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome (OMIM #277000) is a congenital disorder characterized by aplasia of the uterus and the upper two-thirds of the vagina in females showing normal secondary sex characteristics and a 46,XX karyotype. The MRKH syndrome is often first diagnosed during adolescence where patients present with primary amenorrhoea, but they are also troubled with sexual dysfunction and infertility. MRKH syndrome is classified as type 1 (typical/isolated MRKH), type 2 (complex MRKH with renal, skeletal, or auditory malformations), and MURCS association (Müllerian duct aplasia, renal aplasia, and cervical somite dysplasia). The incidence of MRKH syndrome is estimated to 1 in 4,500 live female births. For a long time, the MRKH syndrome was thought to occur sporadically, but cases of familiar clustering led to the hypothesis of an autosomal dominant inheritance with incomplete penetrance. Several candidate genes involved in early embryonic development, such as WNT4, LHX1, WT1, TCF2 (HNF1β) and homeobox genes (HOXA7-HOXA13), have been investigated in earlier studies, but the molecular genetic etiology of the MRKH syndrome remains unknown. We present a familial case of two female cousins (aged 27 and 19) both with type 2 MRKH (with unilateral renal agenesis) that supports the idea of an autosomal dominant heredity. According to basic Mendelian genetics both their mothers, who are siblings, must be carriers of the MRKH causing mutation. The normal uterus and vagina of both mothers can be explained by incomplete penetrance. Interestingly, the two cousins' late grandfather also had unilateral renal agenesis and one of the cousins had a brother, who died shortly after a preterm birth having complete renal agenesis. It indicates that an MRKH causing mutation could be inherited in a family where males also have renal abnormalities. To reveal the etiology of the two cousins' MRKH syndrome and renal abnormalities, we use a genome-wide approach by carrying out Next-Generation Sequencing of their exomes. We also analyse the exomes of the relatives participating in the study. By using a genome-wide approach, we analyse all candidate genes earlier investigated and furthermore have the opportunity of finding new genes associated with MRKH syndrome. We present one of the first studies using a genome-wide approach and the first study using exome sequencing in the investigation of MRKH patients.

2768F

Novel patient with cutis laxa, fat pads and retinopathy due to ALDH18A1 mutations and review of the literature.
E.V. van Asbeck[1], D.F.G.J Wolthuis[1], M. Mohamed[1], T. Gardeitchik[1], E. Morava[1,2]. 1) pediatrics, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands; 2) pediatrics and human genetics, Hayward Genetics Center, Tulane University Medical School, LA, USA.

Autosomal recessive cutis laxa (ARCL) is a connective tissue disorder characterized by wrinkled, inelastic skin; frequently associated with a multisystem disease. We performed next generation sequencing in genetically unsolved patients with progeroid features, neurological and eye involvement to assess the underlying etiology. Here, we describe one of our patients; diagnosed with a novel, homozygous nonsense mutation in exon 18 of the ALDH18A1 gene, and review all reported P5CS patients. Our patient had dysmorphic, progeria-like facial features and severe, generalized sagging skin with visible veins on the abdomen and thorax. In addition, the patient had severe neurological symptoms, including seizures and spasticity. MRI showed microcephaly and hypoplastic corpus callosum. The patient also had a late closing fontanel, adducted thumbs, salt and pepper retinopathy and failure to thrive. The skin abnormalities were comparable to those described in congenital glycosylation defects (CDG). He had abnormal fat pads on the buttocks and upper legs, and the excessive skin wrinkling, late closing fontanel and seizures are also well known features of CDG. However, ATP6V0A2-CDG and COG7-CDG were ruled out on a molecular and biochemical basis. The adducted thumbs, late closing fontanel, blue sclerae and parchment-like skin could be due to a PYCR1 mutation, although the parchment-like feature of the skin is typically more severe in PYCR1 compared to our patient. PYCR1 too, was ruled out on molecular basis. However, a mutation in ALDH18A1 was found. So far 7 patients with mutations in ALDH18A1 were described. All patients had failure to thrive, five had short stature, three had inguinal hernia, five had cutis laxa on hands and feet, five had visible veins on thorax and abdomen, three had cataract, five were microcephalic and six had joint laxity. Four never learned to walk and four never to speak. Unique features included IUGR, corneal clouding, scoliosis, tremor and dystonia. Our patient is unique due to his abnormal fat distribution, retinal abnormalities, undescended testis, late closing fontanel and severe neurological presentation. Retinitis pigmentosa has never been described in ALDH18A1. In conclusion, the clinical phenotype caused by ALDH18A1 mutations is widely variable. Therefore it is difficult to decide who to screen for P5CS. We suggest genetic testing for possible ALDH18A1 mutations in patients with cutis laxa, FTT, microcephaly and joint laxicity.

2769W

Missense mutation of *MAF* in a Japanese family with congenital cataract. *Y. Narumi[1,6], S. Nishina[2], M. Tokimitsu[3], Y. Aoki[4], R. Kosaki[5], T. Kosho[1], T. Murata[3], F. Takada[6], Y. Fukushima[1].* 1) Department of Medical Genetics, Kitasato University Graduate School of Medical Sciences, Sagamihara, Kanagawa, Japan; 2) Division of Ophthalmology, National Center for Child Health and Development, Tokyo, Japan; 3) Department of Ophthalmology, Shinshu University School of Medicine, Matsumoto, Japan; 4) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 5) Division of Medical Genetics, National Center for Child Health and Development, Tokyo, Japan; 6) Department of Medical Genetics, Kitasato University Graduate School of Medical Sciences, Sagamihara, Japan.

Congenital cataract is defined as a crystalline severe blinding disease and onset within the first year of life. One-fourth patients with bilateral cataracts have inherited trait. More than 40 causative genes of cataracts are encoding structural, cytoskeletal, gap junction channel, membrane associated and cell signaling proteins. It considered difficult to establish phenotype-genotype correlations due to genetic heterogeneity. Therefore, comprehensive mutation analysis using whole exome sequencing (WES) is suitable for the diagnosis of inherited cataracts. We report a family with congenital cataract and *MAF* mutation who were identified by WES. All affected family showed congenital cataract and some members had with coloboma and microcornea. Male family members revealed autism spectrum disorder and hernia inguinalis. It was initially considered that the family had an X-linked semi-dominant inheritance pattern. WES was conducted for 3 DNA samples and *MAF* was included of segregated variants. The *MAF* mutation in the family was not reported previously. The mutations have not been identified in 200 alleles in Japanese control and other unaffected members. These data indicated that the mutation in *MAF* is likely the causal mutation in this family. *MAF* mutation had been reported of only 5 families. The family members which were previously reported had shown various type of cataract and rapidity of onset. The patients with *MAF* mutation shared not only cataracts but also anterior segment abnormalities. Our recent data strongly support that congenital/juvenile cataracts with *MAF* mutations showed microcornea and coloboma.

2770T

Clinical and Mutational Spectrum of Type II Collagenopathy Patients in Korea. *E. Ra[1], H. Park[1], S.H. Seo[1], S.I. Cho[1], T.J. Cho[2], N. Park[1], M.W. Seong[1], S.S. Park[1].* 1) Department of Laboratory Medicine, College of Medicine, Seoul National University Hospital, Seoul, Korea; 2) Division of Pediatric Orthopaedics, Seoul National University Children's Hospital, Seoul, Korea.

Introduction: Type II collagen is a helical molecule which is found in cartilage and the vitreous humor of the eye. Due to its wide distribution, type II collagenopathies show a spectrum of significantly variable conditions. Here we present the phenotypic spectrum of Korean type II collagenopathy patients whose diagnosis was confirmed by the mutation analysis of *COL2A1* gene. Methods: Twenty Korean patients who were suspected of type II collagenopathy were screened for the *COL2A1* gene. DNA was extracted from peripheral blood leukocytes, and *COL2A1* gene mutations were screened via direct sequencing of all 54 exons and multiplex ligation-dependent probe amplification (MLPA). Results: Heterozygous *COL2A*1 mutations were detected in 15 of 20 patients. Among 15 detected mutations, 4 were known (c.823C>T, c.1510G>A, c.2659C>T, c.3121G>A) and 11 were novel (c.619G>A, c.835G>C, c.1519G>T, c.1537G>T, c.1574G>A, c.1826G>T, c.1916G>A, c.2267G>A, c.2833G>A, c.3265G>A, c.3490G>A). Two of them were nonsense mutations and 13 were missense mutations, substituting glycine to other amino acid in triple helix domain. Variable skeletal phenotypes were observed; Czech dysplasia, spondyloepiphyseal dysplasia, spondyloepimetaphyseal dysplasia Strudwick type, Legg-Calve-Perthes disease, and Stickler syndrome. Spondyloepiphyseal dysplasia was diagnosed in 9 cases, which was the most common finding in this study. In 2 cases, diagnoses were not clearly categorized. Interestingly, a nonsense mutation which had been previously reported in a case of Stickler syndrome was found in a patient with Legg-Calve-Perthes disease. In addition, five cases with no mutation of *COL2A1* include one patient with Legg-Calve-Perthes disease, another with family history of early onset osteoarthritis and three with suggestive findings of spondyloepiphyseal dysplasia and brachydactyly. Conclusion: Variable phenotype may occur even within the individuals sharing a same mutation in *COL2A1* gene. This may be due to the effect of other genetic factors related in the developmental process. Confirming the diagnosis of a type II collagenopathy by mutation analysis is important in predicting the possible complications of the condition as well as providing an accurate genetic counseling for other family members.

2771F

Clinical spectrum of eye malformations in 4 new patients with Mowat Wilson syndrome. *A. BOURCHANY[1], I. GIURGEA[2], J. THEVENON[3], A. GOLDENBERG[4], G. MORIN[5], D. BREMOND-GIGNAC[6], P-O. LAFONTAINE[1], D. THOUVENIN[8], J. MASSY[9], A. MASUREL-PAULET[3], C. THAUVIN-ROBINET[3], S. EL CHEHADEH[3], S. LYONNET[7], L. FAIVRE[3].* 1) CHU Hôpital d'Enfants, University of Burgundy, Dijon , Burgundy, France; 2) Department of Genetics, Hôpital Henri Mondor, University of Paris Est Creteil, Paris, France; 3) Department of Genetics and Rare Diseases and Development Abnormalitiies and Malformative Syndroms ,CHU Hôpital d'Enfants, University of Burgundy, Dijon, France; 4) Department of Clinical Genetics, Hôpital Charles Nicolle, University of Rouen, Rouen, France; 5) Department of Oncogenetics and Clinical Genetics, hôpital Nord, University of Picardie Jules Verne , Amiens, France; 6) Pediatric Ophtalmology Saint Victor Center, University of Picardie Jules Verne, Amiens, France; 7) Department of Medical Genetics , Hôpital Necker-Enfants Malades, University of Paris 5 René-Descartes, France; 8) Jean Jaurès Ophtalmogic Center, Toulouse, France; 9) Hôpital Charles Nicolle, University of Rouen, Rouen, France.

Mowat-Wilson syndrome (MWS) is a multiple congenital anomaly syndrome characterized by a distinct facial phenotype, intellectual deficiency, Hirschsprung disease and multiple malformations caused by mutations or deletions in the zinc finger E-box-binding homeobox2 gene (ZEB2). To date, nearly 200 cases have been reported in the literature. ZEB2 encodes for Smad-interacting protein 1, a transcriptional co-repressor involved in the transforming growth factor β (TGFβ) signaling pathway and is strongly transcribed at an early stage in the developing central and peripheral nervous system in mice. Eye abnormalities have been rarely described in patients with this syndrome (4% in literature). Herein, we describe 4 patients (2 males and 2 females) with MWS and eye malformations, with a mean age of 7 years. Eye malformations included microphtalmy, iris/retinal colobomas, atrophy or absence of optic nerve, absence of optic disc, total hyphema, and deep refraction troubles, sometimes with severe visual consequences. Malformations were asymmetric in all cases and often unilateral. This spectrum of manifestations, affecting all eye segments, was superimposable with the 8 previous cases mentioned in the literature (iris/chorioretinal/optic disc coloboma, optic nerve atrophy, retinal epithelium atrophy, cataract, korectopia), as well as expression data of ZEB2 in human embryo. With the help of the literature data, we questioned whether or not the presence of eye manifestations could be due to specific type or locations of mutations. We conclude that the presence of eye malformations, although a rare feature in MWS, should not rule out the diagnosis.

2772W

A case of Autosomal Dominant Cutis Laxa within a familial context of cansanguinity. *M.B. DUZ[1], A. GEZDIRICI[1], E. KOPARIR[1], E. YOSUNKAYA[1], H. ULUCAN[1], M. SEVEN[1], A. DE PAEPE[3], P.J. COUCKE[3], B. CALLEWEART[3], M. OZEN[1,2].* 1) Istanbul University Cerrahpasa Medical Faculty, Istanbul, Turkey; 2) Department of Pathology & Immunology Baylor College of Medicine, Houston, TX, 77030, USA; 3) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.

Autosomal Dominant Cutis Laxa is a rare connective tissue disorder characterized by loose, redundant skin folds that may be apparent form birth or appear later in life. Most severely affected areas are the neck, axillar regions, trunk, and groin. Typically, patients presents with characteristic facial features including a premature aged appearance, long philtrum, high forehead, large ears, and a beaked nose. Cardiovascular and pulmonary complications include bicuspid aortic valves, aortic root dilatation, and emphysema. Sporadically, these complications have been documented to cause premature death. Most patients harbor frameshift mutations in the last 5 exons of the ELN gene (ADCL1, OMIM#123700), while one patient was described to have a tandem duplication in the FBLN5 gene (ADCL2, OMIM#614434). We report on a two year-old female patient with severe generalized cutis laxa also affecting the limbs. The skin was mildly hyperexensible. She further presented with sagging checks, large ears, a pectus excavatum, mild thoracolumbar scoliosis, a right inguinal hernia and mild joint hypermobility. Motor development was normal but she had a mild speech delay. She is the first child of consanguineous parents (first cousins). Despite the consanguinity, the clinical phenotype, ADCL seemed more likely and was confirmed by ELN mutation analysis showing a previously undescribed c.2156del (p.Gly719Glufs*36) mutation in exon 30 of the ELN gene. In conclusion, this novel case of ADCL emphasizes that careful clinical assessment remains essential in the evaluation of a patient with cutis laxa, and that consanguinity does not preclude the dominant form in this rare pathology.

2773T

Weaver Syndrome: Variable expression and natural history in a three generation family with documented EZH2 mutation. *H. Hoyme[1], A.R. Mroch[1], P.L. Crotwell[1], Y.E. Hsia[2], L. Hasegawa[2], S. Lee[2], L.H. Seaver[2].* 1) Sanford Research/Office of Academic Affairs, Sanford Health, Sanford School of Medicine of the University of South Dakota, Sioux Falls, SD; 2) Kapiolani Medical Specialists, University of Hawaii John A. Burns School of Medicine, Honolulu, HI.

Weaver syndrome is an autosomal dominant condition characterized by tall stature, macrocephaly, round face in infancy, broad forehead, hypertelorism, large ears and early developmental delay. Recently, mutations in the EZH2 gene have been identified as the cause of Weaver syndrome. The purpose of this report is to provide detailed information on the natural history and intrafamilial phenotypic variability in six individuals in a 3 generation family with Weaver syndrome, including the oldest reported patient, a 53 year old female. Our proband, a 10 year old girl, was evaluated for overgrowth at 10 months with a family history of Sotos syndrome. By age 2.5 years, her features were noted to be more typical of Weaver syndrome. Her mother and sister were evaluated at ages 26 and 2 years, respectively, for overgrowth. The mother is 188 cm tall with an OFC of 59 cm. As an adult, the proband's sister is 183 cm tall and has 4 children, 3 with overgrowth and features suggestive of either Weaver or Sotos syndrome. The two eldest had normal NSD1 testing. EZH2 testing on the two youngest children (both clinically affected) revealed a previously identified pathogenic mutation, c.2050C>T, p.R684C. The mutation was confirmed in our proband. Mutation analysis on the two older children, one clinically affected and one clinically normal, is pending. Term birth weights ranged 3.1-5.2 kg. All had excessive linear growth and macrocephaly during infancy, speech delay and articulation difficulties with milder motor delay. The adults graduated from high school. Medical complications include scoliosis, duplicated ureter, avascular necrosis requiring early hip replacement, pulmonary embolism, tonsillectomy for airway obstruction, cholecystectomy, umbilical hernia repair and neonatal tachyarrhythmia. Behavioral and psychiatric concerns were common. None have had malignancy. Several of the children demonstrated a phenotype more compatible with Sotos syndrome in infancy, progressing to a Weaver syndrome phenotype as they aged. Data from this family further enlarge the phenotype of the Weaver syndrome and highlight the difficulty in differentiating Weaver from Sotos syndrome in isolated cases. The identification of EZH2 mutation as causative for Weaver syndrome will help differentiate the two disorders and allow for more accurate phenotypic characterization of both.

2774F

Xeroderma pigmentosum complementation group B / Trichothiodystrophy spectrum in two siblings with ERCC3 mutations. *M. Migliavacca[1], N. Sobreira[2], S. Bragagnolo[1], M. Ramos[1], D. Valle[2], A. Perez[1].* 1) Clinical Genetics, Federal University of São Paulo, São Paulo, São Paulo, Brazil; 2) Institute of Genetic Medicine , Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Nucleotide excision repair (NER) is the main cellular repair pathway by wich a wide range of DNA lesions are eliminated from the genome.The importance of DNA repair is demonstrated by the existence of at least 15 human genetic disorders associated with defects in DNA repair and defects in approximately 35 different genes. In most cases, there is a greatly elevated cancer incidence and multi-system defects. Xeroderma pigmentosum (XP, OMIM 278700-278780), Cockayne syndrome (CS, OMIM 216400, 133540) and Trichothiodystrophy (TTD, OMIM 601675) are autosomal recessive disorders caused by defects in the repair of DNA, remarkable clinical heterogeneity exists within these diseases. The ERCC3 gene (Excision-Repair Cross-Complementing group 3 gene) is related to them and encodes a DNA helicase that acts in the NER pathway and a component of the transcription factor BTF2/TFIIH that initiates transcription of class II genes. Inactivation of the NER pathway can result in the classic phenotype of XP, while inactivation of transcription can lead to photosensitivity phenotype like XP, CS and TTD. Here we describe a sister and a brother born to a non consanguineous 26 years old mother and a 22 years old father, both with developmental delay, recurrent pneumonia, cutaneous xerosis, moderate sensorineural hearing loss, and retinitis pigmentosa. The sister had head circunference of 53 cm (-1DP), her weight was 47kg (p3-5) and height was 1,56 cm (p10), she had dry hair, sparse eyebrows, and a skin lesion in the right anterior cervical region with a descamative aspect. The brother's head circunference was 53 cm (<p3), his weight was 67 kg (p25-50) and height was 1,67 cm (p10), he had microcephaly, dry hair and ichthyosis. To date no skin tumors were diagnosed in neither one of them. WES was performed in the brother and compound heterozygous mutations were identified in exons 11 (c.1763_1764insG; p.E588fs) and 7 (c.C1004T: p.S335L) of ERCC3. Sanger sequencing showed that the mother was heterozygous for the p.S335L mutation, the father was heterozygous for the p.E588fs and the affected sister was compound heterozygous for both mutations. Our results imply the ERCC3 gene as the responsible for the phenotype in this family. Up to now only six families have been described in the literature with mutations in ERCC3. The absence of skin cancer in our patients may be explained by compensation through an alternative process and by the nature of the compound heterozygous mutation.

2775W

Craniofacial phenotypes in cutis laxa. *Z. Urban[1], C. Lorenchick[1], T.E. Parsons[2], K. Levine[1], S. Madan-Khetarpal[3], S.M. Weinberg[2].* 1) Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 2) Department of Oral Biology, University of Pittsburgh School of Dental Medicine, Pittsburgh, PA; 3) Department of Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Cutis laxa (CL) is a rare connective tissue disorder. Anecdotal reports suggest that individuals with CL may have distinctive craniofacial features, but no quantitative analysis of the craniofacial phenotype has been carried out on this population. Our goal was to apply objective methods to this problem. 23 subjects with cutis laxa were included in this study. Participants underwent physical examination and the medical history and results of the examination were collected using a questionnaire. 3-dimensional (3D) images of the head and face were acquired using a 3DMD portable stereophotogrammetry system. 24 facial landmarks were identified on each subject's 3D facial image, and a geometric morphometric analysis was performed to quantify facial shape. A subset of 6 CL subjects were compared to 18 sex and age-matched controls by principal components analysis. The first principal component captured 39% of the total variation and separated CL subjects from controls. CL patients had significant narrowing and elongation of the face as well as mandibular hypoplasia. Our studies define and quantify craniofacial traits in CL and implicate CL genes as regulators of craniofacial development.

2776T

Expanding the RAD21 mutational spectrum: report of the first intragenic deletion and frameshift mutation in two patients with a mild form of Cornelia de Lange Syndrome. *A. Minor[1], M. Shinawi[2], J.S. Hogue[3], D.R. Hamlin[3], C. Tan[1], K. Donato[1], L. Wysinger[1], S. Botes[1], S. Das[1], D. del Gaudio[1].* 1) Human Genetics, University Chicago, Chicago, IL; 2) Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, Missouri; 3) Department of Pediatrics, San Antonio Military Medical Center, Fort Sam Houston, Texas.

Cornelia de Lange syndrome (CdLS) is a developmental disorder characterized by limb reduction defects, characteristic facial features and impaired cognitive development. In the majority of cases mutations in NIPBL can be identified. However, additional genes have been implicated in the disorder. Recently, missense mutations and whole gene deletions in RAD21 have been identified in children with growth retardation, minor skeletal anomalies, and facial features that overlap findings in individuals with CdLS. We expand the RAD21 mutational spectrum by reporting the first intragenic deletion and the first frame shift mutation identified in two patients presenting with a mild form of CdLS. Patient 1 is a 2 ½ year old boy with developmental delay, hypospadias, inguinal hernia and some dysmorphic features that included mild synophrys, coarse facial features, 5th finger clinodactyly and overlapping of toes two over three. Patient 2 is a 12 year old boy who was diagnosed with CdLS at around a year of life on the basis of developmental delays, characteristic facial features, hirsuitism, and hand and feet anomalies. In patient 1, an intragenic deletion involving RAD21 exon 13 was identified by exon-targeted array comparative genomic hybridization. Breakpoint junction sequence analysis identified a five base pair region of microhomology around the deletion breakpoints, suggesting that the deletion arose through a microhomology-mediated repair mechanism. In patient 2, sequence analysis of the RAD21 coding region revealed a two base-pair duplication in exon 6, c.592_593dup. This frameshift duplication causes premature stop codon at amino acid position 203. mRNA sequences that contain a premature stop codon may be targeted for nonsense mediated decay. The single exonic deletion and the frame shift mutation of RAD21 identified in our patients were associated with a mild presentation of CdLS. As supported by previous literature reports, our data confirm that RAD21 mutation analysis is indicated in patients presenting with milder phenotypes of CdLS.

2777F

Two Novel Mutations in a Patient with Rhizomelic Chondrodysplasia Punctate Type 1. *H. Onay[1], M. Saka Guvenc[1], T. Atik[2], A. Aykut[1], O. Cogulu[1,2], F. Ozkinay[1,2].* 1) Medical Genetics, Ege University Faculty of Medicine, Izmir, Turkey; 2) Pedsiatrics, Ege University Faculty of Medicine, Izmir, Turkey.

Rhizomelic chondrodysplasia punctate (RCDP) is an autosomal recessive peroxisome biogenesis disorder characterized by prominent skeletal manifestations, such as proximal shortening of the long bones, multiple punctate epiphyseal calcifications and multiple joint contractures. Additionally, these patients have congenital cataracts, severe mental deficiency and postnatal growth retardation. Dismorphic features such as hypertelorism, midface hypoplasia, small nose and full cheeks are seen in these patients. Three clinically indistinguishable forms of RCDP have been defined up to now and mutations in three different genes are responsible for these specific forms. PEX7 gene is responsible for the most common form of the disease, Type 1 RCDP. Only a few patients survive beyond first decade and death usually results from generalized neurologic impairment and respiratory complications. PEX7 gene, which comprises 10 exons is located on 6q21-q22. Due to mutations of this gene, transportation of PTS2 containing proteins including peroxisomal 3-ketoacyl-CoA thiolase and PTS2-EGFP protein to the surface of the peroxisomes is impaired. Herein we present a case with dismorphic features such as depressed nasal bridge, short philtrum, high arched palate, anteverted nares; shortened femur, tibia, humerus and fibula which were diagnosed radiologically, multiple punctate epiphyseal calcifications and congenital cataract. There is consanguinity between parents. Mutation screening in the PEX7 gene revealed two new mutations that were never identified before. p.D308Y mutation which is located in exon 10 was inhereted from mother and IVS6 -2 A>G was inhereted from father. In silico analysis indicated that these two mutations were harmful. We present this case to demonstrate two novel mutations detected in PEX7 gene and to remind the importance of having two different mutations in consanguineous families.

2778W

Hereditary multiple osteochondromas. Molecular characterization of three Cypriot families and report of two novel *EXT1* gene deletions. *G.A. Tanteles[1], V. Neocleous[2], C. Shammas[2], E. Ellina[1], L.A. Phylactou[2], C. Sismani[3], V. Anastasiadou-Christophidou[1].* 1) Clinical Genetics, The Cyprus Institute of Neurology and Genetics, Nicosia, Nicosia, Cyprus, Cyprus; 2) Department of Molecular Genetics, Function and Therapy. The Cyprus Institute of Neurology & Genetics, 1683, Nicosia, Cyprus; 3) Department of Cytogenetics and Genomics, The Cyprus Institute of Neurology & Genetics, 1683, Nicosia, Cyprus.

The disorder hereditary multiple osteochondromas (HMO) - previously called hereditary multiple exostoses (HME) - is characterized by the development of multiple osteochondromas (benign cartilaginous bony tumors which grow outward from the metaphyses of long bones) which can lead to a reduction in skeletal growth, bony deformities, restricted joint motion, short stature, premature osteoarthrosis, and compression of peripheral nerves. The risk for malignant transformation (typically osteochondrosarcoma) increases with age, although the lifetime risk is low (~1%). HMO is inherited in an autosomal dominant manner and is caused by mutations in either the *EXT1* or the *EXT2* genes, which account for approximately 60-70% and 20-30% of cases, respectively. This study presents the molecular characterization of three Cypriot HMO patients and their families who underwent genetic analysis of the *EXT1* and the *EXT2* genes. Mutation analysis was performed by direct sequencing of the entire coding regions of both genes. In an attempt to identify deletions or duplications, MLPA and array-CGH analyses were also used. Two novel *EXT1* gene deletions and the previously known *EXT1* p.R701stop mutation were respectively identified in three probands and members of their families. The two novel *EXT1* large deletions ranged from 4.64kb to 0.16Mb. The breakpoints of the first deletion were located within the coding region of exon 1 of the *EXT1* gene while the second deletion removed exons 2 to 11 of the same gene. These observations provide additional evidence of the variability in phenotypic expression and mutational spectrum which is characteristic of HMO, and represent the first genetic analysis of Cypriot HMO patients. Hopefully this analysis can help towards effective diagnosis, improved genetic counseling and be used as a future potential therapeutic platform for patients affected with HMO in Cyprus.

2779T

First molecular study of Kindler syndrome in three Iranian families: novel and recurrent mutations in the FERMT1 gene. *H. Vahidnezhad[1,2], L. Youssefian[2], A. Yazdanfar[3], AM. Kajbafzadeh[4], F. Agha-Hosseini[5], M. Tabrizi[2].* 1) Biotechnology Research Center, Department of Molecular Medicine, Pasteur Institute of Iran, Tehran, Iran; 2) Medical Genetics, Tehran University of Medical Sciences, Tehran, Tehran, Iran; 3) Department of Dermatology, Hamedan University of Medical Sciences, Hamedan, Iran; 4) Pediatric Urology Research Center, Pediatric Center of Excellence, Tehran University of Medical Sciences, Tehran, Iran; 5) Department of Oral Medicine, Dental Research Center, Dentistry School, Tehran University of Medical Sciences, Tehran, Iran.

Bullous acrokeratotic poikiloderma or Kindler syndrome (OMIM 173650) is a rare autosomal recessive genodermatosis characterized by dysphagia, extensive diffuse poikiloderma, minimal trauma-induced blisters, acrokeratosis and cutaneous atrophy. Kindler syndrome is caused by various mutations in the FERMT1 gene encoding kindlin-1, a component of focal adhesions in keratinocytes in which less than 60 mutations have been reported so far. In contrast to most other countries, the burden of this disease seems to be higher in Iran due to consanguineous marriages. For the first time, we conducted a molecular analysis of three separate families from different ethnic populations in Iran toward better genetic characterization and better management of this unique population. After obtaining written informed consent from families, extensive family history was documented and DNA was extracted from patient blood samples. PCR reaction was set up for 15 exons and nearby intronic regions of the FERMT1 gene. PCR products from eight patients in three separate families were bidirectionally sequenced and analyzed. Two affected brothers from central Iran harbored the c.328C>T (p.Arg110X) homozygous mutations in exon 3. Carrier testing was also performed for four other siblings in this family, and in genetic counseling sessions the available reproductive options and recurrent risks for the heterozygotes were discussed. From Azerbaijan, two affected brothers presented with the least severe clinical manifestations with only cutaneous and no extra-cutaneous involvement. The two Azerbaijani brothers seem to be homozygotes for a probable missense mutation c.1577G>A (p.Arg526Lys) after sequencing and analysis of the entire length of the gene. This variant, in heterozygous form, is classified as a polymorphism and non-pathogenic. This is the first report of homozygosity and pathogenicity of this variant. A third family from western Iran with four severely affected siblings were found to harbor a novel mutation c.1383C>A (p.Tyr461X). In conclusion, we provide evidence for novel mutations in the FERMT1 gene from Iran and expand the existing spectrum of mutations in humans. This data also allows us to provide genetic testing to affected individuals and propose the possibility of using PCR and sequencing as a cheap and quick screen of the heterozygotes in the population. Genetic counseling, education, and follow-up could reduce the burden of this disease in Iran.

2780F

Genotype-Phenotype Relationships in Freeman-Sheldon Syndrome. *A.E. Beck[1,2], M.J. McMillin[1], H.I.S. Gildersleeve[1], K.M.B. Shively[1], M.J. Bamshad[1,2,3].* 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington School of Medicine, Seattle, WA; 2) Seattle Children's Hospital, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA.

The distal arthrogryposis (DA) syndromes are a group of ~10 disorders characterized by multiple congenital contractures such as camptodactyly and clubfoot. DA type 2A (i.e., Freeman-Sheldon syndrome) is the most severe of the DA syndromes as it is also manifest by contractures of the facial muscles. DA2A is caused by autosomal-dominant mutations in *MYH3*, the gene that encodes embryonic myosin heavy chain, a protein of the contractile apparatus of skeletal muscle. In our cohort of 46 families with DA2A, we found disease-causing mutations in *MYH3* in 43 kindreds (93%). We then characterized the genotype-phenotype relationships in these mutation-positive families with DA2A. Three common *MYH3* mutations accounted for 39/43 (91%) of the DA2A mutation-positive kindreds including: 8/43 (19%) with p.T178I, 11/43 (26%) with p.R672C, and 20/43 (47%) with p.R672H. Using a quantitative phenotype score that we developed based on physical findings and natural history, we found individuals with p.T178I to have a phenotype score that was more severe than individuals with a p.R672C or p.R672H mutation. Specifically, the p.T178I mutation was associated with more severe facial contractures and congenital scoliosis. The p.R672C mutation was associated with less severe foot and facial contracture scores. Of the total 46 cases screened, 11 were familial cases and 35 were sporadic cases. *MYH3* mutations were found in all of the familial cases and in 32/35 (91%) of the sporadic cases. Severe phenotypes were more common among sporadic cases than familial cases. For example, the mutation with the most severe phenotype scores, p.T178I, was observed in only sporadic cases (8/8). The mutation associated with the least severe phenotype scores, p.R672C, was observed in only 3/11 (27%) sporadic cases but 8/11 (73%) familial cases. Collectively, our analysis suggests that *MYH3* genotype is predictive of phenotype and, to a more limited extent, natural history of individuals with DA2A.

2781W

A 24bp deletion in ELN causing a Marfan-like phenotype. *J. Hoyer, C. Kraus, A. Reis.* Institute of Human Genetics, University Erlangen-Nuremberg, Erlangen, Germany.

We report on a 27 year old male presenting with a Marfan-like phenotype. The patient had a mild myopia, a high palate and broad uvula, skin striae, a plain flat foot, a positive wrist sign as well as long fingers and toes and an increased arm span-to-height ratio. Moreover a threefold patellar luxation was reported. Marfan syndrome was suspected three years ago when funnel chest surgery was performed. A cardiological examination was consecutively recommended but revealed no abnormalities. The recently revised diagnostic criteria for Marfan syndrome integrate information from multiple sources including personal medical history and physical examination. Those features are weighted and grouped to derive a 'systemic score'. In the absence of a family history of Marfan syndrome, a systemic score of ≥7 in combination with an aortic root enlargement is sufficient for the diagnosis. Our patient achieved a systemic score of 7 but did not show an aortic dilatation. We used an individual, Multiplex-PCR based Ion AmpliSeq Kit capturing 98,44 per cent of coding regions to analyze the genes FBN1, ACTA2, ELN, CBS, FBN2, MYH11, COL3A1, SLCA10, SMAD3, TGFBR1, TGFBR2 and TGFB2. Surprisingly we identified a small heterozygous in frame deletion of 24 bp (c.1178_1201del24bp) in exon 20 of the ELN gene leading to a loss of 8 amino acids (p.Gly393_Ala401delinsAla). This result was confirmed by Sanger sequencing. We suppose that expression of a mutated protein leads to the disruption of elastic fibre architecture. Up to now mutations in Elastin are only known for autosomal dominant Cutis laxa (ADCL) and as a cause for aortic aneurysms but are not described as causative for other Marfan phenotype features. For further evaluation physical examination and molecular analysis in the patient's parents is planned.

2782T

Multiple pterygium syndrome, Escobar variant, in a patient with a congenital diaphragmatic hernia and prenatally-diagnosed arthrogryposis. *A.L. Sutton, M.D. Descartes.* Department of Genetics, University of Alabama at Birmingham, Birmingham , AL.

Introduction: The multiple pterygium syndromes (MPS) are characterized by arthrogryposis, webbing of the skin (pterygia), dysmorphic facial features, and other congenital anomalies. There are two forms of MPS: the lethal and non-lethal (Escobar) variant. MPS displays significant locus heterogeneity. Some cases are caused by mutations in the gene encoding the gamma-subunit of the acethycholine receptor (*CHRNG*), which is only expressed during fetal life. Case Report: We describe a patient with prenatally-diagnosed arthrogryposis and a congenital diaphragmatic hernia with MPS, Escobar variant, with heterozygous mutations in *CHRNG*. The mother, a 40 year old G3P2, was referred to our fetal diagnosis center at 25 weeks for multiple congenital anomalies. The detailed ultrasound showed multiple joint contractures, skin edema, and polyhydramnios. Amniocentesis returned a normal female karyotype. She was delivered via cesarean section after a failed induction of labor at 39 weeks. After birth, she was found to have a congenital diagphragmatic hernia, which was repaired on day of life 12. Genetic testing revealed two truncating mutations in *CHRNG* (c.459dupA and c.753_754delCT). After serial casting and physical therapy, she started standing at 16 months of age. Discussion: Diaphragmatic hernias are one of the findings in lethal multiple pterygium syndrome. Although diaphragmatic eventrations have been found in patients with MPS, Escobar variant, diaphragmatic hernias have not previously been described in this syndrome. Additionally, homozygosity for these two truncating mutations have been associated with the lethal form of the disorder, while compound heterozygotes for these two mutations with the Escobar phenotype similar to this case have also been described. This case highlights the phenotypic and genetic heterogeneity in the MPS. MPS should be considered in the differential diagnosis in patients with prenatally-diagnosed arthryogryposis and congenital diagphragmatic hernias.

2783F

Congenital Dyserythropoieic anemia type 1 (CDA 1) presenting with blueberry muffin rash and profound anemia. *L. Turner[1], L. Goodyear[2], L. Bowes[2], S. Fernandez[1], A. Hogg[2].* 1) Medical Genetics, Memorial University of Newfoundland, St. John's, NL, Canada; 2) Pediatric Hematology/Oncology, Eastern Health, Janeway Child/Rehab Centre, St. John's, NL, Canada.

The term blueberry muffin rash was first used to describe a non-blanching purpuric rash that can be seen in neonates with congenital infections. Biopsy findings of the lesions seen reveal extramedullary hematopoiesis. The differential diagnosis of blueberry muffin rash includes infection, neoplasms, and hematologic dyscrasias. More recently, Aicardi-Goutieres syndrome and Mucolipidosis 2 have been added to the list of disorders that can present in the neonatal period with blueberry muffin rash. We present a 37 week gestation singleton male born with a blueberry muffin rash and hepatomegaly. He was depressed at birth requiring intubation and ventilation. Initial CBC revealed a hemoglobin of 56 g/L. Physical examination revealed several digit abnormalities including small finger and toenails, 4-5 toe soft tissue syndactyly and a small 3rd toe on the right foot. The infant was also noted to have a short neck with redundant nuchal skin, bilateral epicanthic folds and mild micrognathia. He had significant hepatosplenomegaly. Initial investigations for congenital infection were negative. Karyotype and microarray were normal. Ongoing investigations including repeat blood smear and bone marrow examination showed abnormalities but were not diagnostic of any one condition. Given the hematologic picture and the digit abnormalities CDA 1 was considered. Molecular analysis revealed compound heterozygous mutations in the CDAN 1 gene. Our patient was started on interferon and is doing well. He is the first reported patient with CDA 1 presenting with blueberry muffin rash. CDA 1 should be added to the differential diagnosis of infants presenting with blueberry muffin rash.

2784W

Whole exome sequencing uncovers mutations in MYH9 associated with expanded phenotype spectrum. *C. Kao[1], J. Liang[2], A. Alodaib[3], Y. Guo[1], L. Tian[1], X. Liu[2], L. Dai[2], B. Keating[1,4,5], M. Menezes[3], W. Gold[3], M. Wilson[6], L. Ades[6], J. Zhang[2], A. Kakakios[7], J. Wang[2,8,9,10], H. Hakonarson[1,4,5], J. Teo[11].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia; Philadelphia, PA, USA; 2) BGI-Shenzhen; Shenzhen 518083, China; 3) Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead; Westmead, New South Wales, Australia; 4) Division of Human Genetics, The Children' s Hospital of Philadelphia; Philadelphia, PA, USA; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania; Philadelphia, PA, USA; 6) Clinical Genetics Dept, Children's Hospital at Westmead; Westmead, New South Wales, Australia; 7) Immunology Dept, Children's Hospital at Westmead; Westmead, New South Wales, Australia; 8) Department of Biology, University of Copenhagen; Copenhagen, Denmark; 9) King Abdulaziz University; Jeddah, Saudi Arabia; 10) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; Copenhagen, Denmark; 11) Haematology Dept, Children's Hospital at Westmead; Westmead, New South Wales, Australia.

Dominantly inherited deleterious mutations in the human gene MYH9 (myosin, heavy chain 9, non-muscle) have been linked to a range of rare hereditary disorders with the overlapping platelet manifestation of macrothrombocytopenia (MTCP), and they can also cause associated symptoms of hearing impairment, nephropathy, leukocyte inclusion and/or cataract. In the current study, we used whole exome capture followed by next generation sequencing to search for causative mutations in two families with MTCP The proband in the first family presented with MTCP, craniosynostosis and hearing loss and was found to have a MYH9 mutation (amino acid change S96L); the proband in the other family presented with MTCP, severe mental retardation, epilepsy, facial malformations and renal calculi and was found to have another MYH9 mutation (A95T). While these two mutations have been reported before, the presentation of developmental, intellectual and psychiatric abnormalities is atypical and new to the MYH9-related diseases. We validated these mutations by traditional Sanger sequencing and the WES data excluded the possibility of other genetic mutations being causal for the syndromes. These results demonstrate that by providing fast and accurate diagnosis at the molecular level, next generation exome sequencing can help broaden the symptomatic spectrum of an established genetic disorder when the presentation is atypical.

2785T

Mutations in *FAM111B* cause Hereditary Fibrosing Poikiloderma (HFP) with tendon contracture, myopathy and pulmonary fibrosis. *S. Mercier[1,2], S. Küry[3], G. Shaboodien[4], D. Houniet[5], N. Khumalo[6], J-M. Mussini[7], E. Salort-Campana[8], D. Figarella[9], N. Bodak[10], A. Munnich[11], R. Gherardi[12], V. Cormier-Daire[11], C. Thauvin[13,14], L. Faivre[13,14], K. Pillay[15], C. Bou-Hanna[16], C. Laboisse[16], A. Hamel[17], A. Magot[18], C. Le Caignec[19], A. David[1], S. Barbarot[20], B. Keavney[5], S. Bézieau[3,16], B. Mayosi[4].* 1) CHU Nantes, Service de Génétique Médicale, Unité de Génétique Clinique, Centre de Référence Anomalies de Développement et Syndromes Malformatifs de l'interrégion Grand-Ouest, 9 quai Moncousu, 44093 Nantes CEDEX 1, France; 2) INSERM UMR1089, Atlantic Gene Therapy Institute, University of Nantes, Nantes, France; 3) CHU Nantes, Service de Génétique Médicale, Unité de Génétique Moléculaire, 9 quai Moncousu, 44093 Nantes CEDEX 1, France; 4) Cardiovascular Genetics Laboratory, Hatter Institute for Cardiovascular Research in Africa, Department of Medicine, Groote Schuur Hospital and University of Cape Town, Cape Town, South Africa; 5) Institute of Genetic Medicine, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; 6) Division of Dermatology, Department of Medicine, Groote Schuur Hospital and University of Cape Town, Cape Town, South Africa; 7) CHU Nantes, Laboratoire d'Anatomopathologie A, Nantes, France; 8) Hôpital de la Timone, Service de Neurologie, Centre de Référence des maladies Neuromusculaires et Sclérose Latérale Amyotrophique, Marseille, France; 9) Laboratoire de Neuropathologie, Faculté de Médecine, CHU Timone, Marseille, France; 10) Hôpital Necker Enfants Malades, APHP, Service de dermatologie, Paris, France; 11) Hôpital Necker-Enfants malades, AP-HP, U781, Fondation Imagine, Paris Descartes-Sorbonne Paris Cité, Service de Génétique, Paris 75015, France; 12) APHP, Service d'Histologie, INSERM U841 , CHU Mondor, Créteil, France; 13) Equipe d'accueil EA 4271 GAD "Génétique des Anomalies du Développement", IFR Santé STIC, Université de Bourgogne, Dijon, France; 14) Centre de Référence Anomalies de Développement et Syndromes Malformatifs de l'interrégion Grand-Est et Centre de Génétique, Hôpital d'Enfants, CHU, Dijon, France; 15) Division of Anatomical Pathology, Department of Clinical Laboratory Sciences, National Health Laboratory Service and University of Cape Town, Cape Town, South Africa; 16) EA Biometadys, Université de Nantes, Nantes, France; 17) Laboratoire d'Anatomie, Faculté de Médecine, Université de Nantes, 1, rue Gaston Veil, Nantes Cedex, 44035, France; 18) CHU de Nantes, Laboratoire d'Explorations Fonctionnelles, Centre de Référence des Maladies Neuromusculaires Rares de l'Enfant et de l'Adulte Nantes-Angers, Nantes, F-44000, France; 19) CHU Nantes, Service de Génétique Médicale, Unité de Cytogénétique, 9 quai Moncousu, 44093 Nantes CEDEX 1, France; 20) CHU Nantes, Clinique dermatologique, Hôtel Dieu, Place Alexis Ricordeau, 44000 Nantes, France.

Hereditary Fibrosing Poikiloderma (HFP) with tendon contracture and pulmonary fibrosis has been described in a South-African family with autosomal dominant inheritance by Khumalo *et al.* in 2006. Here, we report four sporadic cases affected with this syndrome in two males and two females (ages 6-30). Key features consist of: (i) congenital poikiloderma, hypotrichosis, hypohidrosis; (ii) tendon contracture with feet varus deformation; (iii) progressive muscular weakness and (iv) progressive pulmonary fibrosis. Muscle MRI shows extensive fatty infiltration confirmed by muscle biopsy. Microscopy of the skin reveals a sclerodermiform aspect with fibrosis and alterations of the elastic network. Whole exome sequencing performed in the South African family and in the trio of one sporadic case gives rise to the identification of *FAM111B* gene (NM_198947.3), unreported to date. The involvement of this gene has been confirmed by the detection of variations in the three other cases. In total, three independent patients share exactly the same missense mutation (p.Arg627Gly) and one other patient and the South-African family carry two very close mutations respectively p.Ser628Asn and p.Tyr621Asp. All these mutations are missense variants and are predicted to localize in a loop splitting the trypsin-like cysteine/serine peptidase domain of the protein into two parts. We suggest gain of function or dominant negative mutations resulting in FAM111B enzymatic activity changes. Functional studies are ongoing to better understand the pathophysiology of this entity. *FAM111B* sequencing is also ongoing in a series of *RECQL4* negative patients as this syndrome can be misdiagnosed with Rothmund-Thomson syndrome at first sight. In conclusion, HFP with tendon contracture, myopathy and pulmonary fibrosis, a phenotypically recognisable syndrome, is caused by autosomal dominant mutations in *FAM111B* gene. These findings provide genetic evidence for a new pathogenetic pathway for multisystem fibrosis.

2786F

Identification of Human MALT1 Deficiency and Role of the NF-kappa B pathway in a Novel Autosomal Recessive Immune Deficiency and Dysregulation Disorder. *M.L. McKinnon[1,3], J. Rozmus[2,3], S. Fung[2,3], A. Hirschfeld[2,3], K. Del Bel[2,3], L. Thomas[2,3], N. Marr[2,3], S. Martin[4], C. Senger[3,5], A. Tsang[3,5], J. Prendiville[2,3], A. Junker[2,3], M. Seear[2,3], K. Schultz[2,3], L. Sly[2,3], R. Holt[4], M. Patel[1,3], J.M. Friedman[1,3], S. Turvey[2,3].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Pediatrics, British Columbia Children's Hospital, Vancouver, BC, Canada; 3) Child and Family Research Institute, University of British Columbia, Vancouver, BC, Canada; 4) BC Cancer Agency, Genome Sciences Centre, Vancouver, BC, Canada; 5) Department of Pathology, British Columbia Children's Hospital, Vancouver, BC, Canada.

The transcription factor, NF-kappa B (NF-kB) is a chief regulator of lymphocyte activation, survival and proliferation, implicated in immunodeficiency and human inflammatory, autoimmune and neoplastic disorders. Despite its diverse roles, mutations in the NF-kB pathway have not yet been linked to the Hyper IgE syndromes (HIES), a group of primary immunodeficiency disorders characterized by elevated serum IgE, staphylococcal skin infections, eczema and pulmonary infections. Mutations in STAT3 cause autosomal dominant (AD) HIES and mutations in the DOCK8 gene are described in autosomal recessive (AR) HIES. In many HIES cases, the genetic cause remains unknown. Here we describe a novel immune dysregulation phenotype with HIES-like features due to underlying mutations of the MALT1 gene, a key component of NF-kB signaling, highlighting a new role for the NF-kB pathway in human disease. Our patient, a 14 year-old girl born to consanguineous parents, presented in infancy with widespread, treatment-resistant dermatitis. followed by recurrent bacterial and viral skin infections, episodes of pneumonia, chronic lung disease, severe gastrointestinal inflammatory disease, osteoporosis with fractures, chronic granulation tissue of larynx and ears, and severe periodontal disease, along with short stature and non-specific dysmorphic facial features. Laboratory testing showed variable eosinophilia, very low B cell numbers, chronically elevated serum IgE and failure of lymphocyte proliferation upon phytohemagglutinin (PHA) stimulation. By whole exome sequencing, we identified an underlying homozygous missense mutation in the MALT1 gene encoding MALT1/paracaspase, a protein serving dual roles in the NF-kB pathway as caspase-like protease (modifying NF-kB-related substrates), and key scaffold protein of the CARMA1-BCL10-MALT1 (CBM) signalosome complex (critical for CBM-mediated NF-kB signalling). We demonstrate that the MALT1 mutation results in profound deficiency of the MALT1 protein, failure of T-cell receptor-mediated activation of NF-kB and loss of paracaspase proteolytic activity. Characterization of human MALT1 deficiency expands our understanding of the genetic causes of disorders with HIES features and the role the NF-kB pathway in human immune disease. In addition, as MALT1-inhibitors are currently being examined for the treatment of lymphoma and inflammatory conditions, our case provides unique insights to guide the ongoing development of these therapies.

2787W

Familial Occurrence of Multiple Isolated Epidermal Inclusion Cysts: evidence for X-linked inheritance? *J. Jenkins, K. Horii, H. Ardinger.* Divisions of Genetics and Dermatology, Children's Mercy Hospitals & Clinics/ University of Missouri-Kansas City School of Medicine, Kansas City, MO.

Epidermal inclusion cysts (EICs) are among the most common type of subcutaneous cysts and are frequently found on the face, trunk, neck and scalp. EICs originate from the infundibular portion of the hair follicle and are characterized by implantation of epidermal elements into the dermis. Most EICs occur as sporadic, isolated lesions and are typically benign with rare malignant transformation reported. EICs are twice as common in males, show no racial predilection, and commonly arise in adulthood. Multiple EICs have been reported in association with underlying genetic syndromes including familial adenomatous polyposis (FAP), pachyonchia congenita, and nevoid basal cell carcinoma syndrome (NBCCS). Autosomal dominant inheritance of multiple isolated EICs has also been suggested (OMIM #131600); although a specific genetic etiology remains unknown. We report a 17 year-old male with a history of multiple subcutaneous cysts involving the trunk, face, neck and buttocks beginning at age 5. The diagnosis of EICs was confirmed on pathology after surgical excision. The family history was notable for early onset of multiple EICs in the patient's mother, maternal aunts, maternal grandfather, maternal great-grandmother, and maternal-great-great grandfather. There was no family history of polyposis, cancers, nail dystrophy, craniofacial dysmorphism or other findings suggestive of FAP, pachyonchia congenita, or NBCCS. The family history could be compatible with either an autosomal dominant or an X-linked inheritance pattern of multiple isolated EICs. All daughters of affected males in this pedigree are affected, males are affected more severely than females, and lack of male-to-male transmission are all factors which could support an X-linked inheritance pattern. The recurrence risk for multiple EICs in the patient's offspring ranges from 50% for either gender in the case of autosomal dominant inheritance, up to 100% in female offspring in the case of X-linked dominant inheritance. While autosomal dominant inheritance has been previously suggested in familial cases of multiple isolated EICs, few families have actually been reported in the literature. To the best of our knowledge, this is the first report of familial multiple isolated EICs that is compatible with X-linked inheritance. The present case underscores the importance of collecting a detailed multi-generational family history for provision of the most accurate recurrence risk information.

2788T

Intragenic deletion of *NPAS3* in a child with developmental delay. *C. Armour[1,2,3], E. Baxter[2], J. McGowan-Jordan[1,3].* 1) Eastern Ontario Regional Genetics Program, Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON; 3) Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada.

NPAS3 (*neuronal PAS domain-containing protein 3*) is a basic-helix-loop helix transcription factor expressed in the developing brain and implicated in neurodevelopment as it is thought to function in neuron production and maturation. A number of reports have suggested that disruption of *NPAS3* may result in both cognitive impairment and psychiatric illness. Interestingly, this gene has also been shown to contain the largest number of human-accelerated regulatory sequences (non-coding conserved elements that are unique to humans) and thus postulated to have contributed to human brain evolution. This report describes an 8 year old female with developmental delay and an intragenic deletion of *NPAS3*. She walked at 2.5 years and began speaking at 4 years of age and has additional delays in fine and gross motor skills. She currently attends school in a modified class and overall abilities are currently estimated to be 2 to 3 years behind. Additional medical issues include: daytime intermittent enuresis, an asymptomatic spinal syrinx, obesity (BMI of 34) and an astigmatism. Physical exam reveals macrocephaly of approximately 3 to 4 standard deviations (SD) above the mean and height at approximately 2 standard deviations from the mean. She has a round face with full cheeks, mild ptosis and down slanting palpebral fissures. Family history is notable for a maternal history of learning delay requiring special education. The father has macrocephaly, approximately 3 SD above the mean, and has a historical diagnosis of a Fetal Alcohol Spectrum disorder. High density oligonucleotide array in the proband revealed a 0.111Mb intragenic deletion at 14q13.1 involving only *NPAS3*. Prader-Willi MLPA studies and *PTEN* sequencing were normal. Follow-up studies performed so far indicate that the deletion was maternally inherited, correlating with the mother's own reported learning disability. Family studies are ongoing. This case supports a role for *NPAS3* in neurodevelopment and abnormal NPAS3 function may be involved in other cases of non-specific cognitive impairment.

2789F

Dyggve-Melchior-Clausen syndrome In Three Generation. *B. Bozorgmehr, A. Kariminejad.* Dept Clinical Genetics, K-N Pathology & Genetics Ctr, Tehran, Iran.

Dyggve-Melchior-Clausen(DMC) syndrome is a rare autosomal recessive spondyloepimetaphyseal dysplasia.The main features are short trunk dwarfism,short limbs with intellectual disabilities.Radiological findings are:abnormalities of spine ,epiphyses and metaphyses.We are reporting a 15-year-old boy with afghan origin with short trunk dwarfism,short limbs,characteristic radiological findings and severe mental retardation who had three affected maternal uncles and also his mother had one affected maternal uncle.His parents were first cousin and the others, were from one little village so their pedigree showed a pseudo X-linked recessive pattern.

2790W

Novel de novo SPOCK1 Mutation in a Proband With developmental delay, microcephaly and agenesis of corpus callosum. *R. Dhamija[1], J. Graham[2], E. Thorland[1], S. Kirmani[1].* 1) Mayo clinic, Rochester, MN; 2) Cedars-Sinai Medical Center, Los Angeles, CA.

Introduction: Whole exome sequencing of an affected patient and unaffected parents has now made it possible to identify novel de novo mutations in genes possibly linked to human disease. Here we describe a patient with developmental delay and partial agenesis of corpus callosum with a novel de novo SPOCK1 variant. Mutations in SPOCK1 have not been previously described to cause a human phenotype. Case report: The proband's phenotype included severe global developmental delay with dystonia, structural brain abnormality with partial agenesis of corpus callosum, microcephaly and atrial septal defect with aberrant subclavian artery. Extensive chromosomal, genomic and metabolic studies were unrevealing. At age of 13 years, exome sequencing was performed on the patient and her parents at GeneDX, Gaithersburg, MD, and a de novo novel missense mutation was identified in SPOCK1 (coding for Testican-1) on chromosome 5q31: c.239A>T (p.D80V). This mutation affects a highly evolutionarily conserved area of the gene, replacing a polar aspartic acid with hydrophobic nonpolar valine that changes the chemical properties of the protein product, likely representing a pathogenic variant. Online prediction programs SIFT (http://sift.jcvi.org/) and mutation taster (http://doro.charite.de/MutationTaster/index.html) suggest this variant is a deleterious mutation. Previous microdeletions of 5q31 including SPOCK1 have suggested genes on 5q31 as candidates for intellectual disability. No mutations or variants in other genes potentially linked to the phenotype were identified. Conclusion: Testicans are proteoglycans belonging to the BM-40/SPARC/osteonectin family of extracellular calcium-binding proteins. Testican-1 is encoded by the SPOCK1 gene and in mouse models has been shown to be strongly expressed in the brain and modulate important steps in neurogenesis. In humans the expression is more widespread and also seen in heart, blood and cartilage. We hypothesize that because this gene function is critical in neurogenesis, mutations can potentially lead to a phenotype with developmental delay and microcephaly. Further functional studies are under way to confirm these findings, potentially making this the first reported case of human disease related to SPOCK1.

2791T

FOXP1 mutations cause intellectual disability and a recognizable phenotype. *M.F. Hunter[1,11], A.K. Le Fevre[1,2], S. Taylor[3], N.H. Malek[4], D. Horn[5], C.W. Carr[6], O.A. Abdul-Rahman[7], S. O'Donnell[1], T. Burgess[8], M. Shaw[9], J. Gecz[9], N. Bain[10], K. Fagan[10].* 1) Hunter Genetics, Waratah, NSW, Australia; 2) John Hunter Children's Hospital, Newcastle, NSW, Australia; 3) Core Interventions Occupational Therapy Services, Gosford, NSW, Australia; 4) Speech Pathologist, Lisarow NSW; 5) Institute of Medical Genetics, Charité, University of Berlin, Berlin, Germany; 6) Department of Dermatology, Emory University, Atlanta, GA, USA; 7) Department of Pediatrics, University of Mississippi Medical Center, Jackson, MS, USA; 8) VCGS Pathology, Melbourne, Australia; 9) Department of Paediatrics, The University of Adelaide, SA, Australia; 10) Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW, Australia; 11) Faculty of Health, University of Newcastle, Newcastle, Australia.

Mutations in FOXP1, located at 3p13, have been reported in patients with global developmental delay (GDD), intellectual disability (ID) and speech defects. Mutations in FOXP2, located at 7q31, are well known to cause developmental speech and language disorders, particularly developmental verbal dyspraxia. FOXP2 has been shown to work co-operatively with FOXP1 in human development. An overlap in FOXP1 and FOXP2 expression, both in songbirds and the human fetal brain have suggested that FOXP1 may also have a role in speech and language disorders. We report a male child with a 0.19Mb intragenic deletion that is predicted to result in haploinsufficiency of FOXP1. Review of our patient and others reported in the literature reveals an emerging phenotype of GDD/ID with moderate to severe speech delay where expressive speech is most severely affected. Facial features include a broad forehead, down-slanting palpebral fissures, a short nose with broad tip, relative or true macrocephaly, a frontal hair upsweep and persistent fetal finger pads. Autistic traits and other behavioral problems are likely to be associated. Congenital malformations may also be associated with FOXP1 haploinsufficiency.

2792F

CTNNB1 mutation in siblings with intellectual disability, spasticity and microcephaly. *D. Tegay[1,2].* 1) Department of Medicine, NYIT College of Osteopathic Medicine, Old Westbury, NY; 2) Division of Genetics, Nassau University Medical Center, East Meadow, NY.

De novo heterozygous loss-of-function mutations in the beta cadherin-associated protein, beta-catenin (CTNNB1), have recently been identified through whole-exome sequencing in 3 unrelated individuals sharing similar features of intellectual disability, absent or limited speech, microcephaly and spasticity (de Ligt et al., 2012). Beta-catenin is a key downstream component of the canonical Wnt signaling pathway and acts as a negative regulator of centrosome cohesion. While somatic gain-of-function activating mutations in CTNNB1 are found in a variety of tumor types and believed to impart oncogenic activity, animal models suggest germline loss-of-function mutations effect neuronal, particularly cerebellar, development and maturation. We report the first case of siblings, both presenting with variable degrees of unexplained intellectual disability, microcephaly, spasticity and non-specific dysmorphic features, in whom clinical whole-exome sequencing identified a common heterozygous CTNNB1 nonsense mutation (S681X) not detected in either parental sample, suggesting parental germline mosaicism as the likely molecular mechanism. These cases help further delineate and expand the phenotype of CTNNB1 mutation.

2793W

Molecular genetic characterization of *INSR* in a family with Rabson-Mendenhall syndrome. *P. Kantheti[1], S. Agadi[2], Y. Gowda[1].* 1) Centre for Human Genetics, Biotech Park, Electronic City Phase 1, Bangalore, Karnataka, India; 2) Karnataka Institute of Medical Sciences, Hubli, Karnataka.

Rabson-Mendenhall syndrome (RMS) is a rare autosomal recessive disorder caused by mutations in the insulin receptor gene(*INSR*). Sequence analysis of *INSR* in a consanguineous family with all the offspring affected revealed that the two surviving female index cases have a homozygous nucleotide substitution c.768 C > T at the 220th codon of *INSR* predicted to change Proline to Leucine in the cystiene-rich domain of the alpha subunit. The mutation does not appear to affect the transcription of the gene or the binding of insulin to the receptor. Haplotypes inferred from single nucleotide polymorphisms indicated that both siblings inherited identical haplotypes linked to the substitution from the two parents. It is also interesting to note that the affected siblings have inherited homozygous His1085T/T polymorphism from the parents who are heterozygous (His 1085 T/C). This polymorphism may possibly lead to the disruption of an E-box motif (CA**CG**TG) and/or to the disruption of a potential epigenetic signature. This may have relevance to the insulin resistance commonly seen in RMS patients.

2794T

Microarray analysis of a de novo microdeletion involving the Van der Woude Syndrome locus. *E.C. Tan[1], E.C.P. Lim[1], S.T. Lee[2].* 1) KK Research Centre, KK Women's & Children's Hospital, Singapore; 2) Department of Plastic, Reconstructive & Aesthetic Surgery, Singapore General Hospital, Singapore.

Van der Woude syndrome is the most common among syndromes which include cleft lip and/or cleft palate as one of the presentations. This autosomal dominant disorder is usually caused by mutations in the interferon regulatory factor 6 (IRF6) gene, with a few cases of small deletions involving coding sequences in the gene. For larger deletions, there were only four reports of microdeletions in this chromosomal region, one was microscopic while the other three involved microsatellite markers outside the IRF6 gene region. We describe the detection of an interstitial deletion of approximately 2.3 Mb within the 1q32.2 region in a male Chinese patient presenting with typical VWS features using the Affymetrix Human SNP 6.0 Array. Consistent with his negative family history, the same deletion was not present in either parent and his two siblings (a younger sister and a younger brother) were also phenotypically normal. Using microsatellite markers, it had been established that the deletion occurred on the paternal chromosome. Although several known genes (including 11 OMIM genes and several micro RNAs) are deleted, this patient has no other abnormality apart from the orofacial presentations typical of VWS. He has been followed up since birth and there has been no clinically significant abnormality in any other organ system. Other than IRF6, the genes which are deleted in this patient appear to be insensitive to copy number and haploinsufficiency. The deletion in this patient overlaps with another VWS patient which is the only published case with array-CGH data. The female patient is also from this geographical region and her bigger 3 Mb deletion is also de novo. However, she has dysmorphism and some growth retardation in addition to the VWS phenotype. We will present a comparison of the different breakpoints and genes deleted in these two patients and discuss the implications.

2795F

A novel TPM3 gene muation with infantile nemaline myopathy. *S.O. NAM[1], J.H. SHIN[2], Y.J. LEE[1].* 1) Pediatrics, Pusan National University Children's Hospital, Yangsan, Gyeongnam, South Korea; 2) Neurology, Pusan National University Children's Hospital, Yangsan, Gyeongnam, South Korea.

A twelve month-old girl was transferred with respiratory failure. She had no specific perinatal problem with gestational age of 38 weeks 4 days and birth weight of 3,680g. At one month of age, her parents found that her muscle power was decreased. Her parents visited a nearby rehabilitationcenter and the patients had been treated with physical therapy. At 11 months of age, she had caught flu-like symptoms lasting for 2 weeks and progressed to abrupt cyanosis and respiratory failure and visited emergency room. She was admitted to intensive care unit on ventilation therapy after cardiopulmonary resuscitation for 8 minutes. She was tried to be weanedoff her ventilator but failed several times, so she was transferred to the pediatric neurologic clinic of our hospital for the evaluation of underlying disease. There was a family history of motor delay of her mother and her grandfather in her mother's side grown out at adulthood. Physical examination showed elongated face with high arched palate, but had no dysmorphism or visceromegaly. Neurologic examination showed alert mentality and normal cranial nerve function. However, there was marked gross motor delay without involvement of fine motor, language or cognition. Her muscle tone was markedly decreased with frog leg posture, positive head lag sign, and decreased deep tendon reflex. Her laboratory findings including serum creatine kinase and metabolic study were normal. Her brain MRI, electroencephalography, EMG/NCV and cardiac echocardiography were normal. Genetic evaluation including SMN1 gene, DMPK gene, fragile X PCR, and array CGH werealso normal. Light microscopic findings of muscle biopsy revealed selective type 1 fiber atrophy with accumulation of suspicious dense material in modified Gomori trichrome stain. Electromicroscopic findings showed severe loss of myofibrillary structure with nemaline rod accumulation within atrophic fiber. Whole exome sequencing was performed. The results showed 151 variants in genes previously reported to cause neuromuscular disorders. The number of candidate variants was reduced to 22 when screened by normal Korean SNP database. We found a novel missense heterozygous mutation in exon 1 of TPM3 gene (c.32T>A, p.Met11Lys) close to the known pathogenic mutation, compatible with the diagnosis of nemaline myopathy.

2796W

Male infant with Freeman-Sheldon syndrome and an incidental diagnosis of Duchenne/Becker muscular dystrophy. *K. Oishi, E. Quinn, M. Babcock, L. Edelmann, S. Scott, A. Yang, G. Diaz.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Rhabdomyolysis is a condition of rapid breakdown of damaged skeletal muscle triggered from various causes including exercise, hyperthermia, and anesthetic agents. Patients with inherited muscle disorders such as muscular dystrophies are at great risk for malignant hyperthermia and rhabdomyolysis upon using different types of anesthetic agents. This condition can be life threatening and usage of those medications should be avoided for those patients. We report an infant with concurrent diagnosis of Freeman-Sheldon syndrome (FSS) and Duchenne or Becker muscular dystrophies (DMD/BMD), who was admitted to our pediatric ICU for cerebral infarction and elevated CPK after general anesthesia. The patient was a 3-month old male born at full term without complications to non-consanguineous parents from the Dominican Republic with a benign family history. At birth, he had left undescended testis, right inguinal hernia, and developmental dysplasia of the hips. Several days after the cryptorchidism and hernia surgery, he developed a fever, abnormal eye movements, facial nerve palsy and hypertonia. Brain MRI showed watershed cerebral infarction and his CPK was elevated above 45,000 U/L. Physical examination revealed deep-set eyes, low-set ears, high-arched palate, H-shaped cutaneous dimpling of the chin, cortical thumbs, and contractures of wrist and ankle joints. The physical findings supported the diagnosis of FSS, an autosomal dominant disorder due to mutations in *MYH3* gene and characterized by typical facial features, distal arthrogryposis and susceptibility to malignant hyperthermia with rhabdomyolysis following exposure to inhalational anesthetics. Array CGH was done to look for contiguous gene deletion and incidentally identified a *de novo* 74.5-kb deletion on chromosome Xp21.1. The deletion included exons 45-47 of the *DMD* gene, which was predicted to cause DMD or BMD. This is the first reported case of two separate congenital myopathic diseases arising *de novo* in one individual. We speculate that the stroke episode was a complication of malignant hyperthermia with rhabdomyolysis triggered by sevoflurane for general anesthesia. Muscle rigidity has been reported in FSS and malignant hyperthermia with rhabdomyolysis has been described in older DMD/BMD patients, but not in infants. The occurrence of these features in this case, initially thought to be an atypical presentation of FSS, is likely to be a result of the two separate myopathic processes in this patient.

2797T

Whole-exome sequencing identifies a homozygous mutation in the SPG11 gene in patients with Spastic Paraplegia. *V. Adir[1], A. Shalata[2], E. Shahak[1], Z.U. Borochowitz[1].* 1) Simon Winter Institute for Human Genetics, Bnai Zion Medical Ctr., Haifa, Israel; 2) Genetic Institute, Ziv Medical Ctr., Zefat, Israel.

Hereditary spastic paraplegias (HSP) are neurodegenerative diseases which include a heterogeneous group of neurodegenerative diseases. HSP is mainly characterized by lower limb spasticity associated with additional neurological signs in the complicated forms. At least 52 loci and 31 causative genes have been identified. Although mutations in the SPAST gene explain approximately 40%; of the pure autosomal dominant forms, molecular diagnosis can be challenging for the sporadic and recessive forms, which are often complicated and clinically overlap with a broad number of movement disorders. In this study whole-exome sequencing was performed to reveal the genetic cause for the disease in a large family with consanguineous marriage in northern Israel. We sequenced two first cousins patients and the father of one of them. Exome sequencing revealed that both patients carry a novel homozygous nonsense mutation in the SPG11 gene (c.4339C>T; p.Q1447X). The presence of the mutation was confirmed by Sanger sequencing. We developed a PCR RFLP test that would enable easy analysis of the c.4339C>T mutation in the SPG11 gene. The mutation was not found in 100 control chromosomes from the same ethnicity. According to previous studies, the SPG11 encodes the spatacsin protein which is expressed ubiquitously in all tissues. The function of spatacsin remains unknown, but the fact that it is highly conserved among species suggests an essential biological function. Further studies which will reveal the normal function of spatacsin will contribute to the understanding of the pathology of the disease, and better therapy for the patients.

2798F

Identification of a new hypouricemia patient with a pathogenic *GLUT9/SLC2A9* mutation (R380W). *M. Sakiyama[1], H. Matsuo[1], T. Chiba[1], A. Nakayama[1], Y. Kawamura[1], S. Shimizu[1], N. Hamajima[2], N. Shinomiya[1].* 1) Dept. Integrative Physiol. Bio-Nano Med., National Defense Medical College, Tokorozawa, Japan; 2) Dept. Healthcare Admin., Nagoya Univ., Nagoya, Japan.

Renal hypouricemia is a common inherited disease characterized by low serum uric acid (SUA) levels. It is associated with severe complications such as urolithiasis and exercise-induced acute renal failure. We have previously reported that urate transporter 1 (*URAT1/SLC22A12*) and glucose transporter 9 (*GLUT9/SLC2A9*) are causative genes for renal hypouricemia type 1 (RHUC1) and renal hypouricemia type 2 (RHUC2), respectively. In the series of experiments, two families have been revealed to have RHUC2 due to *GLUT9/SLC2A9* missense mutations R198C or R380W. However, there are no reports of other RHUC2 families or patients with these pathogenic mutations until now. Thus, we performed mutational analysis of *GLUT9/SLC2A9* exon 6 (for R198C) and exon 10 (for R380W) in another 50 hypouricemia patients to find other cases of RHUC2. Patients were collected out of more than 2000 samples from J-MICC Study (the Japan Multi-Institutional Collaborative Cohort Study). In this study, one novel hyporuicemia male patient with heterogeneous RHUC2 mutation R380W was identified. His SUA level was 2.6 mg/dL, which is similar to that of our previous case (SUA level: 2.7 mg/dL). Mutation sites in *GLUT9/SLC2A9* (R380W and R198C) locate in highly conserved amino acid motifs in ‘sugar transport proteins signatures’, which is observed in GLUT family transporters. As for glucose transporter 1 (GLUT1) deficiency syndrome which is caused by the corresponding mutations in human *GLUT1* (R333W and R153C), arginine residues in this motif are reported to be an important determinant of membrane topology of GLUT1. The same is probably true in GLUT9 on the basis of membrane topology. In oocyte expression study, their mutants showed markedly reduced urate transport, which would be the result of loss of positive charges of those amino acid motifs. The identification of the new RHUC patient could help to expand the understanding of RHUC pathogenesis and suggest that these *GLUT9/SLC2A9* mutations cause renal hypouricemia by their decreased urate reabsorption on both sides of the renal proximal tubules. In addition, these findings also enable us to propose a physiological model of the renal urate reabsorption and can be a promising therapeutic target for hyperuricemia, gout and associated diseases, such as cardiovascular diseases, cerebrovascular diseases and renal failure.

2799W

The manifestation of Mowat-Wilson syndrome in adult identical twins. *S. Ramanathan[1], S.A. Ashwal[2], R.D. Clark[1].* 1) Pediatric Genetics, Loma Linda University, San Bernardino, CA; 2) Pediatric Neurology, Loma Linda University, San Bernardino, CA.

Mowat-Wilson syndrome (MWS) is characterized by typical dysmorphic facial features, congenital anomalies, including Hirschsprung disease, and moderate to severe intellectual disabilities. Little is known about the features of MWS in adults. We report on 23-year-old monozygotic twin females in whom the diagnosis was recently confirmed. The twins were diagnosed in infancy with an apparently de novo interstitial chromosome 2 deletion: 46,XX,del(2)(q21q23). The parents were seeking re-evaluation for new-onset ‘osteoarthritis’ in the twins when the diagnosis of MWS was made by oligoarray CGH. that showed a 9.7 megabase deletion, including the *ZEB2* gene associated with MWS. The prenatal history was non-contributory and the twins were delivered at 37 weeks with Apgar scores of 8^1 and 9^5. Their developmental delay was initially attributed to the twin gestation and lower birth weight (4lb 11oz, 5lb 2 oz) compared to their siblings. At 14 months, the twins were microcephalic and were not rolling over. MRI of the brain detected agenesis of corpus callosum in both children. Both had recurrent otitis media, upper respiratory infections and urinary tract infections secondary to hydronephrosis with reflux in childhood. Both have epilepsy, which is well-controlled in one. Both have chronic constipation. As adults, they are non-verbal. They babble, follow one-step commands and have some receptive language. They are not toilet-trained. Neither is able to chew food. They need to be restrained at meal-time, as they bat away at spoons. Both have bruxism and bite their wrists. Menarche was at 16 and 18 years and menses are irregular. The twins have had normal echocardiograms, and are awaiting rectal biopsies to rule out Hirschsprung disease. They are being followed by a rheumatologist for swelling of the interphalangeal joints, which has been reported in MWS, but is not well-characterized. Radiographs on the twins do show soft tissue, but no bony changes. Our patients illustrate the value of obtaining a diagnosis, irrespective of age. They add to the meager data on adults with MWS. The authors and the parents of these patients welcome contact with other families who have adults with MWS to collect data on the natural history of this disorder.

2800T

Occurrence of recurrent optic neuritis and cervical cord Schwannoma with Charcot-Marie-Tooth type 4B1 disease: A case report. *P. Scott, F. Almurshedi.* Genetics Dept, Col Med & Health Sci, Sultan Qaboos Univ, Muscat, Oman.

Charcot-Marie-Tooth (CMT) disease is a heterogeneous group of inherited peripheral neuropathies. CMT type 4 is characterized by autosomal recessive inheritance. CMT4B1 due to MTMR2 gene mutations is a rare subtype of CMT4 with reported association of facial weakness, vocal cord paresis, chest deformities, and claw hands. We report the unusual occurrence of recurrent optic neuritis and cervical cord Schwannoma in a male individual with confirmed CMT4B1 disease. This 28-year-old gentleman presented with bilateral, symmetrical limb weakness that started distally and progressed very slowly since the age of 4 years and became wheel chair dependent at the age of 22. Dysphonia with vocal cord paralysis was apparent as he grew. He developed 3 episodes of optic neuritis resulting in a reduced visual acuity. At the age of 27 years, he presented with symptoms suggestive of cervical cord compression with diagnosis of C2/C3 spinal cord Schwannoma. Physical examination was remarkable for facial nerve weakness, chest deformity and claw hands. A positive family history with two affected siblings and the presence of consanguinity suggest autosomal recessive inheritance. The clinical features were consistent with CMT4B1 and MTMR2 gene sequencing revealed a novel nonsense homozygous mutation c.1768C>T (p.Gln590*). The rare association of optic neuritis or Schwannoma with genetically confirmed CMT1A has been individually observed but never with recessive CMT. To the best of our knowledge, the occurrence of recurrent optic neuritis and cervical cord Schwannoma in the same patient has never been reported with any forms of CMT including CMT4B1.

2801F

Novel homozygous missense mutation in the matrix metalloproteinase 2 (MMP2) catalytic domain leading to protein loss of-function in two Italian sibs in the spectrum of Torg-Winchester syndrome. *J. Azzollini[1], D. Rovina[1], C. Gervasini[1], I. Parenti[1], A. Fratoni[2], L. Pietrogrande[3], L. Larizza[1].* 1) Medical Genetics, Department of Health Sciences, University of Milan, Milan, Italy; 2) Pediatrics Unit, G. Salvini Hospital, Rho, Italy; 3) Orthopaedic Division, Department of Health Sciences, University of Milan, Italy.

Torg-Winchester Syndrome (TWS, OMIM #259600) is an extremely rare autosomal recessive disorder characterized by multicentric osteolysis, especially involving hands and feet, subcutaneous nodules, arthropathy with progressive joint contractures and other features such as short stature, hyperpigmented skin lesions, coarse face, corneal opacities and gum hypertrophy. Mutations of MMP2 gene (16q13), encoding Matrix Metalloproteinase 2, have been associated with TWS: only 8 mutations, 3 truncating, 3 missense, 1 splice, 1 in-frame deletion have been described so far in 8 families from different geographic areas. We evaluated two Italian siblings (43♀, 37♂), born to healthy consanguineous parents, who started to display, between age 3 and 6, appendicular osteoporosis and palmar-plantar subcutaneous nodules followed, during adolescence, by tendon retractions, progressive upper limb arthropathies, toes osteolysis and pigmented fibrous skin lesions. The female also suffered from recurrent bilateral eye pterygium after age 30. Based on the clinical presentation and the STRs genotype shared by the sibs at 16q12.2-q21, we performed MMP2 mutation screening on their blood DNA by direct sequencing. Both siblings were found to carry a novel homozygous missense mutation in exon 8, c.1228G>C p.G410R, affecting a highly conserved aminocidic residue within MMP2 catalytic domain. The same mutation was not detected in 260 healthy individuals and is predicted pathological by several bioinformatic tools (Mutation Taster, PolyPhen-2, PMut, Mutation Assessor). Functional assessment was carried out by site-directed mutagenesis on the wild-type MMP2 sequence cloned into an expression vector, trasfection of both wild-type and mutated MMP2 into HEK cell line and gelatin-zymography on the collected conditioned media. As compared to the wild-type MMP2, G410R-MMP2 showed a complete loss of gelatinolytic activity, thus confirming the mutation impact on protein function. Out of 8 described MMP2 mutations, 3 map to exon 8, in the protein catalytic domain (p.E404K, p.V400del, p.G406D), and are associated with Winchester phenotype, mainly characterized by the absence of subcutaneous nodules. p.G410R represents the first inactivating missense mutation, within MMP2 catalytic domain, associated with Torg phenotype, thus reinforcing the complexity of genotype-phenotype correlation in TWS.

2802W

A patient with Simpson-Golabi-Behmel syndrome, Biliary Cirrhosis and successful Liver Transplantation. *B. Demeer[1], G. Guillaume Jedraszak[1], M. Girard[2], A. Mellos[2], D. Djeddi[3], C. Chardot[4], A. Varenterghem[3], M.P. Moizard[5], J. Gondry[6], H. Sevestre[7], M. Mathieu-Dramard[1], F. Lacaille[2].* 1) Genetics Dept, Hopital nord, CHU, Amiens, France; 2) Paediatric Hepatogastroenterology-Nutrition unit, Necker-Enfants-malades Hospital, France; 3) Paediatric hepatogastroenterology-Nutrition unit, C.H.U Amiens, France; 4) Paediatric surgical unit, Necker-Enfants-malades Hospital, France; 5) Service de génétique, INSERM U930, CHRU Tours, France; 6) Prenatal diagnosis unit, Camille Desmoulins maternity hospital / Jules Verne University, C.H.U Amiens, France; 7) Pathology department, C.H.U Amiens, France.

Simpson-Golabi-Behmel syndrome (SGBS) -OMIM 312870-, first reported by Simpson et al in 1975 is a rare X-linked inherited overgrowth syndrome caused by a loss-of-function mutation in the *GPC3* (SGBS type 1) or *GPC4* gene. Patients present a variable phenotype with pre- and post-natal macrosomia, distinctive facial dysmorphism, organomegaly, and numerous congenital anomalies, ie : diaphragmatic hernia, heart or renal defect, genitourinary tract or gastrointestinal malformations, skeletal or hand abnormalities. Intellectual disability is not constant. About 10% of patients develop embryonic tumors in early childhood. A single case of biliary disease (choledochal cyst) has been reported. *GPC3* is localised on Xq26 and encodes for Glypican-3, a glycosylphosphatidylinositol-linked cell surface heparan sulfate proteoglycan, that belongs to the family of glypicans. This report concerns a male infant with Simpson-Golabi-Behmel syndrome type 1, carrier of a *GPC3* mutation. He had neonatal liver disease, and developed early on biliary cirrhosis. Liver transplantation was discussed, considering the risks of cancer and intellectual disability, and performed successfully when the child was 19 months of age. A hypothesis on the role of *GPC3* in this patient's liver disease is proposed.

2803T

Exome Sequencing of a Proband With Bilateral Sensorineural Hearing Loss Furthers Insights into Phenotypic Variability of Waardenburg Syndrome. *A. Wilkens[1], M. Dulik[3], N. Qu[2,4], K. Wang[2,4], J. Brunton[1], V. Jayaraman[1], E. Dechene[1,3], N. Spinner[3], W. Liu[2,4], I. Krantz[1].* 1) Divison of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) BGI@CHOP Genome Center, The Children"s Hospital of Philadelphia, Philadelphia, PA; 3) Division of Pathology, Children's Hospital of Philadelphia, Philadelphia, PA; 4) BGI Health, Shenzhen, China.

Waardenburg Syndrome (WS) is characterized by sensorineural hearing loss and pigmentary differences of the hair, skin and iris with both phenotypic variability, as well as genetic heterogeneity leading to four subtypes. The absence or presence of dystopia canthorum (lateral displacement of the inner canthi) historically has directed clinical testing for two of the four subtypes, as WS type I (presence of dystopia canthorum) is associated with mutations in PAX3, whereas WS type II (absence of dystopia canthorum) is associated with mutations in MITF. We report on a 5-month old proband referred for bilateral sensorineural hearing loss whose physical exam was notable for dystopia canthorum by the presence of a Waardenburg-index of 2.04 (normal being less than 1.95) and whose father was also affected, suggesting dominant inheritance. PAX3 sequencing and deletion analysis was initiated on the proband and was negative. Additional samples from the proband, unaffected sibling, father and mother were obtained and exome sequencing was performed by BGI@CHOP Genome Center on an Illumina HiSeq 2000. Bioinformatic analysis of potentially pathogenic sequence variants shared in the proband and father was completed and led to a single candidate variant as the cause of the phenotype - an MITF p.R255X nonsense mutation previously associated with WS. Initially, this variant was not identified at a read depth coverage of 10x, but was revealed when the read depth coverage was relaxed to 8x. Sanger sequencing was performed to validate the finding and it was present in the proband and father, as well as in the presumed 'unaffected' sibling and was not present in the unaffected mother. This case report of using exome sequencing to aid in the diagnosis for a family with bilateral sensorineural hearing loss highlights the importance of: 1) this technology's ability to identify genetic alterations that would otherwise not have been tested for clinically; 2) amending historically established phenotypes to incorporate a larger clinical spectrum; 3) using lower coverage cut offs in an attempt to decrease false negatives; and 4) how the result of this testing can unearth reduced penetrance in presumed 'unaffected' family members.

2804F

Use of quantitative ultrasound for tibial dysplasia in neurofibromatosis type 1. *D.A. Stevenson, H. Slater, H. Hanson, A. Stevens, J.C. Carey, D.H. Viskochil.* Dept Pediatrics, Univ Utah, Salt Lake City, UT.

Neurofibromatosis type 1 (NF1) is a common autosomal dominant genetic disorder with distinct skeletal manifestations. In particular, tibial dysplasia is a difficult-to-treat morbid manifestation. Tibial dysplasia typically presents with unilateral anterolateral bowing of the tibia with cortical thickening and medullary canal narrowing, with subsequent fracture and non-union. In infancy, physiologic bowing of the lower leg can be confused with pathologic tibial dysplasia in NF1, and early intervention with bracing is thought to improve outcomes. Quantitative ultrasound (QUS) measures speed of sound, avoids radiation, and is reported to be predictive of clinical fracture in the general population. Our aim was to use QUS of the affected and unaffected tibias of individuals with NF1 who presented with suspected tibial dysplasia prior to fracture and pseudarthrosis. Bone quality was assessed on both tibia (the non-bowed and bowed tibia) using the Sunlight Omnisense 7000P scanner (Sunlight Medical, Israel) to measure the speed of sound (SOS) at the mid-shaft. The probe is moved in a sweeping motion, laterally and medially and three to five consecutive measurement cycles are performed, after which the SOS (m/s) is determined and z-scores generated using cross-sectional reference data of the same sex and age provided with the machine. A total of 14 individuals with NF1 with unilateral tibial bowing without fracture/pseudarthrosis were enrolled and both tibia were tested using QUS. The mean SOS z-score for the affected tibia was -2.9. The mean difference in SOS z-score was -2.5 when comparing the bowed tibia versus the individual's contralateral unaffected tibia. One infant with mild anterolateral bowing, radiographically showed no cortical thickening, and the z-score of the 'bowed' bone had a SOS z-score +2.5 difference compared to the other tibia suggesting physiologic bowing rather than pathologic tibial dysplasia. These data show that dysplastic tibia in NF1 prior to fracture and non-union have abnormal bone quality with significant decreases in SOS even though radiographically the tibia show a thickened cortex. These data also suggest that QUS is sensitive to distinguish dysplastic bowing vs. physiologic bowing in infancy in NF1. QUS is an attractive tool as a quantitative outcome measure for future trials aimed at improving tibial bowing to prevent fracture, and potentially for an aid in diagnosis and clinical management in NF1.

2805W

New syndrome of ectrodactyly and lethal pulmonary acinar dysplasia associated with homozygous FGFR2 mutation identified by exome sequencing. *C.P Barnett[1,4], N.J Nataren[2,3], M. Klingler-Hoffmann[2,3], Q. Schwarz[2,4], D.L. Bruno[5], J. Lipsett[8], A.J. McPhee[4,7], A.W. Schreiber[3,6], J.H. Feng[3,6], C.N. Hahn[2,4], H.S. Scott[2,3,4,6].* 1) SA Clinical Genetics, Women's and Children's Hospital/SA Pathology, North Adelaide, South Australia, Australia; 2) Department of Molecular Pathology, Centre for Cancer Biology, SA Pathology, Adelaide, SA, Australia; 3) School of Molecular and Biomedical Science, University of Adelaide, SA, Australia; 4) School of Medicine, University of Adelaide, SA, Australia; 5) Cytogenetics Laboratory, Murdoch Children s Research Institute, Royal Children s Hospital, Parkville, Australia; 6) ACRF Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology; 7) Department of Neonatal Medicine, Women's And Children's Hospital , North Adelaide, SA, Australia; 8) Department of Anatomical Pathology, Women's and Children's Hospital/SA Pathology, North Adelaide SA, Australia.

A female infant from a consanguineous union was born at term but died of respiratory failure on day 2. An autopsy revealed severe pulmonary hypoplasia due to acinar dysplasia/congenital pulmonary airway malformation (CPAM) type 0. The infant also had ectrodactyly involving the hands and feet. Ectrodactyly is genetically heterogeneous and acinar dysplasia is a rare lethal congenital lung lesion of unknown etiology. Recessive inheritance has been suggested. SNP array and whole exome sequencing of this infant's DNA identified 22 homozygous candidate causative sequence variants (SV) after filtering against dbSNP and numerous published and in-house exomes. Furthermore, analyses based on prediction of pathogenicity, protein conservation and expression (especially embryonic lung and limb), highlighted one SV in the FGFR2 gene, homozygous missense variant R255Q. Numerous germline FGFR2 mutations have been described in association with human disease, all heterozygous and mostly de novo autosomal dominant mutations associated with craniosynostosis syndromes. Involvement of the hands is classically spared in some (e.g. Crouzon syndrome) while Apert and Pfeiffer syndromes have both craniosynostosis and syndactyly of the hands and feet. Many FGFR2 missense mutations have been described as constitutively activating mutations, mainly in the extracellular ligand binding third Ig-like IgIII domain (D3) domain. Both parents of the affected infant were heterozygous for R255Q and unaffected which may suggest that R255Q is a loss of function allele. However, functional studies show that R255Q is constitutively activating and leads to rapid dimerizaton. Thus a single copy of R255Q may quantitatively and/or selectively (via altered FGF binding) activate FGFR2 downstream signaling in a tissue specific manner compared to other syndromes associated with FGFR2 mutations, possibly explaining this difference. Others have created a knock-in Crouzon syndrome mouse with a recurrent autosomal dominant activating FGFR2 mutation W290R, also in D3. While heterozygous mice show craniofacial malformations, homozygous mice also showed limb and organ (including lung) agenesis and die of respiratory failure. Our case is the first description of a human with homozygous FGFR2 mutations and is the first with profound lung defects. The combination of ectrodactyly and acinar dysplasia has not been described before and we propose it represents a new syndrome caused by homozygous FGFR2 mutations.

2806T

A GC polymorphism associated with serum 25(OH)D level is the risk for hip fracture in Japanese patients with rheumatoid arthritis. *S. YOSHIDA[1, 2], K. IKARI[1], T. FURUYA[1], Y. TOYAMA[2], A. TANIGUHI[1], H. YAMANAKA[1], S. MOMOHARA[1].* 1) Institute of Rheumatology, Tokyo Women's Medical University, Shinjuku, Tokyo, Japan; 2) Department of Orthopaedic Surgery, School of Medicine, Keio University, Tokyo, Japan.

Background: Vitamin D is important for the maintenance of the musculoskeletal system, is positively associated with muscle strength and physical performance, and is inversely associated with fall and fracture risk. The major determinants of serum 25-hydroxyvitamin D [25(OH)D] concentration are sunlight exposure and intake from diet or supplements. However, genetic variants affecting serum 25(OH)D concentration were recently identified in a meta-analysis of genome-wide association studies in Caucasian populations. We investigated the genetic variants associated with serum vitamin D concentration and the occurrence of hip fracture in Japanese patients with rheumatoid arthritis (RA). Methods: DNA samples of 2068 Japanese patients with RA were obtained from the Institute of Rheumatology Rheumatoid Arthritis cohort study (IORRA) DNA collection. Serum 25(OH)D concentration was measured in 932 of 2068 patients in the spring of 2011. Five of the single nucleotide polymorphisms (SNPs) reported in the recent studies were genotyped: rs2282679, in GC; rs3829251, in NADSYN1; rs12785878, rs1790349, near and in DHCR7; and rs10741657, near CYP2R1. The occurrence of hip fractures was determined from the responses to a patient questionnaire every 6 months from October 2000 to October 2010. After confirmed by review of medical records and radiographs, 39 hip fractures in 39 patients were included into this study. Multivariate linear regression analyses adjusted for the non-genetic factors were performed to investigate the association between serum 25(OH)D concentration and each SNP in 932 patients. The SNP showing a significant association with serum 25(OH)D concentration was included into the following prediction analysis for the occurrence of hip fracture by using a multivariate Cox proportional hazards regression model in 1957 patients with no data deficiency. Results: Multivariate linear regression analyses showed that rs2282679 in GC was significantly associated with lower serum 25(OH)D concentration ($P=5.8\times10^{-5}$). A multivariate Cox proportional hazards regression model indicated that rs2282679 in GC was significantly associated with the occurrence of hip fracture [HR (95% CI): 2.03 (1.03 to 5.94); P=0.042]. Conclusion: Our data indicated that a GC polymorphism is the risk for hip fracture in Japanese patients with RA. These results may contribute to a better understanding of the biological impact of genetic variation within the vitamin D metabolic pathway.

2807F

Fetal alcohol syndrome and assessment of maxillary and mandibular arc measurements. *K. Abell[1], W. May[1], P. May[2,3], W. Kahlberg[3], G. Hoyme[4], O. Abdul-Rahman[1].* 1) University of Mississippi Medical Center, Jackson, MS; 2) University of New Mexico, Albuquerque, NM; 3) University of North Carolina, Chapel Hill, NC; 4) University of South Dakota, Sioux Falls, SD.

Fetal alcohol spectrum disorders (FASD) are a range of physical differences and neurologic deficits due to prenatal alcohol exposure. Diagnosis evaluates characteristics like growth deficiency, microcephaly, and cardinal facial features. Other features like midface hypoplasia are also seen. While traditionally judged subjectively, objective measurements of maxillary and mandibular arcs and the ratio between them may be useful in assessing midface hypoplasia. The Fetal Alcohol Syndrome Epidemiologic Research (FASER) project contains dysmorphology assessments of children for FAS. Data are collected during in-school screenings of first-graders for height, weight, and head circumference. Cases are selected by measurements below the 10th percentile, and matched with normal controls. All subjects are evaluated by a dysmorphologist for FASD physical features. Those with possible features undergo neuropsychiatric evaluation and maternal interview for a diagnosis of FAS, partial FAS, alcohol-related neurodevelopmental disorder (ARND), or non-FAS. Using the FASER database, we investigated the size of the maxillary and mandibular arcs and the arc ratio in respect to FAS diagnosis.

We established normative values for the maxillary arc, mandibular arc, and the maxillary-to-mandibular arc ratio for males and females. In our control group of 545 males and 436 females, the mean maxillary and mandibular arcs for males/females were 24.98/24.52cm and 25.91/25.35cm respectively. The ratio was 0.9643 and 0.9676 for males and females respectively. We evaluated the effect of microcephaly, short stature, and low weight, defined as less than 10th centile, individually on arc measurements in the controls. With a few gender-based differences, the arc measurements were reduced significantly but the ratio did not differ. We compared our controls to 138 males and 135 females FASD cases. We noted a significant difference in arc measurements in male and female groups but not the ratio. The trend was greatest for a diagnosis of FAS but also seen in partial FAS and ARND. We compared non-FAS controls with reduced growth parameters to similar cases with a FASD. We did not find a significant difference in the arc or the ratio measurements. Therefore, we conclude the effect of prenatal alcohol exposure on maxillary and mandibular arc measurements is primarily on growth parameters and less directly on maxillary and mandibular growth.

2808W

Genetic analysis of Gonadal Disorders of Sex Development (46,XY DSD) by cytogenetic and molecular methods. *A. Shojaei[1], F. baghbani-arani[2], R. Ebrahimzadeh-Vesal[1], F. Behjati[3], J. Tavakkoly-Bazzaz[1].* 1) Medical Genetics, Tehran University of Medical Science, Tehran, Iran; 2) Genetics Dept, Varamin-Pishva Branch,Islamic Azad University, Varamin, Iran; 3) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Among Disorders of sex development (DSD), 46,XY gonadal dysgenesis is the most complicated, heterogeneous and rare disease.Various genes have been associated with gonadal dysgenesis. Among these, SRY, NR5A1, DHH, DAX1 and WNT4 genes, are clearly known to be associated with these disorders.Mutations of known genes are responsible for only a few percentages of DSD;therefore, there are probably some other potential loci that play an important role in sexual disorders like gonadal dysgenesis. In this study we recruited38 patients that were clinically suspicious for 46,XY gonadal dysgenesis of unknown etiology. Cytogenetic analysis as well as direct sequencing of the SRY, NR5A1and DHH genes was performed. In continue MLPA was used to detect deletions and duplications in DAX1 and WNT4 and subsequent imbalances were confirmed by real time PCR. Additionally, other potential loci were investigated by whole genome Array CGH method.In this study,one new chromosomal rearrangement and SRY deletions were found in one and five patients, respectively. Previously described NR5A1and DHH allelic variants were observed.A heterozygous partial deletion was presented in NR5A1 gene and heterozygous partial duplication was found in WNT4 gene. These deletions/duplications were subsequently confirmed by real time PCR. Array CGH results confirmed the chromosomal abnormality in the patient with abnormal chromosome 13, derived from a maternal translocation between chromosome 7 and 13 and showed the exact region of rearrangements. Also, one partial deletion was detected in the SOX2OTgene. Autosomal chromosome abnormalities could also play a role in disorders of sex developments. SRY gene deletion still has a significant role in these disorders and has a similar incidence in our patients compared with other reports and should be the first gene for testing in GD. Del/dup mutations found to be more common than point mutations in our patients. Therefore, it might be preferred to check Del/dup mutations prior to point mutations.SOX2OT might have a potential role in gonadal dysgenesis and it should be taken into account in molecular approaches to study GD patients.Array CGH is a valuable tool for finding responsible genes in GD and could unravel some potential loci in this regard.

2809T

BECKWITH-WIEDEMANN SYNDROME: A FAMILIAL CASE REPORT.

M.A. Aceves-Aceves[1], I.M. Salazar-Dávalos[1], M.G. González-Mercado[2], R.E. Jiménez-Arredondo[2], S.A. Alonso-Barragán[2], M. Salazar-Páramo[3], N.O. Dávalos[4], D. García-Cruz[4], C. Roa[5], I.P. Dávalos[2]. 1) Instituto de Genética Humana, Facultad de Medicina, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) Doctorado Genética Humana, IGH, CUCS, Universidad de Guadalajara, CIBO-IMSS, Guadalajara, México; 3) Depto. Fisiología, Universidad de Guadalajara, Div. Investigación, UMAE, HE, CMNO, Guadalajara, México; 4) Doctorado Genética Humana, Instituto de Genética, CUCS, Universidad de Guadalajara, México; 5) UMAE, Hospital de Pediatría, CMNO, IMSS, Guadalajara, Jalisco, México.

Beckwith- Wiedemann Syndrome (BWS)(OMIM 130650) is a genetic disorder with complex inheritance mode. Possible patterns include autosomal dominant inheritance with variable expressivity, contiguous gene duplication on chromosome 11p15, and genomic imprinting resulting from a defect or absence of the maternal copy gene. BWS has a prevalence of 1 in 13,700 births; it is characterized by overgrowth, macrosomia, macroglossia, organomegaly, exomphalos and predisposition to embryonal tumor development, most common are Wilms tumor or nephroblastoma. Clinical diagnosis is primarly, requiring 3 major criteria or 2 major and 3 minor criteria. OBJECTIVE: To present a familial case of BWS. CASE REPORT: Family data: Mother with antecedent of 4 pregnancies, 2nd correspond to case 1 and 4th to case 2, she presents ear pits. Case 1: 8 years 5 months female, product of the 2nd pregnancy term, obtained by caesarian section from non-consanguineous parents, with 28 years (he) and 29 years (she) at birth time. Apgar 9-10, weight and height >pc97, macroglossia, depressed nasal bridge, posterior helical ear pits and exomphalos surgically corrected, presented neonatal hypoglycemia. During the first year of age presented overgrowth, renal ultrasonography reported right pyelic duplication. At the age of 3 years was performed transversal and anteroposterior reduction glossectomy. Actually weight and height pct 97 and hemihypertrophy was detected.Case 2: Two years 11 months female, product of the 4th pregnancy, obtained by cesarian section at 30 weeks, weight at birth 2500g. Physical examination: weight and height > pct 97, posterior helical ear pits, macroglossia, umbilical hernia; has not required surgical procedure. Renal ultrasonography without abnormalities. Conclusion: Clinical findings of both cases have BWS criteria diagnosis. The treatment consisted of surgical reduction of exomphalos and anterior-transversal reduction glossectomy in case 1. The patient of case 2 has not required surgical intervention. The renal ultrasonography in patient 1 reports right pyelic duplication, meanwhile patient 2 renal ultrasonography was normal. The early diagnosis of BWS allows a complete treatment and a genetic counseling. The early and continuous monitoring of glycemia level in the first days after birth provide an opportune treatment of hypoglycemia and could prevent neurologic complications. The abdominal and renal ultrasonography support early diagnosis of kidney tumor.

2810F

Gastric necrosis and dilation with or without rupture a cause of death in Prader-Willi Syndrome: Three new case reports with a review of the mortality literature and an update in causes of death in PWS.

J.A Gold[1,2], R.T Rivera[3], J. Heinemann[4], A. Scheimann[5,6], P.J Gold[7], S.B Cassidy[8], J.L Miller[9]. 1) Pediatrics, Loma Linda University Medical Center, Loma Linda, CA; 2) Division of Pediatrics, Department of Genetics and Metabolism University California Irvine Medical School, Irvine , CA; 3) Department of Surgery, Providence Tarzana Medical Center, Tarzana, CA; 4) Prader-Willi Syndrome Association (USA), Sarasota, FL; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Department of Pediatrics, John Hopkins School of Medicine, Baltimore, MD; 7) Department of General Surgery, Warwick University Medical School, Warwick, UK; 8) Department of Pediatrics, Division of Medical Genetics, University of California, San Francisco, CA; 9) Department of Pediatrics, University of Florida, Gainesville, FL, USA.

Prader-Willi syndrome (PWS) is a neurodevelopmental disorder. While obesity is only part of the disorder, binge eating and periods of rapid weight gain and weight loss are common. Individuals have a high pain threshold and rarely vomit. A 3% mortality due to gastric necrosis and rupture was previously reported. We report three individual's age range 8-50 year's one female and two males with PWS and death due to gastric necrosis or dilation and rupture in one case with confirmed binge eating in one with suspicion of a binge in another. One of the complications of PWS is constipation and gastroparesis. It has previously been speculated that in individuals with PWS, episodes of food foraging and episodic hyperphagia cause acute dilation. This is likely to result in relative ischemia or occlusion of blood vessels, ischemia and gastric rupture. Stomach muscle integrity due to prior weight loss has also been hypothesized to be contributory factor (Wharton 1997). A thorough literature review was undertaken. Although rare, it appears that gastric necrosis, dilation and gastric rupture is more common in individuals who have an eating disorder such as bulimia and anorexia nervosa. It appears that a low or near normal basal metabolic index (BMI) is more common in those who suffer this complication after a prior episode of obesity, but is not always present. These cases and the literature review support strict adherence to dietary control from birth and prevention of vacillating weights. Vigilance on the part of emergency departments to be aware of the possible complication and need for CT and emergency laparotomy. Future long term outcome studies in PWS and other eating disorders including bulimia and anorexia nervosa is imperative. It would be helpful to determine if the use of growth hormone aids this goal by improving weight management. I hypothesize that a gastrostomy tube (GT) may cause scaring and a weakness making this a complication of GT insertion and should be avoided in individuals with PWS. It will be important to document if individuals who receive gastrostomy tubes in the neonatal period could have a nidus for gastric rupture at a weakened GT site or due to scaring.

2811W

The Coordination of Rare Diseases at Sanford (CoRDS) patient registry for all rare diseases and those undiagnosed.
E.A. Donohue, N.A. Simpson, R.M. Bourscheid, D.A. Pearce. Sanford Children's Heath Research Center, Sanford Research, Sioux Falls, SD.

BACKGROUND Sanford Children's Health Research Center at Sanford Research has established a rare disease registry named CoRDS (Coordination for Rare Diseases at Sanford). CoRDS houses de-identified contact and clinical information for patients who have been diagnosed with any rare disease. CoRDS' mission is to accelerate rare disease research by creating a resource of rare disease data for researchers and a mechanism by which participants can be contacted about research opportunities. METHODS CoRDS serves as a rare disease patient registry with several components. First, there is a data collection component, in which the patient-reported data is collected. Second, data management involves the archiving, collation, and accumulation of data. Third, dissemination of data ensures that researchers can access and utilize information for their research. A Scientific Advisory Board has been established to provide oversight and review applications to ensure that researchers have IRB approval and the research proposal is sound. CoRDS utilizes Velos eResearch, a clinical research information system supporting the collection of data and processes in study design/monitoring/execution, patient recruitment, reporting, data integration, compliance and safety monitoring for its software solution. Participant recruitment is closely tied to the process by which CoRDS partners with Patient Advocacy Groups (PAGs) that support patients diagnosed with rare diseases. PAGs communicate enrollment information to their membership and educate their researchers about the opportunity to utilize CoRDS as a resource. RESULTS CoRDS has 853 participants (446 enrolled, 407 in the screening), representing 127 rare diseases. CoRDS has partnered with 44 PAGs and organizations in the medical and rare disease field. These partnerships help increase participant enrollment and disease numbers. DISCUSSION Collecting and collating data on patients diagnosed with any rare disease offers the opportunity to perform a comparative analysis to better understand and treat the diseases. Many treatments are symptomatic, thus treatment strategies for one disease may be beneficial in application to other diseases with similar clinical profiles. CoRDS serves as a source of data to help researchers better understand the particular disease they are studying and can help them identify potential participants for research studies or clinical trials with the potential to accelerate their timelines.

2812T

Homoplasmy of a mitochondrial 3697G>A mutation causes Leigh syndrome.
Y. Negishi[1], A. Hattori[1], E. Takeshita[2], C. Sakai[2], N. Ando[1], T. Ito[1], Y. Goto[2], S. Saitoh[1]. 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan.

Leigh syndrome (LS) is a subacute necrotizing encephalomyelopathy characterized by bilateral symmetrical necrotic lesions in the basal ganglia and brainstem. We report that three siblings with LS harbor the m.3697G>A mutation in a homoplasmic fashion. Three siblings from healthy parents are all affected with LS of different severity. Patient 1 is a 9-year-old girl presenting with progressive gait disturbance from 18 months. Currently, she can walk without support, and speak sentences. Patient 2 is a 7-year-old boy. Progressive rigidity of four limbs started at 2 years. Currently, he cannot stand without support and does not have meaningful words. Patient 3 is a 5-year-old girl. She presented with progressive gait disturbance from 18 months. Her condition severely deteriorated after viral infection at 2 years, and has been bedridden with tracheostomy since then. Brain MRI showed necrotic lesions in the bilateral basal ganglia in all patients, but brainstem involvement only in patient 3. Entire mitochondrial DNA sequencing of a biceps brachii muscle specimen from patient 1 revealed that this patient harbored the m.3697G>A substitution (G131S) of the ND1 gene. This mutation was homoplasmic in all three siblings and heteroplasmic in their mother. Heteroplasmic m.3697G>A has been reported in patients with Leber's hereditary optic neuropathy (LHON) and mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS). Since all of our three patients carried m.3697G>A in a homoplasmic fashion and showed LS, it is conceivable that homoplasmy of m.3697G>A causes the LS phenotype. The mechanism by which the three siblings developed homoplasmy while the mother carried the mutation in a heteroplasmic fashion can be explained by the bottleneck theory. This phenomenon should be taken into account for genetic counseling of mitochondrial disorders.

2813F

Defining the Undefined Congenital Myopathies: Pathological Findings and Clinical Features.
L.C. Swanson[1], P.D.S.C. Ciarlini[1, 2], A.H. Beggs[1, 2]. 1) Boston Children's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA.

The congenital myopathies are a group of rare and genetically heterogenous neuromuscular disorders that affect skeletal muscle, often causing numerous symptoms including, but not limited to, non-progressive muscle weakness, hypotonia, respiratory insufficiency, scoliosis, facial weakness, and feeding difficulties. The severity of symptoms and age of onset are variable among individuals with the same diagnosis. Mutations in different genes can lead to similar histopathological features, and mutations in the same gene can give rise to variable clinical and pathological phenotypes. Several specific congenital myopathy diagnoses include nemaline myopathy, congenital fiber type disproportion, myotubular/centronuclear myopathy, and central core/multi-minicore disease. The diagnoses are based on histopathological analysis, by which the muscle is reviewed for specific features associated with a congenital myopathy diagnosis, such as central nuclei or nemaline rods. When none of these defining features are present and a distinct congenital myopathy diagnosis cannot be determined based on histopathological analysis, the patient may be said to have an undefined congenital myopathy. We report the retrospective review of 61 patients enrolled in the Beggs Congenital Myopathy Research Program with a diagnosis of undefined congenital myopathy who had pathology reports and muscle biopsies available. After this review, 34 cases were reclassified with a pathological diagnosis of a specific congenital myopathy including centronuclear myopathy, congenital fiber type disproportion, and multi-minicore disease or another neuromuscular condition like congenital muscular dystrophy. Of the remaining 27 patients, 7 patients had varying clinical presentations including myasthenic syndrome, chronic pain, and facial weakness, whereas 20 patients presented with weakness or hypotonia at early age as a unifying feature. Seven of this latter group had additional findings not limited to muscle, including a patient with Vici syndrome. The remaining 13 patients had a clinical history consistent with a congenital myopathy, with delayed motor milestones, feeding difficulties, and diffuse hypotonia, but with normal or nonspecific findings on biopsy. We explore the clinical courses, possible genetic causes, and the pathological findings present in the patients originally classified as undefined congenital myopathy.

2814W

Molecular and clinical study of 30 Angelman syndrome patients with *UBE3A* mutations.
K. Hosoki[1], S. Saitoh[2]. 1) Department of Pediatrics, Hokkaido University Graduate School of Medicene, Sapporo, Japan; 2) Department of Pediatrics and Neonatology, Nagoya City University School of Medical Sciences. Nagoya, Japan.

Angelman syndrome (AS) is a neurodevelopmental disorder related to genomic imprinting at 15q11-q13. Genetically, AS is classified in five genetic groups; approximately 70% are caused by deletion at 15q11-q13, 5% paternal uniparental disomy 15, 5% imprinting defects, 10% *UBE3A* mutations, and 10% unknown etiologies. Deletion positive patients have been reported to show more severe clinical features than those with patients in other genetic groups including *UBE3A* mutation group. In order to delineate molecular and clinical characteristics in AS patients with *UBE3A* mutations, we reviewed identified mutations in the *UBE3A* gene in our series of molecular testing for patients with AS, and tried find genotype-phenotype correlation. We detected 26 *UBE3A* mutations in 30 patients with AS, including 7 familial and 23 sporadic patients. It is of note that 7/26 (27%) mutations were present in exon 16. Frameshift mutations were most predominant (18/26, 69%). Seventy three % of mutations have not been reported previously. Next, we compared clinical features of the patients with *UBE3A* mutations to those of 10 AS patients with typical 5Mb deletions. There were no significant difference between two groups for major clinical features of AS. However, motor developmental milestones including head control, independent sitting and independent walking were significantly less delayed in patients with *UBE3A* mutations than in patients with typical deletions. Therefore, our study indicates that patients with *UBE3A* mutations show milder phenotype than patients with typical deletions. Typical deletion encompasses three genes (*GABRB3*, *GABRA5* and *GABRG3*) encoding subunits of $GABA_A$ receptor. Haploinsufficiency of these genes may explain more severe phenotype in patients with typical deletions.

2815T

Stargardt Disease, a Clinical Description. *L. Mora[1], M. Tamayo[1,2], F. Rodriguez[2], M. Valencia[2].* 1) Instituto de Genética Humana, PontificiaUniversidad Javeriana, Bogotá, Colombia; 2) Fundación Oftalmológica Nacional, Bogotá, Colombia.

INTRODUCTION. Stargardt's disease or juvenile macular degeneration, first described in 1909 by Karl Stargardt, is known as a nosologic entity of autosomal recessive (AR) inheritance. It is the most common cause of macular degeneration in childhood, due to accumulation of lipofuscin in the retinal pigment endothelium and limited energy exchange at level of the photoreceptors; clinically evident between 7 and 12, is shown as a progressive deterioration of visual acuity. It has been reported worldwide incidence of 1/10000, but no data are available at the Colombian population. **MATERIALS AND METHODS.** * 10 patients (4 men - 6 women) with clinical diagnosis of Stargardt disease. They were obtained from the database of patients with genetics eye diseases of the National Ophthalmological Foundation (FUNDONAL). * Average age now 39.2 years, range 12-69 years. * The patients were evaluated with a complete medical history, fundoscopy under pupil dilation and best corrected visual acuity using Snell scale and fluorescein angiography. * 6 patients were evaluated with OCT (Optical Coherence Tomography) images, which were performed using 5 lines radiated scanner, manually centered on the fovea. **RESULTS.** * 100% of patients present with best corrected visual acuity of > 20/100. * 80% of those tested were found in stage I / II. * Inverse correlation was found between visual acuity and foveal thickness. * Symptoms start in the first two decades of life in 90% of the patients. * We consider the AR inheritance mechanism as the most likely in the study population. **DISCUSSION.** Stargardt's disease is a progressive disease, affecting individuals in the first decade of life. In the clinical and paraclinical description of 10 individuals in collaboration of the Institute of Human Genetics at the Javeriana University and National Ophthalmological Foundation, it was established that for our population the most frequently reported symptoms in the first decade of life include nictalopia and tunnel vision. Symptoms have slow progression, most patients were in stage I and II, this being compatible with a visual acuity of > 20/100, with an average of evolution of 20 years. The OCT findings showfoveomacular reduced thickness and loss of photoreceptors in the macular area, which can be useful in the diagnosis of Stargardt disease, it was found strong correlation between foveal thickness and visual function in patients sample.

2816F

Mixed phenotype of incontinentia pigmenti and anhidrotic ectodermal dysplasia with immunodeficiency in a patient with duplication of the IKBKG gene. *A. Ramalingam, E. van Asbeck, T.J. Chen, E. Morava.* Human Genetics, Tulane University School of Medicine, New Orleans, LA.

/We report on a 30 year old female patient with a mixed phenotype of incontinentia pigmenti and anhidrotic ectodermal dysplasia with immunodeficiency in combination with normal intelligence and a progressive neurologic disease. The patient presented with recurrent benign tumors, multiple hemangiomas and cysts and arteriovenous malformation. She has cysts in her ear canal and neurinoma. Besides fatigue, muscle pain and generalized muscle weakness she also developed rapidly progressive peripheral neuropathy, macrocephaly with absence of seizures and hearing loss. She has experienced sudden visual loss on her right eye due to retinal hemorrhage. She presented with progressive problems with gastric emptying and gastroparesis necessitating a G-tube. Skin symptoms with color changes in a linear pattern, decreased hair growth, patchy alopecia and decreased sweating appeared after puberty. She also has severe and recurrent life threatening infections with compromised immunity but without obvious T or B cell deficiency or low immunoglobulin levels. A duplication of approximately 136 kb at Xq28 was identified by array CGH, and confirmed by qPCR. G6PD and IKBKG genes were fully duplicated while FAM3A and GAB3 were partially duplicated. Deletions, partial duplications and point mutations in IKBKG have been detected in patients with X-linked anhidrotic ectodermal dysplasia with immunodeficiency (XL-EDA-ID). The clinical phenotype of our patient supported our postulation that the fully duplicated IKBKG also played a pathogenic role in XL-EDA-ID patients. Our patient shows a progressive, mixed phenotype of both incontinentia pigmenti and anhidrotic ectodermal dysplasia with immunodeficiency, suggesting more complicated pathogenic mechanisms, such as reported somatic mosaicism, variable expression in different tissues, and skewed X-inactivation. IKBKG is likely to involve in these mechanisms. Further characterization of this duplication and its pathogenic role is in progress.

2817W

Single Molecule Targeted Sequencing of Long Fragments (>1kb) for Ovarian Hypersensitivity Syndrome. *F. Orkunoglu-Suer[1,2], A. Harralson[3], D. Frankfurter[4], P. Gindoff[4], E. Hoffman[1,2], T.J. O'Brien[5].* 1) Children's National Medical, Research Institute, Genetics in Medicine, NW, Washington DC, 20010; 2) Dept of Integrated System Biology, Sciences Department, the George Washington University Medical Center, Washington, DC 20037; 3) Department of Pharmacogenomics, Bernard J. Dunn School of Pharmacy, Shenandoah University, Ashburn, VA; 4) Department of Obstetrics and Gynecology, The George Washington University Medical Center, Washington,; 5) Department of Pharmacology and Physiology, the George Washington University Medical Center, Washington, DC 20037.

Targeted resequencing using next-generation sequencing technology is being rapidly applied to the molecular diagnosis of genetic diseases. To identify novel predictive genetic biomarkers for ovarian hyper stimulation syndrome (OHSS) that is multi-genic in nature and requires rapid genomic techniques for the identification of risk alleles we employed targeted single molecule sequencing technology. To perform a proof-of-concept study, we selected four patients with OHSS. We developed a custom Targeted Single Molecule Sequencing (T-SMS) Panel; total 44 loci and 3756 primer pairs using long 1 kb standard fragment size with droplet PCR technology (Rain DanceTM). T-SMS was carried out using single molecule real time (SMRT) DNA sequencing (Pacific BiosciencesTM). Secondary analyses were conducted using the Genome Analysis Toolkit for SNP discovery embedded in Smart Portal followed with ANNOVAR for functional annotations. Filtered functional variants further validated using conventional Sanger sequencing in original samples. Target enrichment PCR yielded amplicons averaging 1 kb fragment size. The total targeted sequence was 3.18 Mb per sample. Mean mapped CCS read accuracy was 97%, with filtering set default >10X CCS. Using this approach, we have identified a non-synonymous variant LHCGR (NP_000224.2:p.Asn291Ser; rs12470652) in two severe OHSS cases. Critically, combining larger fragments and droplet PCR targeted approach helped us to overcome gDNA and other fragments that may occupy unnecessary sequencing cells, and cause reduction of reads. Single Molecule Sequencing technology provided a new resolution by providing significantly longer reads, reads of GC reach sequences, sequencing of many target genes in a short time (45 min on instrument), low error profile that has promises for clinical diagnosis. Best of our knowledge this is the first clinical study reporting effective sequencing of 1 kb size amplicons utilizing droplet PCR technology. This data show excellent utility for follow-up validation and extension studies in ovarian hypersensitive syndrome and, panel that can be further modified and extended to assisted reproductive therapy and infertility genomics.

2818T

Myocardial thickness and ventricular Tei index by echocardiography among normal developing fetuses. *Y-H. Zhang[1,2], Q-Y. Cao[1,2], J. Ge[1,2], N. Zhong[1,2,3,4].* 1) Ultrasound, The 4th Municipal Hospital of Shijiazhuang, Shijiazhuang, China; 2) Center of Translational Medicine for Maternal and Children Health, The Maternal Health and Obstetric Hospital, Shijiazhuang, China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, New York.

Congenital heart disease (CHD) has become the top list of birth defect in China in the recent years. In order to accurately monitor the fetal heart development, normal developmental parameters are necessary to be established. For which, we have undertook a study to obtain the normal fetal myocardial thickness and ventricular Tei index, and to explore its growth regularities as well as reference ranges of normal fetal growth among 257 normal singleton fetuses. These fetuses were randomly selected from normal pregnancies that underwent prenatal ultrasound screening for birth defect in our hospital. The fetal myocardial thickness of left and right ventricular wall, ventricular septum, and left and right ventricular apex were measured from a four-chamber view during the relaxation period. The blood flow frequency spectrums of left and right ventricular inflow and outflow were obtained respectively using pulse Doppler ultrasound. The corresponding time intervals were measured and corrected by heart rates, and then Tei index was calculated. The fetal myocardial thickness and ventricular Tei index varied in different pregnant weeks. Fetal myocardial thickness grew with the increase of pregnant weeks and there was a linear correlation between them. Right ventricular wall was thicker than other parts. The ventricular Tei index decreased gradually with the increase of gestational age and Tei index of right ventricle was lower than that of the left ventricle. Fetal myocardial thickness and Tei index could be estimated accurately by echocardiography. Reference ranges of normal fetuses with different gestational ages are helpful in early detection and timely management of fetal abnormalities.

2819W

Harlequin Ichtyosis: A rare case of two consecutive harlequin ichthyosis in the offspring of a sequindigravida. *O. Franciska, R. Santos.* Obstetric Gynecology department, Jose R Reyes Memorial Medical Center, Rizal Avenue, Manila, Philippines, MD.

ABSTRACT: Harlequin fetus is rare and the most severe form of congenital ichthyosis, inherited as autosomal recessive trait with an incidence of about 1 in 300,000 births. It is characterized by hyperkeratosis and desquamation of the epidermis which begins prenatally. The skin barrier is severely compromised, leading to excessive water loss, electrolyte imbalance, temperature dysregulation and an increased risk of life threatening infection. Here in our institution, we report a case of a pregnant woman with a previously born child affected with Harlequin Ichthyosis. For this second pregnancy, they were expecting to have a normal baby since the congenital sonographic scan done prenatally was unremarkable. However, the mother delivered at 29-30 weeks age of gestation with the fetus showing signs of Harlequin Ichtyosis. The child was a 1200 gram female neonate born prematurely by partial breech extraction. Clinical manifestation of harlequin ichthyosis were present at birth. Furthermore, the fetus suffered from respiratory distress and was very ill, few hours after delivery and died few days after birth. Prenatal Diagnosis of HI is made with the use ultrasound guided fetal skin biopsy, imaging technique with 2D and 3D real time sonography as well as genetic mutational analysis. Prenatal genetic counseling is very essential in her case because of the serious implications to consider for her offspring. The complex management of our patient can be best achieved by using a multidisciplinary approach characterized by strong communication, both among the medical team and with the family. Key words: Harlequin ichtyosis, congenital, prenatal diagnosis, ultrasound.

2820T

Prenatal presentation of fetal anemia associated with εγδβ -thalassemia: two new cases. *E. Gagne[1], Y. Sabr[1, 2], M.F. Delisle[1, 2], N. Kent[1, 2], T. Nelson[1, 3], L.A. Brown[1, 3], K. Schlade-Bartusiak[1, 3], P.C. Rogers[1, 4], S. Pritchard[1, 4], L.D. Wadsworth[1,3], N. Au[1,3].* 1) Children and Women's Health Centre of British Columbia, Vancouver, BC, Canada; 2) Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada; 3) Department of Pathology and Laboratory Medicine, University of British columbia, BC, Canada; 4) Department of Paediatrics, University of British Columbia, Vancouver, BC, Canada.

εγδβ-thalassemia is rare hematological disease caused by either a deletion of the beta-globin gene cluster, or deletion of the upstream β locus control region (LCR). It results in a phenotype of beta-thalassemia trait in adulthood. In the fetal and newborn period, the phenotype can range from self-limiting anemia to life threatening anemia requiring intra-uterine transfusions. Of the 30 εγδβ-thalassemia deletions reported so far, only 4 have been associated with severe congenital anemia requiring intervention or intrauterine transfusion. Here, we report the clinical course and molecular characterization of two new cases. Case 1- A 26 year G1P0, Filipino (Southeast Asian) who presented at 19 weeks and 4 days of gestation with polyhydramnios, fetal pericardial effusion and thickened myocardium. Fetal karyotype was normal. Cordocentesis showed a fetal hemoglobin of 30g/L, which required two intra-uterine transfusions. Delivery was induced at 36 weeks of gestation. Cord blood hemoglobin was 150g/L. Hemoglobin levels improved slowly (Hb at 9 months 96g/L) and the patient did not require further blood transfusion. Alpha and Beta thalassemia molecular genetic testing was normal. Microarray analysis (Affymetrix Cytoscan HD) revealed a deletion of 54 kb encompassing the β LCR, HBE, HBG1 and HBG2 genes. The deletion was confirmed by FISH and shown to be maternally inherited. Case 2- A 42 years G2T1L1, Caucasian (Irish/Scottish) who presented at 33 weeks and 3 days of gestation with polyhydramnios, umbilical cord cyst, thick placenta (7.8cm) and elevated MCA peak systolic velocity (94cm/sec). Patient was delivered by cesarean section at 33 weeks and 5 days gestation because of fetal hydrops. Cord blood hemoglobin was 49g/L, and multiple blood transfusions were needed over a 6 months period. Microarray analysis showed a deletion of 771 kb of the entire beta-globin gene cluster. The deletion was shown to be inherited from the father who had no prior history of anemia. Detailed family history however revealed a paternal aunt previously diagnosed with εγδβ-thalassemia 24 years ago and was reported as the >185 kb Canadian deletion (Diaz-Chico, 1988). In conclusion, we report the first εγδβ-thalassemia deletion diagnosed in a patient of South Asian ethnicity and present evidence of the inter-individual variation in the clinical presentation between carriers of the same deletion. In addition, we show that microarray can be a valuable tool to diagnose these deletions.

2821W

Aicardi-Goutières syndrome carrier screening in Ashkenazi Jewish families.
C. Jalas[1], A. Fedick[2,3], C. Landau[1], C. Halberstam[1], A. Shaag[4], W.K. Chung[5], J.G. Pappas[6,7], E.S. Moran[6], J. Roberts[8], N.R. Treff[2,3], M. Szynkiewicz[9], G. Rice[9], Y.J. Crow[9], O. Elpeleg[4], S. Edvardson[4]. 1) Bonei Olam, Center for Rare Jewish Genetic Disorders, Brooklyn, NY; 2) Department of Microbiology and Molecular Genetics, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ; 3) Reproductive Medicine Associates of New Jersey, Basking Ridge, NJ; 4) Monique and Jacques Roboh Department of Genetic Research, Hadassah, Hebrew University Medical Center, Jerusalem, Israel; 5) Department of Pediatrics, Columbia University Medical Center, New York, NY; 6) Center for Children, NYU Hospital for Joint Diseases, New York, NY; 7) Clinical Genetic Services, New York University School of Medicine, New York, NY; 8) Maimonides Medical Center, Brooklyn, NY; 9) Manchester Academic Health Science Centre, University of Manchester, Genetic Medicine, Manchester, UK.

Aicardi-Goutières syndrome (AGS) is a genetically determined disorder, most particularly affecting the brain and the skin. In its most characteristic form, AGS is a remarkable mimic of the sequelae of congenital infection. In such cases, the prominent clinical features consist of subacute onset of a severe encephalopathy, often after a period of apparently normal initial development, characterized by irritability, psychomotor regression, and slowing of head growth. The initial period is followed by apparent stabilization during which appendicular spasticity, truncal hypotonia and dystonic posturing are prominent along with severe cognitive deficits. Imaging is an important part of diagnosis, typically showing calcification of the basal ganglia, particularly the putamen, globus pallidus, and thalamus. White matter calcifications may also be seen. Other common imaging abnormalities consist of dysmyelination, with a predilection for the frontotemporal regions and cerebral atrophy. Laboratory abnormalities observed in most patients are chronic cerebrospinal fluid (CSF) lymphocytosis and increased CSF interferon alpha (IFN-α). Increased CSF neopterin, a marker of cellular immune system activation, has been shown to be prominent in a number of patients. Although the majority of recognized patients conform to the relatively stereotyped 'classical' phenotype just described, there is also now an extensive literature reporting a broader spectrum of disease presentation, progression and outcome encompassing a variety of apparently non-specific 'inflammatory' features. Consequently, it is likely that AGS is underdiagnosed. AGS is caused by mutations in at least six genes, with the SAM domain and HD domain-containing protein 1 (SAMHD1) gene accounting for 10 percent of mutations in an unselected population. We have identified a total of five mutations in SAMHD1 that cause AGS in thirteen Ashkenazi patients among our cohort and published cases. The most common mutation is a deletion of exon 1, which has a carrier frequency of 1/138 derived from genotyping over 1000 healthy Ashkenazim. Based on these findings, and considering the severity of the condition, we recommend including the SAMHD1 mutations on the expanded carrier screening panels for members of the Ashkenazi Jewish population.

2822T

Quantification of total and fetal cell-free DNA in patients with abnormal Maternal Serum Screen (MSS) parameters.
I. Manokhina[1, 2], T.K. Singh[2, 3], M. Peñaherrera[1, 2], W.P. Robinson[1, 2]. 1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Child & Family Research Institute, Vancouver, BC, Canada; 3) Faculty of Biology, University of British Columbia, Vancouver, BC, Canada.

Determining the relationship between total/fetal cell-free DNA (cfDNA) in maternal plasma and gestational characteristics may help in early detection of fetal abnormalities and the prediction of pregnancy complications. For this purpose 27 plasma samples between 14 and 33 weeks gestational age were collected from women with normal (N=8) or abnormal maternal serum screen parameters (N=19; 9 with low PAPPA and 10 with abnormal hCG, AFP or inhibin A). Placenta and pregnancy outcome data were collected at term: 7/19 women from the abnormal MSS group developed pregnancy complications (1 early-onset preeclampsia, 3 intrauterine growth restriction, 1 placental pathology associated with low birth weight, 1 preterm labour/chorioamnionitis and one trisomy 21). Fetal and total cfDNA was quantified using TaqMan qPCR assays specific to the *SRY* (Chr Y) as these were all male pregnancies and *RPP30* (Chr 10) genomic regions. The mean values of total and fetal cfDNA in the whole population were 2426 genome equivalents (GE)/mL, (range: 357-9767 GE/mL) and 149 GE/mL (range: 18-739 GE/mL), respectively. We investigated possible associations of quantified levels of total and fetal cfDNA with gestational age and pregnancy characteristics, such as abnormal maternal serum screen results, conditions associated with placental insufficiency (preeclampsia, intrauterine growth restriction), and pathological outcomes in total. No significant association between gestational age and the levels of cfDNA was detected, which could be attributed to small sample size. No significant association was detected between total or fetal cfDNA levels in the control and abnormal MSS groups. However, the levels of fetal cfDNA were higher in pregnancies complicated by placental insufficiency (N=4; mean: 321 GE/mL, range: 57-738 GE/mL) relative to uncomplicated pregnancies (N=23; mean: 118 GE/mL, range: 18-488 GE/mL) but this trend did not reach significance (ANOVA, p=0.062). Additionally, total cfDNA was significantly higher in cases with any pathological outcome (N=7; mean: 5691 GE/mL, range: 1372-9767 GE/mL) compared to pregnancies with normal outcomes (mean: 1223 GE/mL, range: 356-2498 GE/mL) (ANOVA, p <0.0001). No significant associations with the available MSS parameters (AFP, PAPP-A, hCG) were detected. We are currently expanding the sample and comparing several approaches for cfDNA quantification to further evaluate these relationships.

2823W

***De novo* mutations in embryonic development and early lethality.**
A. Hoischen[1], C. Serra-Juhé[1,2], C. Gilissen[1], M. Steehouwer[1], J.A. Veltman[1], L.A. Pérez-Jurado[2]. 1) Department of Human Genetics, Radboud University Medical Centre, Nijmegen, Netherlands; 2) Unitat de Genètica, Universitat Pompeu Fabra & CIBER de enfermedades raras (CIBERER), Barcelona, Spain.

We and others have recently shown the importance of de novo mutations for rare [1-3] and common [4-6] sporadic disorders by the direct detection of de novo mutations by exome sequencing. This has revolutionized the genetics of congenital malformations, intellectual disability, and clinically defined syndromes. There are however several reasons why (Mendelian) disease genes are not always easily detected, one reason can be early lethality. Pathologic pregnancies are one of the major health burdens of our society, as just 25% of all conceptions lead to a liveborn child. While the role of (de novo) chromosomal aberrations as a cause of pathologic pregnancies has been intensively studied over the last decades, the role of (de novo) point mutations remains largely unexplored, leaving a large proportion of pregnancy losses unexplained. In this study we have selected cases of pregnancy loss exhibiting well-defined malformations (14 fetuses with hypoplastic left heart (HLH), 14 fetuses with neural tube defect and Arnold-Chiari malformation and 10 fetuses with multiple malformations) in order to identify (de novo) mutations that cause embryonic lethality. In a first analysis we e.g. identified de novo mutation in *USP32* and *NCAPD3*, and subsequently private variants in the same gene or pathway in additional cases, allowing to speculate about the role of these genes in human heart development. The systematic study of the role of point mutations in pathologic pregnancies will allow identification of genes with functions crucial for early development in humans, such mutations would remain undiscovered when only studying liveborn patients. In addition this will improve diagnostics and counseling for affected families in the future. References: 1.Hoischen A et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. Nat Genet. 2011 Jun 26;43:729-31. 2.Hoischen A et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet. 2010 Jun;42:483-5. 3.Ng SB et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010; 42: 790-793. 4.Vissers LE et al. A de novo paradigm for mental retardation. Nat Genet 2010; 42: 1109-1112. 5.O'Roak BJ et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 2011; 43: 585-589. 6.de Ligt J et al. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability.N Engl J Med. 2012 Oct 3. *AH and CSJ contributed equally.

2824T

Non Invasive Prenatal Screening - Are we providing a false sense of security? *S. Klugman[1], B. Suskin[1], K. Erskine[1], N. Kirshenbaum[2], S. Dolan[1].* 1) Reproductive Genetics, Montefiore Med Ctr, Bronx, NY; 2) Maternal-Fetal Medicine, Montefiore Med Ctr Bronx, NY.

Background: HIstorically, a patient who has had a high risk for aneuploidy after first trimester screening was offered invasive prenatal testing via chorionic villus sampling or amniocentesis. In the past year, non invasive prenatal screening has been offered by some in lieu of diagnostic testing. Currently, four companies offer this testing. Case: A 37 year old woman had a first trimester screen at an outside clinic with an increased risk for DS of 1/5 and an increased risk for trisomy 13/18 of 1/25. She was counseled by her primary obstetrics provider and chose to do non invasive prenatal screening. The results were low risk for trisomies 21, 18 and 13. The patient reports she was relieved at that time and chose to continue the pregnancy. The patient was referred to our facility at 19 weeks gestational age for amniocentesis after an anatomy scan showed fetal hydrops and a hypoplastic left heart. The patient was confused because she was told, 'everything was good on the new special blood test.' Amniocentesis revealed a 45XO karyotype. Discussion: This case highlights the limitations of non-invasive prenatal screening. It also reiterates the importance of pre-test counseling and post -test counseling when using non invasive prenatal screening as outlined in the recent ACOG/SMFM and ACMG guidelines. Non-invasive prenatal screening tests for a limited number of chromosomal aneuploidies, and they differ by platform. In this case specifically, choosing a test that included sex chromosome aneuploidy would have been preferred. In addition, this case highlights the fact that in those patients with a positive first trimester screen who undergo invasive testing, the resultant karyotype is not always consistent with the abnormal screening result. A full karyotype and/or microarray, i.e. a diagnostic test, is recommended after a positive screen and will provide more information than a non invasive prenatal screen.

2825W

Two unrelated cases of female fragile X carriers with proximal duplication of non-repetitive DNA sequence within the CGG repeat region of the *FMR1* gene. *J. Skeen[1], J. Coppinger[1], J. McCarver[1], R. Cao[2], S. Filipovic-Sadic[2], L. Schnetzler[3], D. Sevilla[3], J. Weisberger[3], A. Hadd[2], S. Nolin[4], R. Zimmerman[3], G. Latham[2].* 1) Pharmacogenomic and CLIA Service, Asuragen, Austin, TX; 2) Research and Technology Development, Asuragen, Austin, TX; 3) GenPath, BioReference Laboratories, Inc, Elmwood Park, NJ; 4) Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

Nearly all individuals with fragile X syndrome have a full expansion mutation of more than 200 CGG trinucleotide repeats in the 5'untranslated region of the *FMR1* gene. Females with fragile X premutations (55-200 CGG repeats) have a risk of transmitting a fragile X full mutation to their offspring. New high resolution PCR methodologies enable precise sizing of the CGG repeat region and reveal non-repeat sequences, such as AGG interruptions. Here, we report two fragile X carrier females with unexpected DNA sequence findings in the CGG repeat region of the *FMR1* gene detected by these novel PCR assays.

Two clinical diagnostic testing laboratories utilizing the same *FMR1* PCR technology independently identified an unusual electrophoretic reading suggesting a non-CGG sequence within the CGG repeat regions of two unrelated, pregnant fragile X carriers. The samples were further analyzed for AGG interruptions and methylation status, and sequenced across the repeat region. Patient #1 is an African American female who underwent fragile X carrier testing and was found to have an *FMR1* allele size consistent with 78 repeats, but with an unusual electrophoretic trace suggesting a segment of non-CGG DNA within the CGG repeat region. Patient #2 is a female of Taiwanese descent identified by fragile X carrier testing to have an *FMR1* allele size consistent with 58 repeats, but also with a pattern consistent with a non-CGG DNA insert in the repeat segment. Sanger sequencing on both maternal DNA samples confirmed the insertion of a unique non-repetitive sequence of 71 nucleotides in Patient #1, and 37 nucleotides in Patient #2. In both cases, the inserted sequence was homologous to a 36 bp sequence within the *FMR1* gene that is directly upstream of the CGG repeat region.

This type of intra-repeat insertion-duplication has not been previously reported in the medical literature to our knowledge. Both patients have reportedly normal health and intelligence, and both are currently pregnant with male fetuses, one of whom has been confirmed to also carry the atypical *FMR1* allele without any expansion. These findings suggest that the increasing use of more sensitive, high resolution PCR technologies will reveal rare genotypes with uncertain clinical consequences, raising new questions and challenges for researchers, clinical providers, and patients.

2826T

Non-invasive prenatal testing (NIPT): Proceed with caution. A case of trisomy 18 mosaicism in a phenotypically normal newborn. *H. Welsh[1,3], H. Ardinger[1,3], M. Begleiter[1,3], L. Zhang[2,3], L. Cooley[2,3].* 1) Department of Pediatrics, Children's Mercy Hospitals and Clinics, Kansas City, MO; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospitals and Clinics, Kansas City, MO; 3) University of Missouri-Kansas City School of Medicine, Kansas City, MO.

Prenatal screening is used to identify fetuses with chromosomal anomalies so parents and their physicians are able to make critical care decisions. The newest screening methods use massively parallel sequencing of circulating cell-free fetal DNA in maternal plasma to detect chromosomal aneuploidy. This non-invasive prenatal screening (NIPT) allows for the detection of chromosome anomalies without the risk of miscarriage. However, as a new front line screening test, a positive NIPT result must be confirmed by conventional methods prior to making major or irreversible medical decisions. We present a 41-year-old woman, who had MaterniT21™ testing at 11 weeks gestation due to advanced maternal age. The test was positive for Trisomy 18. Amniocentesis was performed at 15 weeks gestation. FISH analysis of uncultured amniocytes showed three chromosome 18 centromere signals in 2.5% of nuclei. Chromosome analysis found an extra small marker chromosome in one of 15 colonies. FISH analysis of cultured amniocytes found the extra small marker chromosome in 3 of 21 colonies. Serial sonograms were performed with normal fetal anatomy and growth documented. The female child, born at term, weighed 7 pounds 9 ounces, was 19 inches in length, and was non-dysmorphic. Random chromosome analysis of cord blood, cord, membranes and villi found a normal diploid karyotype in 98%, 65%, 100% and 100%, respectively. Two abnormal cell lines were found: one with the extra small marker chromosome (positive with chromosome 18 centromere probe), and one with complete trisomy 18 (positive for chromosome 18 centromere and 18q21 probes). The small extra marker was present in 2% of blood, and 0% of cord, membranes and villus metaphase cells. Trisomy 18 was present in <1% of blood, 35% of cord, and 0% of membranes and villus metaphase cells. FISH analysis found trisomy 18 in 4% and 8% of blood metaphase cells and nuclei, respectively; 15.5% of cord; 24.5% of membranes and 32% of villus nuclei. This case shows NIPT may have high sensitivity for chromosomal aneuploidy and reinforces NIPT as a screening test. Positive NIPT results must be carefully followed up with gold standard prenatal classic cytogenetic diagnostic methods before critical decisions are made. This new methodology may reveal fetuses with mosaic conditions that previously may have gone undetected.

2827W

Detection of Carriers of Rare and Novel Mutations using Next Generation DNA Sequencing. *V. Greger, C.J. Kennedy, K. Casey, N. Chennagiri, C. Perreault-Micale, S. Hallam.* Good Start Genetics, Cambridge, MA 02139.

Traditional carrier screening assays can identify only the most common mutations within a gene, typically those recommended by the American College of Medical Genetics, the American Congress of Obstetricians and Gynecologists and societies representing the Ashkenazi Jewish (AJ) population. This methodology can yield high detection rates in specific populations such as the AJ, but a targeted and limited carrier screening approach can miss many mutations prevalent in other ethnicities, as well as any novel or rare mutations. Next generation DNA sequencing (NGS) enables deep gene sequencing in a cost-effective manner, and it is not dependent on an ethnicity-derived panel of mutations. In this study we evaluated the spectrum of mutations found in a pan-ethnic population in the following genes: HEXA, ASPA, BLM, G6PC, FANCC, CFTR, BCKDHA, BCKDHB, DLD, CLRN1, PCDH15, SMPD1, MCOLN1, IKBKAP, and ABCC8. We sequenced the exons and exon-intron junctions of these genes in more than 10,000 patients referred for carrier testing by IVF clinics across the US. To date we have identified a total of 87 unique mutations and 449 carriers. AJ founder mutations or other common mutations, such as the Caucasian Tay-Sachs mutation c.1073+1G>A in HEXA, were often found in multiple carriers. However, we also detected a variety of rare variants, such as ASPA c.2T>C (M1T), ASPA c.79G>A (G27R), and SMPD1 c.1280A>G (H427R) that we observed only once. We also identified novel variants that are predicted to be pathogenic because they are expected to disrupt gene or protein function. Such truncating mutations include nonsense mutations, frameshift mutations and mutations at invariant splice site positions. Examples of some of the novel truncating variants we have identified include: CFTR c.580G>T (G194X), CFTR c.1526delG, IKBKAP c.2128C>T (Q710X), ABCC8 c.579+2T>A, G6PC c.975delG, BCKDHA c.288+1G>A, BCKDHB c.634-2A>G, MCOLN1 c.244delC, CLRN1 c.634C>T (Q212X), BLM c.757C>T (Q253X), BLM c.4076+1delG, and FANCC c.705delC. A significant proportion of the disease mutations we found are rare or novel. Most of these mutations would have not been detected by traditional carrier screening, in particular the novel mutations which are only detectable with comprehensive sequencing. NGS has significant advantages over traditional carrier screening and is expected to yield higher detection rates, irrespective of patient ethnicity, resulting in fewer missed carriers.

2828T

Cree Leukoencephalopathy and Cree Encephalitis Carrier Screening: Retrospective Evaluation of a Population-Based Program.
A.M. Laberge[1-3], J. Leclerc-Blain[1,2], G.A. Mitchell[1-3], B. Wilson[4], A. Bearskin[5], V. Gosselin[6], J. Torrie[6], A. Richter[1-3]. 1) Medical Genetics Division, CHU Sainte-Justine Montreal, QC, Canada; 2) CHU Sainte-Justine Research Center, Montreal, QC, Canada; 3) Department of Pediatrics, University of Montreal Montreal, QC, Canada; 4) Dept of Epidemiology and Community Medicine, University of Ottawa, Ottawa, ON, Canada; 5) Eeyou Awaash Foundation, Chisasibi, QC, Canada; 6) Cree Board of Health and Social Services of James Bay, Chisasibi, QC, Canada.

STATEMENT OF PURPOSE: Cree encephalitis (CE) and Cree leukoencephalopathy (CLE), two severe neurodegenerative autosomal recessive diseases, are found in the James Bay Cree communities (Northern Quebec, Canada). A population-based carrier screening program for CE/CLE (CSP) was developed by local health authorities in collaboration with a community family support group (Eeyou Awaash Foundation). Two groups are targeted: high school students (≥14 years) and women of reproductive age and their partners, mostly in prenatal settings. The CSP has not been evaluated since its start in 2006. OBJECTIVE: To describe screening outcomes using existing data collected by the CSP. METHODOLOGY: We obtained ethics approval to use data collected for management purposes by the CSP (12/2006-04/2013). Available variables include demographic information, carrier screening decision and carrier screening status. We determined uptake of screening and prenatal diagnosis, and carrier rates. Data from school-based and prenatal/preconception screening were analyzed separately. RESULTS: 1340 students were offered screening in school-based CSP. Of those, 650 (48.5%) were screened. Complete data are currently available for 454 screened students. Most had no known family history of CE (98.5%) or CLE (95.8%). 1.2% were screened in the context of an ongoing pregnancy. 22 students were identified as carriers for CE, 48 for CLE, and one for both. For CLE, one potentially at-risk couple was identified; there was no prenatal diagnosis. In the prenatal/preconception setting, 379 individuals were offered screening. Most had no known previous family history of CE (94.7%) or CLE (80.2%). 326 (86.0%) individuals chose carrier screening for CE and/or CLE; 115 (69.3%) women had an ongoing pregnancy. 46 individuals were identified as CLE carriers and 15 as CE carriers, 10 were carriers for both. Of 8 at-risk couples with an ongoing pregnancy, 6 chose prenatal diagnosis. We estimate the population-based carrier rate as 1/11 for CLE and 1/17 for CE, in contrast to incidence-based carrier rates previously calculated at 1/10 and 1/30, respectively. CONCLUSION: High uptake of carrier screening in target populations confirms its acceptability in the communities, as well as individual interest in knowing their carrier status. We plan to evaluate the CSP prospectively to assess screening outcomes and participant knowledge and satisfaction.

2829W

Y-chromosome partial deletions and male infertility in Indian sub-continent.
S. Rajender[1], D. Jaiswal[2], K. Kumar[3], D.S. Rani[4], K. Singh[2], R. Dada[3], K. Thangaraj[4]. 1) Endocrinology, Central Drug Research Institute, Lucknow, UP, India; 2) Banaras Hindu University (BHU), Varanasi, India; 3) All India Institute of Medical Sciences (AIIMS), New Delhi, India; 4) Centre for Cellular and Molecular Biology (CSIR), Hyderabad, India.

Complete AZFc deletions are a proven risk factor for male infertility; however, partial deletions are now emerging as risk factors. No concrete data on AZFc partial deletions in Indian populations are available. We have analyzed 2184 individuals in three cases-control groups from different corners of the Indian sub-continent. Kolkata cohort included 775 infertile and 287 normozoospermic fertile men from Kolkata and adjacent regions. Uttar Pradesh cohort included 547 infertile and 300 fertile control subjects from Lucknow, Varanasi, and adjoining areas. Delhi cohort included 200 cases and 75 control individuals recruited from Delhi and adjacent areas. All subjects were analyzed for complete and partial deletions in the AZFc region using STS and SNV markers recommended by the European Academy of Andrology (EAA). This was followed by meta-analyses on gr/gr and b2/b3 deletions. Classical Y-deletions were observed in a few individuals only. Microdeletions involving the entire AZFc region (b2/b4 recombination) were observed in all cohorts and their presence increased the risk of infertility. Most interesting observation concerned gr/gr deletions, which were observed in all cohorts at comparable frequency and increased the risk of infertility. B1/b3 partial deletions were less common and did not affect infertility risk significantly. Interestingly, b2/b3 deletions were absent in most of our cohorts and were not a risk factor for infertility. In conclusion, partial deletions in the AZFc region are risk factors for male infertility in Indian populations. The data from these cohorts suggest that gr/gr partial deletions are a strong risk factor and that b2/b3 deletions, which are very common and increase infertility risk in Chinese and few other populations, are almost completely absent in Indian populations. Population specific differences in the existence of AZF partial deletions and associated infertility risk are evident from comparison of our data with other populations, particularly Chinese/East Asians. Meta-analysis on partial deletion data suggested that gr/gr deletions are important in infertility and may be considered for inclusion in clinical testing of infertile individuals. On the other hand, meta-analysis on b2/b3 deletions found them to be of least significance for testing in a clinical setting.

2830T

MTHFR C677T POLYMORPHISM IN MEXICAN PATIENTS WITH POLYCYSTIC OVARY SYNDROME.
I.M. Salazar-Dávalos[1], N. Suárez-Magaña[2], J.P. Mena[3], M. Salazar-Páramo[4], E. Chávez[4], M.A. Aceves-Aceves[1], N.O. Dávalos[1], M.G. González-Mercado[3], F. Grover-Páez[4], I.P. Dávalos[3]. 1) Instituto de Genética Humana, DGH, Facultad de Medicina, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) UMAE, HGO, IMSS, Guadalajara, México; 3) Doctorado Genética Humana, CUCS, Universidad de Guadalajara;CIBO-IMSS, Guadalajara,México; 4) Depto. Fisiología, CUCS, Universidad de Guadalajara; Div. Investigación, UMAE, HE, CMNO, IMSS, Guadalajara, México.

Introduction: Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in reproductive age patients and one cause of anovulatory infertility. There are findings of high homocysteine levels in PCOS, the C677T polymorphism of the MTHFR gene have been associated with hyperhomocysteinemia and decrease of folic acid. Objective: To determine the association between MTHFR C677T polymorphism in mexican patients with polycystic ovary syndrome. Material and methods: We included 64 patients with PCOS diagnosis according to Rotterdam criteria and 101 Mexican mestizos as reference group (M). The genotyping was performed by PCR/RFLP technique. The M group presented Hardy-Weinberg equilibrium. Results: In the PCOS group (n=64)the MTHFR C677T genotypic frequencies %(n)were distributed as follows: CC 33 %(21), CT 48 %(31) y TT 19 %(12). The allelic frequency of the C allele was 57 %(73) and T was 43 %(55). The GF in the reference group (n=101) were CC 31 %(31), CT 50 %(51), TT 19% (19) and AF for C allele 56 %(113) and for T allele 44 % (89). The AF comparison between both groups were no statistically different ($p>0.05$). Conclusion: The genotypic and allelic distribution between both groups (group-PCOS vs group-M) were similar. This study showed no association between MTHFR C677T polymorphism and PCOS.

2831W

Gene-Environment Interaction in Adverse Reproductive Outcomes.
B.Dev. Banerjee. Biochemistry, University College of Medical Sciences, Delhi, India.

Many human disorders result from a complex interaction between an individual's genetic make-up and environmental stressors Humans are exposed to numerous xenobiotics constantly and unavoidably such as pesticides, metals, PCBs etc. Endocrine disruption, genetic predisposition, altered immune surveillance, inflammation and subsequent oxidative stress may antedate adverse reproductive outcomes and contribute to its pathogenesis. Although environmental factors are important, genetics clearly plays a role in adverse reproductive outcomes. Identification of genetic susceptibility variants will lead to better understanding of the role of variable factors in adverse reproductive outcomes. It has been observed that a lot of women with genetic polymorphism do experience normal delivery while some do not. It can be hypothesized that genetic polymorphism requires the presence of certain environmental stimuli to have consequences of clinical significance. The recent abundance of epidemiologic research examining associations between polymorphic genes that code for enzymes involved in xenobiotic biotransformation and disease has on occasion generated interesting findings. Recent studies from our laboratory clearly showed the importance to assess the role of variations in the human genome (polymorphisms) in modifying the effect of exposures to xenobiotics to define 'Gene-Environment Interaction', which render some individuals or groups in the population more or less likely to develop adverse health effect. Current and future efforts to identify new polymorphisms in genes involved in environmental response with larger sample size will broaden the scope of potential genetic effect modifiers. Currently, our laboratory is involved in studying the role of 'Gene-Environment Interaction' with reference to xenobiotic metabolism and oxidative stress related genes in various diseases and we have reported the association of xenobiotics with many of adverse reproductive outcomes such as preterm birth, intrauterine growth retardation, recurrent miscarriage, etc. Our effort in this area may also lead to the development of possible biomarker(s) to screen individuals, exposed to xenobiotics and preventive measures for safe reproductive outcomes. Determining the effect of these polymorphisms along with xenobiotics burden will be of paramount importance in an early diagnostic strategy and preventive measures for adverse reproductive outcomes with reference to environmental toxins.

2832T

Trans-ethnic GWAS of pelvic organ prolapse among African American and Hispanic post-menopausal women of the Women's Health Initiative.
D.R. Velez Edwards[1, 2], R.M. Ward[3], A. Giri[4], K.E. Hartmann[1], A.J. Park[5], C.L. Avery[6], R. Wallace[7], G. Bareh[8], L. Qi[9], M.J. O'Sullivan[10], A. Reiner[11], T.L. Edwards[2,4], .JM. Wu[12]. 1) Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Department of Obstetrics and Gynecology, Division of Female Pelvic Medicine and Pelvic Reconstructive Surgery, Vanderbilt University Medical Center, Nashville, TN; 4) Institute for Medicine and Public Health, Division of Epidemiology, Department of Medicine, Nashville, TN; Vanderbilt University; 5) Department of Obstetrics and Gynecology, and Urology Georgetown University School of Medicine Staff, Female Pelvic Medicine & Reconstructive Surgery Medstar Washington Hospital Center; 6) Department of Epidemiology, University of North Carolina, Chapel Hill NC; 7) Department of Epidemiology, University of Iowa, Iowa City, IA; 8) University of Texas Health Science Center, San Antonio, TX; 9) Department of Public Health Sciences, University of California Davis, Davis, CA; 10) Department of Obstetrics and Gynecology, University of Miami, Miami, FL; 11) Department of Epidemiology, University of Washington, Seattle, WA; 12) Department of Obstetrics and Gynecology, Division of Urogynecology, University of North Carolina-Chapel Hill, NC.

Pelvic organ prolapse (POP) impacts 40% of post-menopausal women and has a profound negative impact on a woman's quality of life. Despite POP being common and having acute effects, the pathophysiology is not well understood. As much as 40% of POP risk may be attributed to genetic factors. We conducted a genome-wide association study (GWAS) of POP in post-menopausal African-American (AA) and Hispanic (HP) women from the Women's Health Initiative (WHI) SNP Health Associated Resource (SHARe) GWAS. This is the first GWAS of POP in minority populations. WHI POP outcomes were obtained from pelvic exams. WHI data included 1427 participants with POP (stage 1 through 3) and 1274 controls (stage 0). After imputing to the most recent 1000 Genomes reference haplotypes, race stratified single SNP logistic regression on POP outcomes (for POP stage 0 versus 1-3 and stage 0 versus 2&3, respectively) was performed adjusting for continuous axes of ancestry and known POP risk factors (age, parity, BMI [AA] or WHR [HP]). We used fixed effects models to meta-analyze across AA and HP cohorts. We identified one novel locus at chromosome 4q32.3 (carboxypeptidase E [CPE] rs28573326, OR = 2.36, 95% CI 1.78 to 3.15, $p = 3.5\times10$-9) in AAs using POP stage 0 versus 2&3 and another at chromosome 4q34.3 when AA and HP data were meta-analyzed (intergenic, rs113518633, OR = 2.41, SE = 0.16, pHet = 0.722, $p = 2.19\times10$-8) using POP stage 0 versus 1-3. Both associations exceeded genome-wide significance and neither has previously associated with POP. No genome-wide significant associations were observed in HPs, however, we found some suggestive associations at chromosome 11q23.3 (myelin protein zero-like 2 [MPZL2], rs3018366, OR = 0.37, 95% CI 0.25 to 0.55, $p = 4.68\times10$-7) using POP stage 0 versus 1-3. Although neither SNP was previously associated with POP, the CPE SNP associated in the same direction in both AA and HP and the gene has been associated with risk for rectal prolapse in prior studies. These results provide the first evidence of GWAS-associated SNPs with POP in minority populations. We are also evaluating POP in European Americans and AAs participants from BioVU, the Vanderbilt electronic medical record DNA biorepository.

2833W

The Genetic Predisposition for Uterine Leiomyomas in Recently Admixed Populations: A Preliminary Study in Individuals From Electronic Medical Records. *J. Jeff[1], G. Belbin[2], D. Ruderfer[2,3], E. Stahl[2,3], S. Purcell[2,3], E. Bottinger[1], R. Loos[1,4], O. Gottesman[1], E. Kenny[1,2].* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Institute for Genomics and Muliscale Biology, Department of Genetics and Genomics, Ichan School of Medicine at Mount Sinai, New York, NY; 3) Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai; 4) The Mindich Child Health and Development Institute at The Icahn School of Medicine at Mount Sinai.

Uterine leiomyomas (UL) are common benign tumors of the uterus and are the leading cause of hysterectomies. UL affects 25-77% of women and women of African descent have 2-3 fold increased risk. In addition to African ancestry, risk factors for UL include increased age, obesity, endogenous hormonal factors, and family history. Heritability studies in Europeans suggest 26-79% of UL can be explained by genetic factors. Despite the high heritability, established common genetic variants have yet to be identified. Identifying common genetic risk factors for UL in particular has been a challenge. First, obtaining a large population of UL cases that have been clinically diagnosis has been difficult to achieve in survey based cohort studies. Furthermore, there are few genetic studies that have been performed in UL in minority populations. We aim to address these limitations by using a diverse clinical population to identify and characterize the genetic predisposition of UL in African Americans (AAs) and Hispanics (HAs). Here, we report preliminary findings from a association study using Illumina OmniExpress and Human Exome BeadChip arrays in AAs and HAs from BioMe Biobank Program at Icahn Medical School at Mount Sinai, a clinical care cohort of consented patients with genotype data linkable to their electronic medical records (EMR). Using medical billing codes as a broad preliminary definition of case status, we identified a total of 325 cases and 650 controls from BioMe that were self-reported AA and 306 cases and 606 controls that were self-reported HA. We tested 844,372 common SNPs for an association with UL. All analyses were adjusted for age, body mass index, and informative principle components. While no SNP reached genome-wide significance at the discovery stage, we identified four loci at p<5.0E-6: AASDHPPT on chr11 (OR=1.6, p=1.4E-6) and an intergenic region on chr4 (OR=1.7, p=4.1E-6) in AAs as well as GAB4 on chr22 (OR=1.9, p=3.0E-6) and an intergenic region on chr2 (OR=0.51, p=2.5E-6) in HAs. We then meta-analyzed the results of AAs and HAs and detected six associations at p<5.0E-6. Notably, one locus harbors a tumor suppressor gene, FOXP1 (p=2.9E-6) and trends towards significance in both AAs and HAs (p=1.2E-4 in AAs, p=2.4E-3 in HAs). To expand this limited study, we are expanding the BioMe cohort and developing a robust algorithm to extract UL cases and controls from the EMRs of BioMe participants.

2834T

Role of Shigella in etiology of Endometriosis. *M. Latha[1,2,3], V. Kutala[1], Q. Hasan[3], V. Kodati[2].* 1) Biochemistry, CCMB-NIMS, Hyderabad, Hyderabad, India; 2) Genetics and Molecular medicine,Vasavi hospital , Hyderabad, Hyderabad, India; 3) Genetics and Molecular medicine,Hyderabad Science society,Hyderabad,India.

Endometriosis is defined as a disease characterized by the presence of endometrial cells, gland and stroma at ectopic sites outside the uterine cavity in addition to their normal presence as the innermost lining of the uterus. This is responsible for menstrual related problems, infertility and morbidity in women of reproductive age. There are different etiological factors implicated in the pathogenesis of endometriosis like mechanical, hormonal, environmental, immunological and genetic. In addition, our group has demonstrated the role infection in pathophysiology of endometriosis for the first time particularly identifying the novel DNA sequence obtained from ectopic endometrium which is having 96% homology to the Shigella sps genomic sequence.Present study has demonstrated that the mRNA was isolated from ectopic and eutopic tissues of three endometriosis cases and Differential Display RT- PCR (DD-RT-PCR) revealed sequences which were analyzed using the BLAST search .A 60/65 bp, 96% homology with Shigella dysenteriae, flexineri and sonnei species. (Figure- 13, 17) was obtained.Hence "bacterial hypothesis" provides a novel explanation of the molecular etiopathogenesis of endometriosis.

2835W

Identification of differentially expressed non-peptide metabolomic molecules with metabolomics approach in pregnancy-induced hypertension syndrome. *X-C. Luo[1,2], J. Pan[1,2], X-G. Tao[2,3], X-L. Zhao[2,3], Y. Gu[1,2], Q-X. Shi[1,2], N. Zhong[1,2,3,4].* 1) Lianyungang Maternal and Children's Hospital, Lianyungang, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, Lianyungang, China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in DD, Staten Island, NY.

Pregnancy-induced hypertension syndrome (PIH) is the most common medical condition encountered during pregnancy and is reported to affect up to 9.4% ~10.4% of all gestations. PIH is defined as a systolic of 140mmHg or greater or a diastolic of 90mmHg or greater. Earlier studies showed that several risk factors have been described as predisposed to PIH, however, the pathophysiology of the PIH is not clear yet. The most well accepted hypothesis is that placental ischemia/hypoxia results from inadequate utero-placental vascular remodeling, which leads to a decrease in placental blood flow and the most of the data reported earlier were related to large peptide molecule metabolism. In this study, we have applied a metabolomics approach with HPLC-MS to investigate the differential metabolomic molecules presented in HIP. A total number of 537 molecule peaks (fractions) have been detected and computerized at the p<0.01 value. Among which, 84 were identified to be shared by all PIH1 (BP:140-150/95-100), PIH2 (BP: 160-175/105-125), and PIH3 (BP: >180/100) groups, 10 were overlapped between PIH1 and PIH2, 123 between PIH2 and PIH3, and 6 between PIH1 and PIH3. There were 113 peaks were down regulated and 37 up regulated when the group of control was compared against that of PIH1, 259 down regulated and 8 up regulated against PIH2, and 311 down regulated and 16 up regulated against PIH3. Two molecules, one (Feature 897.2/466) from the subgroup PIH-1 that was detected as up regulated and another (Feature 736.1/1870) that was down regulated, from PIH-3 showed excellent separation one from another between the PIH and control groups with a relative narrow intragroup variation within the two comparison groups. These two molecules could be potentially considered as bio-signature candidates for further investigations. Two molecules, Feature 894.2/467 and Feature 418.2/1345, showed a good separation, both intergroup and intragroup, and the remaining 5 showed a fair separation. Our findings justified a prospective assessment of metabolomic technology as a screening tool for PIH, and may lead to an improved method in future to monitor PIH including preeclampsia using non-peptide biochemical markers.

2836T

Admixture mapping study of uterine fibroids finds evidence for fibronectin and diabetes genes. *K.S. Tsosie[1], D.R. Velez Edwards[1,2], T.L. Edwards[1,3].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 3) Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, TN.

Uterine fibroids are benign tumorous growths that affect up to 77% of women by menopause and account for $9.4 billion in yearly healthcare costs in the U.S. Fibroids are highly heritable, with up to 69% of risk attributable to genetic factors, as estimated by family history and supported by racial differences in prevalence and severity. Compared to European Americans (EA), African American (AA) women have earlier age-of-onset, more numerous and larger fibroids with a higher lifetime incidence. Here we further evaluated the racial disparities in fibroid risk through admixture mapping. Utilizing the Vanderbilt BioVU DNA databank, which links adult DNA samples with electronic medical records, 1,237 AA and EA subjects (185 cases and 1,022 controls) were identified whose fibroid status were confirmed with pelvic imaging. Admixture mapping was performed among AAs using GWAS data and identified regions were further evaluated in both EAs and AAs for association with fibroid risk. Local ancestry was estimated using LAMP-ANC, and logistic regression models relating local ancestry to fibroids were adjusted for principal components of ancestry. Analyses revealed associations that link uterine fibroid risk and African ancestry on chromosome 3q26 at two novel loci, TRAF2 and NCK-interacting protein kinase (*TNIK*; rs13315469, odds ratio [OR] = 0.41, 95% confidence interval [CI] 0.27, 0.62; $p = 2.6 \times 10^{-5}$) and solute carrier family 2, facilitated glucose transporter member 2 (*SLC2A2*; rs11924032, OR = 0.42, 95% CI 0.28, 0.64; $p = 3.6 \times 10^{-5}$). *TNIK* has an essential role in the Wnt signaling pathway but also phosphorylates *TCF7L2*, which modulates blood glucose homeostasis and is associated with type 2 diabetes. *SLC2A2* has also been implicated as a susceptibility gene for noninsulin-dependent diabetes mellitus. Of note, complex interactions between insulin, insulin-like growth factors (*IGF-I*), estrogen, and progestogen pathways have been shown to lead to fibroid growth. Diabetes has been previously associated with protection from fibroids and has a racial disparity in prevalence, supporting a potential role for *SLC2A2* in fibroid disparity. When models were further adjusted for age, an association was observed in the gene fibronectin type III domain containing 3B (*FNDC3B*; rs6763884, OR = 0.42, 95% CI 0.28, 0.64; $p = 5.6 \times 10^{-5}$). Previous studies have already implicated increased expression of fibronectin in uterine fibroids and *FNDC3B* has a known role in oncogenesis.

2837W

The Kallmann Syndrome gene WDR11 and Binding Partner EMX1 Are Expressed in Hypothalamic and Gonadal Tissues. *EK. Ko[1], SD. Quaynor[1], LP. Chorich[1], HG. Kim[1], SH. Kim[2], RS. Cameron[1], LC. Layman[1].* 1) Institute of Molecular Medicine & Genetics, Georgia Regents University , Augusta , GA 30912; 2) Division of Basic Medical Sciences, St. George's Medical School, University of London, London SW17 0RE United Kingdom.

The hypothalamic-pituitary-gonadal (HPG) axis is regulated by pulsatile secretion of gonadotropin releasing hormone (GnRH). When HPG dysfunction occurs at the level of the hypothalamus, normosmic hypogonadotropic hypogonadism (nHH) results. Patients with IHH manifest absent puberty with low serum levels of sex steroids and gonadotropins. When sense of smell is impaired in IHH patients, it is termed Kallmann Syndrome (KS). Approximately 40% of nHH/KS patients have mutations in a variety of genes that affect GnRH neuron migration and/or GnRH action. One such gene is *WDR11*, which was identified by positional cloning in a KS patient with a 10;12 chromosome translocation. Approximately 3% of nHH/KS patients had mutations in *WDR11*, which interferes with WDR11's binding to EMX1. EMX1 is a transcription factor known to be expressed in the developing forebrain, olfactory sensory neurons, branchial arches, and the embryonic kidney~all of which are affected in KS patients. Recently it was reported that some nHH/KS genes (FGFR1 and NR0B1) could have physiologic effects at both the hypothalamus and gonads. WDR11 is a cytoplasmic protein, which was sequestered in the nucleus following leptomycin B treatment, which inhibits nuclear export. The mechanism of how WDR11 causes KS is not known, but we hypothesize that WDR11 shuttles from the cytoplasm to nucleus, binds with EMX1, and affects expression of downstream genes in GnRH neurons. Therefore, we first wanted to determine if mRNA expression of *Wdr11, Emx1*, and related gene Emx2 were present in hypothalamic GnRH neurons and gonads. RNA was extracted from mouse GnRH neurons (migratory NLT and postmigratory GT1-7 cells), human GnRH olfactory neuroblasts (FNCB4-hTERT cells), testes, and ovaries, and then subjected to RT-PCR with confirmation by DNA sequencing. In mouse GnRH neurons, *Wdr11*, *Emx1*, and *Emx2* were expressed. However, in human GnRH olfactory neuroblasts, *WDR11* was expressed, but *EMX1* and *EMX2* were not detected. In mouse testes and ovaries, *Wdr11* and *Emx1* were expressed. Protein analysis is ongoing. These findings indicate that WDR11 and EMX1 are co-expressed in GnRH neurons, testes, and ovaries. Interestingly, *EMX1* and *EMX2* were not detected in human GnRH olfactory neuroblasts that were derived from an earlier embryological stage than the mouse GnRH neuronal cells used. These findings suggest that WDR11 could have reproductive functions at both the hypothalamus and gonads similar to other nHH/KS genes.

2838T

Understanding the genetics of spermatogenic failure by resequencing the sex chromosomes of infertile men. *R. George[1], J. Hughes[1], L. Brown[1], L. Lin[2], D. Koboldt[2], R. Fulton[2], R. Wilson[2], R. Oates[3], S. Silber[4], S. Repping[5], D. Page[1].* 1) Whitehead Institute, Cambridge, MA; 2) The Genome Institute at Washington University, St. Louis, MO; 3) Boston University School of Medicine, Department of Urology, Boston, MA; 4) Infertility Center of St. Louis, St. Louis, MO; 5) Academic Medical Center, Department of Reproductive Medicine, Amsterdam.

The sex chromosomes harbor a large number of genes involved in spermatogenesis and are hemizygous in males. For these reasons, mutations that cause spermatogenic failure~the production of very few or no sperm~are disproportionately likely to be found on the X and Y chromosomes. While several genetic causes of spermatogenic failure, such as microdeletions of the Y chromosome and large-scale cytogenetic abnormalities (e.g. Klinefelter's syndrome; XXY), have been identified, they only account for 20-30% of cases and the majority of genetic causes remain unknown. To identify new mutations involved in spermatogenic failure, we have captured and sequenced the coding and conserved non-coding regions of the X and Y chromosomes in 300 men with nonobstructive azoospermia and 300 controls. Our targeted regions include 838 and 53 protein coding genes from the X and Y chromosomes respectively, along with 22 Mb of non-coding sequence that contain ncRNAs and putative regulatory regions. We will describe the initial analysis of these sequences, and the identification of putatively causal coding mutations and copy number variants.

2839W

Mutations in the Kallikrein related peptidase-3 (KLK3) gene affect semen parameters in Indian men. *N. GUPTA[1], D.V.S. SUDHAKAR[2], S.N. SANKHWAR[3], N. GUPTA[4], K. THANGARAJ[2], S. RAJENDER[1].* 1) ENDOCRINOLOGY, CSIR-CDRI, LUCKNOW, UTTAR PRADESH, INDIA; 2) CENTRE FOR CELLULAR AND MOLECULAR BIOLOGY, HYDERABAD, ANDHRA PRADESH, INDIA; 3) DEPARTMENT OF UROLOGY, KING GEORGE MEDICAL UNIVERSITY, LUCKNOW, UTTAR PRADESH, INDIA; 4) INSTITUTE OF REPRODUCTIVE MEDICINE, KOLKATA, WEST BENGAL, INDIA.

Kallikrein related peptidase-3 (KLK3) is a highly abundant serine protease in the prostatic epithelium and seminal plasma. It has been known to play a role in fragmentation of semenogelins, resulting in the dissolution of semen coagulum and activation of sperm progressive motility. Among male reproductive organs, KLK3 expression was earlier thought to be restricted only to the prostate gland. Recent studies have reported KLK3 expression in adult testis, suggesting its possible participation in the process of spermatogenesis. Genetically compromised activity of KLK3 might alter spermatogenesis, but analysis of KLK3 sequence to identify mutations in infertile men has not been undertaken. Therefore, we have analyzed the coding region of KLK3 in ethnically matched 875 infertile and 290 fertile men. Interestingly, this study identified thirty three substitutions, of which thirteen were novel (not yet available in public databases). None of these substitutions has been studied in correlation with male infertility. Among the thirty three substitutions, nine resulted in non-synonymous changes while rest of them were either in the intronic regions or resulted in synonymous changes. In silico analysis on novel substitutions revealed four to be missense variants, six to be intronic variants, one to be a 5'UTR variant, one to be a 3'UTR variant, and one to be a downstream gene variant with reference to the transcript ENST00000326003. Variant effect prediction analysis on the non-synonymous substitutions identified p.Met139Lys, p.Glu174Lys, and p.Ile179Thr to be deleterious. Statistical comparisons of genotype frequencies showed that six SNPs were differentially distributed between fertile and infertile men. The frequency of genotypes 'CA+AA' (rs266881, OR = 2.92, $P = <0.0001$), 'TC+CC' (rs11573, $P = 0.022$), 'GA+AA' (rs266875, OR= 1.44, $P = 0.016$), and 'TC+CC' (rs174776, OR = 1.91, $P < 0.0001$) were significantly higher in infertile men as compared to fertile controls. However, the genotypes 'TC+CC' at intronic SNPs c.206+235T>C (OR = 0.44, $P = 0.002$) and 'GA+AA' at c.631-74G>A (OR = 0.26, $P = 0.02$) were more frequent in fertile controls as compared to infertile men. In conclusion, our findings suggest that genetic alterations in the KLK3 gene may lead to altered semen parameters and affect male fertility; highlighting the importance of this gene in spermatogenesis/testicular function.

2840T

Expression Quantitative Trait (eQTL) Mapping in Mid-Secretory Phase Endometrial Cells Identifies Candidate Genes for Recurrent Early Pregnancy Loss (REPL). *C.L. Kagan[1], G. Kosova[1], K. Patterson[1], M.D. Stephenson[2], C. Ober[1,2].* 1) Dept of Human Genetics, University of Chicago, Chicago, IL; 2) Dept of Obstetrics and Gynecology, University of Chicago, Chicago, IL.

Fertility traits are heritable in humans; however, little is known about the genes that influence reproductive outcomes or genetic variants that contribute to inter-individual differences in these traits. To address this gap in knowledge, we performed an unbiased genome-wide study to identify functional (regulatory) SNPs in endometrial biopsies collected during the mid-secretory phase, the time at which implantation occurs, in 34 women of European ancestry with a history of at least two documented intrauterine miscarriages of less than 10 weeks of gestation. Gene expression was measured in RNA using Illumina Human HT12v4 arrays; DNA samples were genotyped using the Affymetrix Axiom™ Genome-Wide CEU 1 Array, which contains 674,517 SNPs. To identify *cis* eQTLs, we included SNPs within 200kb of each gene and considered an additive model as implemented in Matrix eQTL. Transcripts from 10,191 genes were detected as expressed in these samples. Out of 532,538 association tests, 1,088 *cis* eQTLs in 308 genes were significant at a false discovery rate (FDR) of <5%. We then genotyped the 80 most significant *cis* eQTLs in 256 women with REPL (not included in the eQTL study) and 232 women with at least one successful pregnancy and no history of infertility or miscarriage (fertile controls), and tested for differences in genotype frequencies between the two groups using the Cochram-Armitage trend. 64 of these SNPs were successfully genotyped using iPlex (Sequenom). Nine SNPs had p-values < .05 (based on permutation) when only 3.2 were expected by chance, reflecting a significant enrichment for small p-values (p = 0.0037). The most significant association was with a SNP (rs56274787; p = 0.0087) that is an eQTL for the *TBCD* gene, which encodes a tubulin-folding protein that promotes epithelial cell detachment. Overall, 9 SNPs showed independent evidence both for being functional in the endometrium (i.e., associated with gene expression) and for being significantly associated with REPL patients compared to fertile controls; This study demonstrates that eQTL mapping in a relevant tissue is a promising strategy for identifying genes and associated alleles that are risk factors for REPL, which affects approximately 5% of couples. The eQTLs identified in our study could also serve as candidates for other reproductive disorders such as primary infertility and preeclampsia. Supported in part by NIH grant HD21244.

2841W

The Clinical Analysis of Relationship between Sperm DNA Damage and Sperm Parameters. *F. Kaplan[1], S. Aydos[1], B. Altinok Zaim[1], I. Yükselen[1], Y. Yükselten[1], A. Sungurolu[1], K. Aydos[2].* 1) Ankara University Faculty of Medicine,Department of Medical Biology Ankara, Turkey; 2) Ankara University Faculty of Medicine, Reproductive Health Research Center Ankara, Turkey.

Purpose:Although fertilization rate by using intracytoplasmic sperm injection (ICSI) is increasing, the probability of fertilization with DNA damaged sperm creates risk for both pregnancy success and also health of the new born.Recently, disputes about whether sperm parameters give an opinion related to sperm DNA damage or not have become more of an issue.The aim of the study is to investigate the relationship between sperm DNA damage and sperm parameters. Methods: 298 infertile men,between 24-38 years,whose sperm motility percentage and sperm counts are known,were included to the study.140 of them had also sperm morphology records. In all couples,female factor was eliminated.DNA damage was studied with Comet assay. Observations were made at magnification 400 X using an epifluorescent microscope.100 cells were analyzed visually from each 3 slide.Each image was classified according to nucleus scale and tail length given a value of 0,1,2,3 or 4(from undamaged class 0 to maximally damaged class 4),so that the total DNA damage score of the slide could be between 0 and 400 arbitrary units (AU).Sperm DNA damage ratios were correlated with sperm motility,sperm count and sperm morphology.Statistical analysis was made by Mann-Whitney u and Kruskal-Wallis tests. Results:Our results showed a significant negative correlation between the sperm counts ≥ 15 million/ml(p<0.001),progressive motility ≥%32(p<0.05),and total sperm DNA damage score. Total sperm DNA damage scores of the samples with sperm counts ≥15,progressive motility ≥%32 were found 185.3±40.08 and sperm counts ≤15,progressive motility ≤%32 were found 222.4±56.88 respectively.According to our findings there was no correlation between Kruger morphology results and sperm DNA damage scores. Discussion:According to our results,there was no relation between sperm morphology and sperm DNA damage.We found significant relation between sperm count,motility,and total sperm DNA damage score.The significant increase of spermatozoa rates with DNA damage in infertile males comes into prominence due to being overcome obstacles by using assisted reproductive technology(ART) in these infertile males.Indeed,in case of natural fertilization, spermatozoa reached the fallopian tube have much healthier DNA when compared to which cannot.The results of our study emphasize the risk of using sperm cell with DNA damage in males having abnormal sperm parameters,because of randomly chosen morphologically normal sperms during IVF/ICSI.

2842T

Quantitative Analysis of Mixtures by Deep Sequencing of HLA Gene Amplicons Using Next Generation Systems. *B.N. Hoglund[1], M. Rastrou[1], D. Goodridge[2], H.A. Erlich[1,3], C.L. Holcomb[1].* 1) Human Genetics, Roche Molecular Systems, Pleasanton, CA; 2) Conexio Genomics, Perth, Australia; 3) Children's Hospital Oakland Research Institute, Oakland, CA.

Purpose: The clonal property of next generation sequencing systems allows for the quantitative analysis of mixed samples by simply counting sequences corresponding to the component alleles. Previously, we developed a method using 454 amplicon sequencing and Conexio software for high resolution and high throughput genotyping of the HLA class I and class II loci. This system was used to analyze the blood of a severe combined immunodeficiency (SCID) child and estimate the proportion of maternal cells by counting HLA-C allelic sequence reads corresponding to the non-transmitted maternal allele. Here we report the development of a system that allows us to quantify HLA allelic mixtures in plasma. Methods: DNA from plasma or contrived mixtures of cell lines was amplified using primers that targeted a short region (~150 bp to amplify small DNA fragments in plasma) of HLA DPB1 or DQB1 exon 2. Amplicons were further amplified by emulsion PCR and sequenced on a 454 GS FLX or GS Junior. Sequences were examined using modified Conexio Assign ATF 454 software. This software allowed for rapid digital analysis of each DPB1 or DQB1 allele. We also analyzed mixtures using the Illumina MiSeq. Results: Using the modified Conexio software, minority HLA alleles in a mixture were readily identified and separated from background 'noise' i.e., sequences generated by PCR or sequencing errors. In mixtures of two heterozygous cell lines, the minority HLA alleles could be detected at 0.5% with 1 ng (~140 diploid genomes) DNA input. In one plasma sample from a pregnant woman in the third trimester, fetal HLA alleles were detected at 11.6% of the total of maternal plus fetal alleles. Conclusion: We have developed a system that allows the sensitive and precise analysis of mixtures using either 454 or Illumina MiSeq sequencing technology. This method could be useful in numerous clinical applications; we have demonstrated the quantification of fetal DNA in maternal plasma during pregnancy by detection of HLA alleles. Such quantification has potential uses in noninvasive prenatal diagnostics.

2843W

Screening models for early detection of late-onset preeclampsia with various markers in low-risk pregnancy population. *H.J. Park, D.H. Cha, S.H. Kim, S.S. Shim, J.Y. Kim, Y.K. Cho, K.J. Lee.* CHA Gangnam Medical Center, CHA University, Seoul, South Korea.

Objective : Our primary objective was to establish a cut-off value for the soluble fms-like tyrosine kinase 1(sFlt-1)/placental growth factor (PlGF) ratio measured using the Elecsys assay to predict late-onset preeclampsiain low risk women. A second objective was to evaluate the ability of combination models that included Elecys data, second trimester uterine artery (UtA) Doppler, and fetoplacental proteins measurements during Down syndrome screening. Methods:A prospective cohort study was carried of 262 women at low risk for preeclampsia. Maternal plasma samples were obtained for measurement of pregnancy-associated plasma protein-A (PAPP-A),alpha-fetoprotein, unconjugated estriol, human chorionic gonadotrophin, inhibin-A, and the sFlt-1/PlGF ratio. All women underwent UtA Doppler at 20-24 weeks of gestation. Results: Of the 262 women, eight (3.0%) developed late-onset preeclampsia. ROC curves demonstrated that the sFlt-1/PlGF ratio in the third trimester yielded the best detection rate (DR) at a fixed false positive rate (FPR) of 10%. This was followed in rank order by the sFlt-1/PlGF ratio in the second trimester, sFlt-1, and then PlGF. Binary logistic regression was used to determine the five best significant combination models for screening. For a FPR of 5 and 10%, the combination of PAPP-A with the sFlt-1/PlGF ratio in the second trimester yielded a DR of 87.5%, identical to that of serial sFlt-1/PlGF ratios for late PE. The combination of maternal BMI and second trimester sFlt-1 yielded a DR of 87.5% at a 10% FPR. Combining PAPP-A with inhibin-A yielded a 50% DR at a 10% FPR. The combination of PAPP-A and the sFlt-1/PlGF ratio in the third trimester yielded a DR of 62.5% for late preeclampsia at a 10% FPR. Conclusion: The combination of the second trimester sFlt-1/PlGF ratio and PAPP-A data or BMI, and the second trimestersFlt-1 can predict late onset preeclampsia more effectively than any single marker alone.

2844T

First live birth in Hong Kong after preimplantation genetic diagnosis on a disease-predisposition mutation carrier with a novel genomic deletion in BRCA2. *Q. Wang[1,2], J. Chow[1,2], W. Yeung[1,2], E. Lau[1,2], V. Lee[1,2], E. Ng[1,2], P.C. Ho[1,2].* 1) Dept. of Obs. & Gyn., The University of Hong Kong, Hong Kong; 2) Dept. of Obs. & Gyn., Queen Mary Hospital, Hong Kong.

Preimplantation genetic diagnosis (PGD) is an option for couples bearing disease-causing mutations to avoid the bearing of an affected child. The use of PGD on disease-predisposition mutation carriers, such as BRCA mutation is controversial. Here we report the first live birth in Hong Kong, after PGD upon a request from a woman with an early onset of breast cancer. She carries a paternally derived mutant allele with an unknown genomic breakpoint. Due to the limited availability of familial data, sperm haplotyping was conducted on the woman's carrier brother, leading to the identification of the mutant haplotype. Embryo biopsy was performed on 8 good-quality embryos on day 3, followed by whole genome amplification and linkage analysis with 4 microsatellite markers and 2 intragenic SNP markers. Among the five unaffected embryos identified, one morula and 1 blastocyst were replaced on day 5, resulting in a singleton livebirth. To increase the diagnostic accuracy in future cycles and to better understand the cancer risk among the family, efforts were made to identify the mutation breakpoint, which was found to be novel. The data showed that a genomic sequence of 2596 nucleotides including exon 15 and 16 was deleted. In conclusion, the first live birth in Hong Kong after PGD on disease-predisposition mutation marked the beginning of a new phase of PGD practice in Hong Kong.

2845W

Embryo SNP array genotyping: a model for preimplantation diagnosis in human. *F. Campagnari[1,4], Y.T. Utsunomiya[3], A.S. Carmo[4], J.A. Vinsintin[2], J.F. Garcia[3], C. Rosenberg[1], R.V. Alonso[2,4].* 1) Genetics and Evolutionary Biology Dept , University of Sao Paulo, Sao Paulo, SP, Brazil; 2) FMVZ - University of Sao Paulo, Sao Paulo, SP, Brazil; 3) FMVA - UNESP, Araçatuba, SP, Brazil; 4) Deoxi Biotecnologia, Araçatuba, SP, Brazil.

Preimplantation diagnosis and screening is gradually been incorporated to routine for couples with increased risk of particular disorders or to increase in vitro fertilization (IVF) success by selecting euploid embryos. Developments in animal reproduction biotechnology, such as embryo in vitro production (IVP), micromanipulation and preimplantation genetic diagnosis (PGD) can be used as models for application in human PGD. The aim of this study was to perform PGD in bovine embryos through SNP arrays genotyping (BeadChip BovineLD - 6,909 SNP). The small amount of genomic DNA (gDNA) obtained from embryo biopsy is the main limitation for SNP array analysis. Whole Genome Amplification (WGA) (Repli-g® Mini Kit, Qiagen, Hilden, Germany) was used to increase the amount of gDNA from embryo biopsy and allow the analysis of thousands of SNP simultaneously. Eighty-eight IVP bovine embryos were subjected to micromanipulation by microaspiration, forming three groups based on numbers of biopsied cells: G1) 05-10 (n=28); G2) 10-20 (n=37); and G3) > 100 - hatched blastocyst (n = 23). All samples were subjected to the same WGA protocol, and 4µL of each sample were genotyped on iScan/Illumina platform. The genotyping quality was assessed using the Call Rate (CR), GenCall Score (GC10), Alelle Drop In (ADI) and Alelle Drop Out (ADO). Kruskal-Wallis test was applied to investigate differences in the distribution of variables among the groups. Spearman's rank coefficient was calculated to assess the correlation of each possible variables pair. The results showed a positive correlation between CR and GC10 (0.99/P <0.001), while ADI and ADO rates were negatively correlated with CR and CG10 (ADI/CR: -0.87; ADI/GC10: - 0.88; ADO/CR: -0.87; ADO/GC10: -0.86), P<0.001 for all variables. Kruskal Wallis pointed to significant differences in all variables (CR, GC10, ADO and ADI) among the 3 groups of biopsies (G1, G2 and G3). The CR average was 59.26 percent;, 78.47 percent; and 95.97 percent; for G1, G2 and G3, respectively, revealing that the number of cells recovered from the embryo biopsy is crucial for obtaining satisfactory results in SNP analysis. The development of an algorithm based on Mendel's first law (Law of Segregation) increased CR average to 79.69 percent;, 88.20 percent; and 97.28 percent; for groups 1, 2 and 3, respectively. The present study showed that, based on animal model, it is possible to apply SNP arrays in embryo stage, enabling early genotyping for PGD.

2846T

KaryoLite® - A rapid single cell screening assay to simultaneously detect aneuploidies for all chromosomes from whole-genome amplified DNA from 3 day blastomeres. *S. Dallaire, R. Walker, M. Schermer.* PerkinElmer, Molecular Diagnostics, 940 Winter Street, Waltham, MA. 02451.

Aneuploidy is a genetic defect where the number of chromosomes is either greater or less than 46. Aneuploidies have been detected in a majority of miscarriages. The ability to test for aneuploidies prior to IVF implantation may allow better management and outcome of the pregnancy. The FISH and microarray tests that are used in clinical labs to detect aneuploidy on all chromosomes are impractical for many labs because of cost, labor, and time-to-result. We have developed the KaryoLite® kit, which is a bead-based multiplex assay in PerkinElmer's BACs on Beads® family, which will detect aneuploidies for chromosomes 1-22, X and Y. KaryoLite BoBs utilizes a new concept of composite beads having DNA from three different BAC clones on each bead type. The composite clone format expands the region of chromosomal DNA interrogated by each bead. The assay is fluorescence based and uses encoded multiplex beads which have been coupled to BAC derived DNA from defined loci on all 24 chromosomes. This allows detection of aneuploidies on all chromosomes from a clinical sample in a single well of a 96-well microtiter plate with results in as little as 16 hours. A study was performed using 117 DNA samples derived from blastomeres at day 3. DNA from the biopsied cells was processed by whole genome amplification. 240ng of the amplified DNA from each sample was fluorescently labeled and analyzed. The assay was robust with typical signal/background greater than 4, and the KaryoLite assay was completed within 24 hours. Results indicated 68 female and 49 male samples. Of these, 16 were normal, 87 indicated whole chromosome aneuplodies, 9 with arm specific aneuploidies and 4 were -X and 1 was +Y. These results show KaryoLite to be a sensitive, multiplex, high throughput assay that can detect all aneuploidies on all chromosomes from single cell blastomere samples. This demonstrates KaryoLite as a useful tool in preimplantation genetic screening research.

2847W

Quebec perspectives on the medical and social uses of preimplantation genetic diagnosis and on the current service deliveries. *F. Duplain-Laferrière, R. Drouin, C. Bouffard.* Division of Genetics, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Quebec, Canada.

Since 2010, Quebec's healthcare system has covered costs related to preimplantation genetic diagnosis (PGD) for the purpose of detecting specific chromosomal or single-gene defects. However, this measure has several shortcomings. In order to identify problems and to better understand the situation, we consulted the Quebec geneticists, obstetricians-gynecologists and genetic counselors likely to be involved in PGD. Among the objectives of this study, we wanted to know their positions on the medical and social uses that could justify PGD, and hear their proposals to improve the current conditions of the delivery of services.

Methodology: Qualitative Research Design ~ Online questionnaire containing 34 questions has been completed by 15 obstetricians, 15 geneticists and 17 genetic counselors engaged in activities related to prenatal diagnosis in the province of Quebec. Mixed data analysis: quantitative and descriptive for closed questions; qualitative (inductive) and thematic approach for open questions.

Results: The participants propose to: 1) certify laboratories, 2) restrict PGD to lethal or disabling diseases, 3) proscribe PGD for multifactorial diseases or social uses, 4) direct the decision-making process of the patients, 5) create a provincial discussions table or committee to organize the development and the implantation of PGD 6) regulate, standardize and make PGD more accessible. As for conditions that may justify PGD: 1) obstetricians seem less restrictive, 2) geneticists more conservative and 3) genetic counselors more focused on genetic counseling and the needs of the patients.

Conclusion: Despite certain differences between professionals, there is a clear consensus on the necessity of improving the conditions of PGD services, and of reducing problems of accessibility. All also agree on the need to regulate and standardize PGD and to ensure the availability of knowledge relating to the situation.

2848T

***Notchless* Impacts Multiple Signaling Pathways During Pre-Implantation Development.** *C.-L. Lo[1, 3], A.C. Lossie[1, 2], J.B. Sherrill[4].* 1) Department of Animal Science, Purdue Univ, West Lafayette, IN; 2) Dept of Medicine, Indiana University School of Medicine, Indianapolis, IN; 3) PULSe Interdisciplinary Graduate Program, Purdue University, West Lafayette, IN; 4) Department of Biological Sciences, Purdue University, West Lafayette, IN.

Peri-implantation is a critical stage in mammalian development. Genes and genetic pathways that are necessary for establishing and maintaining maternal-fetal interactions play crucial roles in this process. Several members of the WNT pathway are detected in normal embryos and maternal tissues during this transitional stage. Recent studies in our laboratory demonstrated that mutations in *Notchless (Nle1)*, a member of the vast WD40-repeat family and a putative NOTCH signaling molecule, disrupts expression of several genes in the *Wnt* pathway. Mutant embryos also express high levels of *Cdkn1a*, suggesting that they are under severe cellular stress that ultimately leads to caspase-mediated apoptosis at the hatched blastocyst (E4.5) stage. To determine if mutants were undergoing cell cycle arrest prior to apoptosis, we examined expression of *Cdk9*, a downstream target of CDKN1A that promotes cell cycle recovery from replication arrest. At E3.5, mutant embryos showed reduced steady-state levels of Cdk9, suggesting that they were in cell cycle arrest at this stage. Since mutant embryos showed evidence of apoptosis at E4.5 and TRP53 induction often precedes apoptosis, we analyzed *Trp53* expression at the mRNA and protein levels. Although we did not detect altered mRNA levels of *Trp53* by qRT-PCR, immunofluorescence studies using an antibody that detects an active form of TRP53 (Acetyl K376), demonstrates that the acetylated form is only detected in mutant blastocysts at E3.5 and E4.5. Together, these data suggest that at E3.5, Nle1 mutants are undergoing cell-cycle arrest, while at E4.5 the mutants initiate caspase-mediated apoptosis. These two mechanisms are regulated by activation of TRP53. Our data implicate NLE1 in WNT signaling, cell cycle arrest and apoptosis via TRP53-mediated signaling. Intriguingly, WNT signaling is critical for gastrulation in mice. Deletion of Wnt3 leads to failure prior to primitive streak formation, and multiple ligands and receptors are detected in blastocysts and the uterus during peri-implantation. These pathways could provide novel targets for the design of therapeutic interventions for infertility.

2849W

The first report of a viable, 35 week gestation pregnancy following the transfer of a genetically normalized blastocyst; embryo normalization can occur during differentiation to the blastocyst stage. *W.G. Kearns[1,2], M. St. Amant[3], B. Welch[4], J. Carter[4], A. Potts[4], P.R. Brezina[5], A.T. Benner[1], K.J. Tobler[2], G.R. Cutting[6], R.P. Dickey[4].* 1) Center for Preimplantation Genetics, LabCorp, Rockville, MD; 2) Department of Gynecology and Obstetrics, Division of Reproductive Endocrinology and Infertility, Johns Hopkins Medical Institutions, Baltimore, MD; 3) Women's Hospital, Baton Rouge, LA; 4) The Fertility Institute of New Orleans, Mandeville, LA; 5) Fertility Associates of Memphis, Memphis, TN; 6) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, MD.

Case Report. Here we report the progress of a genetically normal 35 week fetus that resulted from the embryo transfer of a genetically normalized blastocyst. A 36 year old female with 9 years of secondary infertility was treated with in vitro fertilization and preimplantation genetic screening (PGS). Two cleavage stage embryos underwent embryo biopsy by laser and a single cell was removed from each embryo for PGS. The single cells underwent lysis and DNA whole genome amplification using a modified random priming method. Following this, comparative genomic hybridization (aCGH) microarray analysis was performed. The single cell from each cleavage stage embryo was aneuploid; embryo one was 48, X, +17, +19, +21 and embryo two was 47, XY, +15. During further embryo culture, only embryo two differentiated to the blastocyst stage of development and then underwent a trophectoderm (TE) biopsy and repeat aCGH analysis. The aCGH results showed embryo two to be euploid, 46, XY. Because no cleavage stage euploid embryo was available, the 'normalized' embryo was transferred after obtaining informed consent from both parents regarding the risk of delivery of an aneuploid baby. A fetal heartbeat was detected at 6 weeks. Amniocentesis performed at 16 weeks; confirmed a normal karyotype: 46, XY. Due to the aneuploidy in the single cell from the cleavage stage embryo, uniparental disomy analysis was performed to confirm that one copy of chromosome 15 came from each parent. Masked aCGH microarray reanalysis of the cleavage stage embryo DNA and the trophectoderm DNA confirmed karyotypes of 47, XY, +15 and 46, XY respectively. The pregnancy is progressing normally at 35 weeks. DNA fingerprinting analysis is ongoing to confirm that the transferred, genetically normalized blastocyst resulted in the viable pregnancy. This report supports the hypothesis that mosaic cleavage stage embryos, with aneuploid cells, can normalize into euploid blastocysts capable of viable pregnancies.

2850T

Stability Testing of a Noninvasive Prenatal Test (NIPT) in a Clinical Setting - the MaterniT21™ PLUS Laboratory-Developed Test. *R.C. Tim[1], J.A. Tynan[2], T.J. Jensen[1], L. Cagasan[1], V. Lu[1], L. Liu[1], S. Sovath[1], M. Riviere[1], P. Oeth[2], M. Ehrich[2].* 1) Sequenom Center for Molecular Medicine, LLC, San Diego, CA; 2) Sequenom, Inc., San Diego, CA.

Excellent clinical performance of a noninvasive test for fetal aneuploidies using next-generation sequencing has been demonstrated by several laboratories. This study demonstrates the robustness of the processes involved in one such assay, the MaterniT21™ PLUS laboratory-developed test, focusing on reproducibility and stability of the results. The study was divided into two portions to determine the robustness of the MaterniT21™ PLUS test. The first was designed to measure reproducibility of chromosomal representation in a pool of circulating cell-free (ccf) plasma DNA isolated from women at increased risk for fetal aneuploidy with a known euploid fetus as determined by fetal karyotyping. For these experiments, ccf DNA obtained from 976 women was combined and used to prepare over 1000 libraries. These libraries were sequenced and analyzed for variability of chromosomal representation. The second portion of the study was designed to investigate the stability of the post-PCR workflow processes. For this part of the study, a set of libraries was generated for use throughout the entire series of experiments, comprised of 44 known euploid and 44 known trisomy 21 samples. For each experimental subset, all factors were kept constant, including operator, reagent lots, and instruments, except for the particular variable of interest. All flow cells from both portions of the study were clustered on the Illumina® cBot in 12-plex and sequenced on the HiSeq™ 2000 (Illumina, San Diego, California). Sequencing reads were de-multiplexed, aligned to the human genome with Bowtie2 and chromosomal representations were determined. Results demonstrate that while library concentrations and raw aligned counts may be variable from plate to plate, chromosomal representation is remarkably stable for the pooled maternal ccf DNA samples processed by multiple operators, across library batches, and measured on multiple sequencing instruments. From the second part of this study, no significant differences in classification z-score values were found across flow cells as a function of library storage time, flow cell storage time, reagent lot, or sequencing instrument. Both sensitivity and specificity for each of these experimental subsets were determined to be nearly 100%. This study demonstrates the stability for use of the MaterniT21™ PLUS test across operators and instruments and reveals the low variability for discrete process steps of the assay.

2851W

Identification of Pathogenic CNV by arrayCGH in Prenatal Cases with Oral Clefts. *Y. CAO[1], Z. LI[3], J. ROSENFELD[4], A. PATEL[5,6], J. HUANG[1], X. SUN[3], T.Y. LEUNG[1], S.W. CHEUNG[5,6], K.W. CHOY[1,2].* 1) Department of Obstetrics and Gynecology, The Chinese University of Hong Kong, Hong kong, China; 2) CUHK-Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; 3) Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China; 4) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, Washington, USA; 5) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 6) Medical Genetics Laboratories, Baylor College of Medicine, Houston, TX, USA.

PURPOSE: Oral clefting is one of the most common congenital malformations. With improvements in sonography, it is now being detected more frequently and accurately during routine fetal morphology scan. Prenatal identification of oral cleft raises questions about prognosis, as oral cleft may be part of syndromic presentations, associated with other congenital malformations, possibly detectable by ultrasound; and/or with neurodevelopment problems, which could not be assessed prenatally. Identification of a genetic cause may help clarify the prognosis and improve clinical management. A systematic review showed that up to 7.9% of oral clefts cases which were combined with different cleft categories both detected prenatally and postnatally was identified with chromosomal defects by karyotype. However the role of submicroscopic chromosomal variants (CNVs) still remains in limited knowledge in prenatal oral cleft cases. METHOD: We performed a retrospective multi-centre study of prenatal cases undergoing clinical microarray-based comparative genomic hybridization (aCGH) with different categories of oral cleft. RESULTS: There were 119 isolated oral cleft cases and 159 syndromic oral cleft cases. We further subdivided these cases into these subgroups: isolated cleft lip(CL) (45/278, 16.2%); isolated cleft lip and palate(CLP) (74/278, 26.6%); syndromic CL (46/278, 16.5%); and syndromic CLP (113/278, 40.7%). Cardiac defects, limb defects and brain abnormalies were the most frequently accompanying defects in syndromic cleft. The overall detection rate of pathogenic CNVs was 12.9%. Pathogenic CNVs were more commonly identified among syndromic groups, (29/159, 18.2%), especially in syndromic cleft lip, while less (7/119, 5.8%) among isolated cleft. Four pathogenic CNV regions, 4p16, 15q11.2, 16p12.2, 17p13.3 were recurrent in our cases. CONCLUSION:aCGH is a powerful tool to detect submicroscopic pathogenic CNVs in prenatal cases with oral clefts. It is highly recommended and should be a first-tier test prenatally, especially for suspected syndromic oral cleft cases.

2852T

Assessing the utilization and distribution of an evidence-based resource recommended in the 2013 American College of Medical Genetics and Genomics statement on noninvasive prenatal screening for fetal aneuploidy. *S. Meredith[1], B. Skotko[2], C. Brasington[3].* 1) Human Development Institute, University of Kentucky, Lexington, KY; 2) Massachusetts General Hospital Down Syndrome Program, Boston, MA; 3) Down Syndrome Clinic at Carolinas Medical Center, Charlotte, NC.

Introduction and Purpose: The purpose is to share data for the first time on the utilization and distribution of an evidence-based resource as implemented by medical providers delivering a prenatal diagnosis of Down syndrome (DS). "Understanding a DS Diagnosis" (www.lettercase.org) was prepared with assistance from representatives of ACMG, ACOG, NSGC, and the national DS organizations. It is part of the National Center for Prenatal and Postnatal DS Resources at the University of Kentucky and was recently included as a recommended resource in the 2013 ACMG policy statement. Specifically this assessment determines: 1. Trends in US geographic areas 2. Professional discipline trends 3. Referral sources used by medical providers 4. The impact of professional guidelines, professional newsletters, and internet promotions Methods: This study examines the number of medical providers who requested/purchased printed copies of "Understanding a DS Diagnosis." When requesting a free printed book, medical providers are required to submit their name, email, medical practice, address, area of practice, and can optionally indicate how they learned about the book. The data was collected and organized from 681 medical requests/purchases from June 1, 2012-June 1, 2013 to show the number of medical providers from different disciplines requesting books from different regions. Results: The data reveals that the top states where medical providers request/purchase the Lettercase book include Washington(85), California(82), Pennsylvania(57), Minnesota(55), and Illinois(47). The top discipline requesting books is genetic counselors(154) while the least requests come from OB/GYNs(49) and Family Practitioners(20). Data indicates that the majority of medical providers obtain information about patient resources via professional organizations, colleagues, and email. Conclusion: The top states requesting booklets typically include higher population states and states where local Down syndrome organizations and state health departments are conversely less active in distributing books. Genetic counselors are most likely to utilize a patient resource and receive the most DS training. Yet, OB/GYNs and family practitioners are least likely to utilize the book and receive the least training about DS. Data indicates that 'word of mouth' and professional organizations are most influential in the sharing of a patient resource, and online professional newsletters prompt the most requests.

2853W

Exploring Placental Gene Expression Pattern in Abnormal Fetal Growth. *A. Sabri[1,2], C.H.M. Ng[1,3], D. Lai[2], A. D'Silva[1], J. Kaur[1], J.A. Hyett[1,3].* 1) Department of Obstetrics, Gynecology and Neonatology, Queen Elizabeth II Research Institute for Mothers and Infants, The University of Sydney, Sydney Australia; 2) Molecular Biology Facility, Bosch Institute, The University of Sydney, Sydney, Australia; 3) RPA Women and Babies, Royal Prince Alfred Hospital, Sydney, Australia.

Objective: Extremes of fetal growth are associated with increased perinatal mortality and morbidity and a higher prevalence of cardiovascular disease, obesity and diabetes in later life. The placenta plays a pivotal role in fetal growth and development. We aimed to identify changes in placental gene expression in term pregnancies with evidence of growth dysfunction and to identify candidate genes that may be used to identify abnormal patterns of growth prior to delivery.

Materials and methods: Placenta samples were collected and classified from pregnancies that were small for gestational age (SGA; <10th centile; n=5), large for gestational age (LGA; >90th centile; n=6) or normal (AGA; 40-60th centile; n=5). All pregnancies were 39 weeks' gestation, with male infants delivered by caesarean section. RNA was extracted from the placental samples prior to microarray gene expression analysis (Affymetrix HG-U219 array). Microarray data were analysed using Partek Express and Ingenuity Pathway Analysis. Significant differential gene expression was defined by >2 fold-change with p<0.05.

Results: Significant differential gene expression was found in both SGA and LGA placentas, with up-regulation in 68 and 18 genes and down-regulation in 9 and 9 genes respectively. Differential expression in SGA placentas was seen in genes involved in embryonic and connective tissue development and in developmental, metabolic and neurological disorders. In contrast, differential expression in LGA placentas involved genes involved in cardiovascular and haematological disease, cell-mediate immune responses, endocrine and developmental disorders. Individual findings of interest include gremlin 2 (*GREM2*) and zinc finger protein 711 (*ZNF711*) (SGA) and leptin (*LEP*) and matrix metallopeptidase 12 (*MMP12*) (LGA), playing important roles in embryonic and organ development, cell proliferation and tissue differentiation.

Conclusions: Dysfunctional growth is associated with differential placental gene expression and affects genes involved with a whole spectrum of developmental and cellular functions. *GREM2* and *ZNF711* appear to be good candidates for prediction of SGA in term pregnancies. *LEP* appears to be decreased in LGA infants, which may be of significance for long term outcome.

2854T

Diagnostic dilemma: fetal cardiomyopathy presenting at a late gestational age. *B. Suskin[1,2], K. Bajaj[1,2], M. Rosner[1,2], P. Dar[1,2], S. Klugman[1,2].* 1) Department of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center, Bronx, NY; 2) Albert Einstein College of Medicine, Bronx, NY.

Background: Cardiomyopathies are diseases of the myocardium that affect the mechanical or electrical function of the heart. American Heart Association has stated that most are due to genetic causes. Elucidating a cause, expediting testing and explaining a prognosis is challenging in the fetal setting. Case: A 30 year-old multiparous woman at 30 weeks with gestational diabetes was referred from an outside clinic for fetal surveillance. Her pregnancy was otherwise uncomplicated and she denied any significant family history. Oligohydramnios was found. At her follow-up ultrasound with fetal medicine specialists, oligohydramnios was seen again. They also found a grossly enlarged heart, a small pericardial effusion and an increased PR interval and suspected cardiomyopathy. The fetus did not have hydrops. Fetal echocardiograms by pediatric cardiologists confirmed these findings. TORCH and Lupus work-ups were sent. The patient was lost to follow-up, but represented in labor. A female infant was delivered at 39 weeks at 3585 grams with apgars 5/6/8 at 1, 5 and 10 minutes respectively. She was intubated and transferred to the pediatric hospital. Echocardiograms show bilateral ventricular hypertrophy and moderately depressed function. The work-up is pending. Discussion: Due to the fact that the patient presented at 30 weeks with sparse prenatal care, the work-up was limited to infectious and maternal causes. Neonatal cardiomyopathies can be seen in high-output failure, anemia, volume overload and myocardial damage. A common non-genetic cause of neonatal cardiomyopathy is hyperinsulinemia secondary to maternal diabetes. There are also many possible genetic causes. These include syndromic, metabolic, neuromuscular, mitochondrial and sarcomere mutations. Despite our abilities to evaluate for many such causes, we are often limited in the prenatal setting by time and cost. It is important to consider both genetic and acquired causes of cardiomyopathy of fetuses during pregnancy. The increased risk of fetal demise is known, especially when hydrops is present. This poster presents the genetic considerations, work-up and limitations of prenatally diagnosed fetal cardiomyopathy.

2855W

Improving efficiency and cost of next generation sequencing of maternal cell free DNA for the detection of fetal aneuploidy.
L. Chitty[1,3], J. Weir[2], C. Boustred[3], S. Fielding[3], F. McKay[3], H. White[4], Z. Kingsbury[2], S. Humphray[2], E. Tsogka[2], N. Lench[3], J. Betley[2]. 1) Clinical Molecular Genetics Unit, UCL Institute of Child Health, London, London, United Kingdom; 2) Illumina Inc, Chesterford Research Park, Little Chesterford, Essex, CB10 1XL, UK; 3) North East Thames Regional Genetics Laboratory, Great Ormond Street NHS Foundation Hospital, London; 4) Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, Wiltshire SP2 8BJ.

We have recently reported preliminary data describing a new PCR-free method of preparing cfDNA for next generation sequencing (NGS), which may remove PCR-bias and potentially improve accuracy of trisomy 13 and 18 detection as well as decreasing costs and time to results. Before use in clinical practice, it is necessary to ensure that this method is accurate for the detection of aneuploidies, and is not affected by the various protocols used for maternal blood collection. Here we aim to investigate the robustness of this new method in a larger number of normal and trisomic pregnancies with plasma prepared in different ways to inform clinical implementation. Maternal blood was taken into both EDTA and cell-stabilising tubes from women undergoing invasive tests for clinical indications. cfDNA was extracted from plasma double spun at varying time intervals. Samples were prepared for sequencing (HiSeq 2500) using a modified version of Illumina's PCR-free protocol, eliminating amplification bias as a source of variable accuracy. For each sample preparation method, percentages of mapped reads by chromosome were compared to determine whether plasma or library preparation method had a significant effect. Duplicate samples were taken and prepared using a modified version of Illumina's TruSeq PCR workflow for comparison between the two library prep methods. Our results showed that after alignment with iSAAC 1x36 bp (hg19), all 3 collection/spin workflows produced similar standard deviation (SD) values after removing aneuploidy samples from SD values in each case; including for Chromosome 13 all were < 0.03 standard deviation (SD) of the percentage aligned. In addition when the data was combined into a single data set, the standard deviation (SD) values were similar showing robust data when using a combination of collection strategies is possible. The z scores were plotted based on the average and SD of each data set and all the data sets called correctly Trisomy results. We have shown that omitting PCR enrichment of libraries for NGS of cfDNA does not affect sequencing performance for cfDNA collected into EDTA or cell-stabilising tubes, and appears to give reliable results for trisomy calls in plasma separated up to 24 hours after blood draw. Omitting amplification eliminates any possibility of PCR-bias and the results presented here indicate that this approach may improve accuracy as well as decreasing time and costs for NGS of maternal plasma cfDNA.

2856T

Cordocentesis: an alternative prenatal procedure for women who missed amniocentesis in developing regions of China.
Q. Cao[1, 2], J. Ge[1,2], Y-Y. Peng[1,2], E.C. Jenkins[4], W.T. Brown[4], N. Zhong[1,2,3,4]. 1) The 4th Municipal Hospital of Shijiazhuang, Shijiazhuang, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in DD, Staten Island, NY.

Objective: To assess the efficacy of cordocentesis for rapid karyotyping in a high-risk obstetric population. Methods: Cordocenteses were performed on 64 pregnant women with different indications for prenatal diagnosis in an outpatient setting from November 2009 to September 2012. Fetal chromosome karyotypes were examined. Results: The most frequent indication for cordocentesis was a fetal abnormality on sonography (92.2%). All of the procedures were performed using the free cord loop. The success rate of cordocentesis was 92.2% (59/64) and cell culture was 94.9% (56/59). Transient bleeding was common at the puncture site, but all stopped spontaneously within one minute. Seven chromosomal abnormalities (12.5%) and 4 chromosomal variations (7.1%) were identified among 56 specimens cultured. Six of the seven chromosomal abnormalities (11.8%) were identified among 51 cases with abnormal ultrasound findings. Trisomy, the most prevalent chromosomal abnormality, was present in 6 of the 7 (85.7%) abnormalities detected. One case of 69, XXY was also identified among the 7 chromosome abnormalities. Conclusions: Cordocentesis is a rapid and reliable procedure for prenatal diagnosis, especially for mid and late pregnancies.

2857W

Single nucleotide polymorphism (SNP)-based non-invasive prenatal testing (NIPT) detects triploidy: two case studies.
Z. Demko, M.P. Hall, M. Hill, B. Zimmermann, S. Sigurjonsson, M. Rabinowitz. Natera Inc., San Carlos, CA.

Objective: To report two NIPT cases with evidence of triploidy. **Design:** Two triploid samples were analyzed using the non-invasive prenatal Next-generation Aneuploidy Test Using SNPs (NATUS) algorithm under a commercial protocol (reporting detection of Trisomies 13, 18, and 21, and Monosomy X) or as part of an IRB-approved research protocol. **Materials and Methods:** Cell-free DNA (cfDNA) isolated from plasma from pregnant women was analyzed via sequencing after amplification using a multiplex PCR protocol targeting 19,488 SNPs. Sequencing results were analyzed with the NATUS algorithm. Laboratory personnel (and the NATUS algorithm) were blinded to sample karyotype. **Results:** NATUS analysis identified more than two fetal haplotypes present for multiple chromosomes, indicating twins or triploidy. **Case 1** (fetal fraction: 6.4%): multiple paternal haplotypes on multiple chromosomes (13, 18, 21, X) were detected, indicating either paternally-inherited triploidy or fraternal twins. **Case 2** (fetal fraction: 20.8%): multiple paternal haplotypes on multiple chromosomes (13, 18, 21, X) were detected, similarly indicating either a paternally-inherited triploidy or fraternal twins. Ultrasound with chorionic villus sampling **(Case 1)** or amniocentesis **(Case 2)** confirmed single gestations affected with triploidy. **Conclusions:** This SNP-based approach examines the relative distributions of different alleles at polymorphic loci, thus not requiring a reference chromosome, and therefore offers the unique ability to detect triploidy. Here, this method correctly flagged both triploidy cases. This method is expected to distinguish twins from triploidy based on the presence (in the case of twins) or absence (in the case of triploidy) of an extra fetal haplotype. Detection of paternal triploidy is particularly clinically relevant as it is correlated with partial molar pregnancy, and molar pregnancy (partial or complete) can develop into choriocarcinoma. Studies are ongoing to differentiate all twin and triploidy cases; in the interim ultrasound will be required to confirm which cases may be at risk of triploidy. **Support:** NIH 4R44HD062114-02.

2858T

Prenatal diagnosis of 23 cases of microduplication 22q11.2.
C. DUPONT[1], F. GRATI[2], K.W. CHOY[3], S. JAILLARD[4], J. TOUTAIN[5], M.L. MAURIN[6], J.A. MARTINEZ-CONEJERO[7], C. BENETEAU[8], D. MOLINA-GOMES[9], N. HORELLI-KUITUNEN[10], A. ABOURA[1], C. BAUMANN[11], E. BLONDEEL[9], B. BESSIERES-GRATTAGLIANO[6], A.C. TABET[1], G. SIMONI[2], B. BENZACKEN[1,12], F. VIALARD[9]. 1) Cytogenetic, APHP-Hopital Robert Debre, Paris, France; 2) TOMA, Advanced Biomedical Assays Laboratory-Busto Arsizio, Varese, Italia; 3) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong; 4) Cytogenetics Department, Pontchaillou University Hospital, Rennes, France; 5) Cytogenetic Laboratory, Medical Genetics Department, Pellegrin Maternity, CHU Bordeaux, Bordeaux, France; 6) Service d'histologie, embryologie et cytogénétique- APHP, Hôpital Necker-Enfants malades- Paris, France; 7) Fundación IVI-Instituto Valenciano de Infertilidad (IUIVI)-University of Valencia-Valencia, Spain; 8) Department of clinical genetics- CHU de Nantes, France; 9) Department of Cytogenetics, Obstetrics and Gynaeacology, CHI Poissy St Germain, Poissy, France; 10) United Medix laboratories Ltd. Helsinki, Finland; 11) Genetic Department-Robert Debré Hospital- Paris, France; 12) Service d'Histologie-Embryologie et Cytogenetique, Biologie de la Reproduction- APHP, Hopital Jean Verdier- Bondy, France ; UFR-SMBH, Paris XIII, France.

Microduplication 22q11.2 is mainly characterized by a high variable clinical phenotype that ranges from mild dysmorphic features with very moderate learning disabilities to severe malformations with profound mental retardation. One of the main characteristics of this new chromosomal pathology is its extreme phenotypic variability. Therefore genetic counselling is very difficult when a microduplication 22q11.2 is found in prenatal diagnosis. We report twenty three prenatal cases of 22q11.2 duplication mostly diagnosed by new molecular cytogenetic technology: the Prenatal BACs on Beads™. We analysed phenotypic profiles of fetuses with microduplication 22q11.2, including indication of prenatal diagnosis, and outcomes of the pregnancy when available. Several fetuses' DNAs (16/23) were analysed by microarray. Out of the 23 cases, 15 are inherited, 6 are de novo and 2 of unknown origin. Termination of pregnancy was mainly due to ultrasound findings and occurred in 7 cases. We discussed about the difficulty of genetic counseling in this syndrome and about different hypothesis to explain its clinical heterogeneity. We examined in particular, the co-existence of additional CNVs and their contribution to the phenotypic variation in the 22q11.2 microduplication syndrome. In conclusion we report the first prenatal serie of microduplication 22q11.2. Our data are consistent with the extreme variability of this syndrome. New molecular cytogenetics technologies as Prenatal BoBs™ and microarray would probably increase the detection rate of this cryptic chromosomal rearrangement. Moreover the clinical follow-up during childhood of patients with 22q11.2 microduplication diagnosed prenatally would help the understanding of this pathology.

2859W

Validation of epigenetic marker for noninvasive prenatal diagnosis of fetal trisomy 18. *DE. Lee[1], SY. Kim[1], JH. Lim[1], HJ. Kim[1], SY. Park[1], HM. Ryu[1, 2].* 1) Laboratory of Medical Genetics, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, Korea.

Background: The quantification of cell-free fetal DNA by methylation-based DNA discrimination has been used in noninvasive prenatal diagnosis of fetal chromosomal aneuploidy. Maspin (SERPINB5) gene, located on chromosome 18q21.33, is hypomethylated in the placenta and completely methylated in maternal blood cells. The objective of this study was to evaluate the accuracy of noninvasive fetal trisomy 18 detection using tissue specific-methylation of maspin in the first trimester of pregnancy. Methods: A nested case-control study was conducted using maternal plasma samples collected from 66 pregnant women carrying 11 trisomy 18 and 55 normal fetuses. Using real-time quantitative methylation-specific PCR, the concentrations of unmethylated maspin (U-maspin) and methylated maspin (M-maspin) were measured in first trimester maternal plasma. Results: U-maspin concentrations were significantly elevated in women with trisomy 18 fetuses compared with controls (27.2 vs 6.7 copies/mL; P<0.001). M-maspin concentrations were significantly higher in women with trisomy 18 fetuses compared with controls (96.9 vs 19.5 copies/mL, P<0.001). The specificities of U-maspin and M-maspin concentration for noninvasive fetal trisomy 18 detection were 96.4% and 74.5%, respectively, with a sensitivity of 90.9%. In the risk assessment for fetal trisomy 18, the adjusted odds ratios of U-maspin and M-maspin concentration were 325.2 (95% confidence interval: 17.9-5903.8, P<0.001) and 19.0 (95% confidence interval: 2.1-175.6, P=0.009), respectively. Conclusions: Our results suggest that U-maspin and M-maspin concentration may be useful as potential biomarkers for noninvasive fetal trisomy 18 detection in the first trimester of pregnancy, irrespective of the sex and genetic variations of the fetus.

2860T

Highly accurate non-invasive detection of fetal aneuploidy for chromosomes 13, 18, 21, X and Y. *B. Levy[1], S. McAdoo[2], B. Zimmermann[2], M. Banjevic[2], B. Pettersen[2], M. Hall[2], Z. Demko[2], M. Hill[2], M. Rabinowitz[2].* 1) Department of Cell Biology and Pathology, Columbia University, New York, NY; 2) Natera, Inc, 201 Industrial Rd, San Carlos, CA, 94070.

Objective: To develop a non-invasive prenatal test (NIPT) that will detect fetal chromosome 13, 18, 21, X, and Y copy number through analysis of cell-free (cfDNA) isolated from maternal plasma using the Next-generation Aneuploidy Testing Using SNPs (NATUS) algorithm. **Methods:** cfDNA was isolated from 699 maternal plasma samples at ≥9 weeks gestation, including 70 aneuploid samples, under an institutional review board approved protocol. DNA at 19,488 polymorphic loci covering the target chromosomes was amplified in a single multiplex-PCR assay and sequenced. Results were analyzed using the NATUS algorithm, which uses Bayesian statistics to accurately identify copy number. NATUS considers parental genotypes and HapMap data to calculate expected allele distributions for a large number of possible fetal ploidy states. It then compares the predicted allele distributions to the actual allelic distributions as measured in the cfDNA sample, sums the likelihoods of each ploidy state hypothesis (monosomy, disomy or trisomy) based on the sequencing data, and calls the hypothesis with the maximum likelihood as the ploidy state. Unlike other methods, NATUS calculates a sample-specific accuracy without requiring a reference chromosome. **Results:** In samples that passed quality control thresholds (94.2%), 99.8% of calls were correct. The mean confidence calculated by NATUS for these samples was greater than 99%. Samples typically failed to meet the quality control threshold due to poor DNA quality and/or low fetal fraction (less than 4%). **Conclusions:** The NATUS algorithm is the only NIPT method that calculates sample-specific accuracies and detects chromosomally abnormal fetuses from maternal plasma with high accuracy at chromosomes 13, 18, 21, X and Y as early as 9 weeks gestation.

2861W

Disease-specific characteristics of fetal epigenetic markers for non-invasive prenatal diagnosis of trisomy 21. *J. Lim[1], D. Lee[1], S. Park[1], D. Kim[1], H. Kim[1], H. Ahn[2], S. Lee[2], K. Choi[2], M. Kim[2], H. Ryu[1,2].* 1) Laboratory of Medical Genetics, Medical Research Institute, Cheil General Hospital and Women's Healthcare Center, Seoul, South Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine, Seoul, South Korea.

BACKGROUND: Non-invasive prenatal diagnosis of trisomy 21 (T21) is being actively investigated using fetal-specific epigenetic markers (EPs) that are present in maternal plasma. Recently, 12 EPs on chromosome 21 were identified based on tissue-specific epigenetic characteristics between placenta and blood, and demonstrated excellent clinical performance in the non-invasive detection of fetal T21. However, the disease-specific epigenetic characteristics of the EPs have not been established. Therefore, we validated the disease-specific epigenetic characteristics of these EPs for use in non-invasive detection of fetal T21. METHODS: We performed a high-resolution tiling array analysis of human chromosome 21 using a methyl-CpG binding domain-based protein (MBD) method with whole blood samples from non-pregnant normal women, whole blood samples from pregnant normal women, placenta samples of normal fetuses, and placenta samples of T21 fetuses. Tiling array results were validated by bisulfite direct sequencing and qPCR. RESULTS: Among 12 EPs, only four EPs were confirmed to be hypermethylated in normal placenta and hypomethylated in blood. One of these four showed a severe discrepancy in the methylation patterns of T21 placenta samples, and another was located within a region of copy number variations. Thus, two EPs were confirmed to be potential fetal-specific markers based on their disease-specific epigenetic characteristics. The MBD results of these EPs were consisted with the results obtained by bisulfite direct sequencing and qPCR. However, the methylation level of these EPs showed inter-individual variability. CONCLUSION: We validated that two EPs have the potential to be fetal-specific EPs which is consistent with their disease-specific epigenetic characteristics. The findings of this study suggest that disease-specific epigenetic characteristics should be considered in the development of fetal-specific EPs for non-invasive prenatal diagnosis of T21.

2862T

The concerns of health professionals and pregnant women involving the non-invasive prenatal diagnosis of trisomy 21 in Quebec and in France. *A.K. MAGLO[1], R. DROUIN[1], JM. MOUTQUIN[2], C. BOUFFARD[1].* 1) Division of Genetics, Department of Pediatrics, Université de Sherbrooke, Faculty of Medicine and Health Sciences, Sherbrooke, Quebec, Canada; 2) Department of Obstetrics and Gynecology, Université de Sherbrooke, Faculty of Medicine and Health Sciences, Sherbrooke, Quebec, Canada.

Introduction: In France, the Agence Nationale de Sécurité du Médicament currently assesses the relevance of authorizing the non-invasive prenatal diagnosis (NIPD) through blood testing of trisomy 21 (T21) (PrenaTest®). As Quebec begins to be confronted with this issue, our objective was to identify and compare the concerns of health practitioners, pregnant women and couples involving the NIPD of T21 from probative data available for France and Quebec. Methodology: Qualitative aspect: 1) Systematic narrative review: research through key-words in probative data (scientific publications and gray literature) in databanks: PubMed, ERIC, JSTOR, SCOPUS, etc. 2) Thematic and comparative analysis of the content: coding, categorization and classification of information in themes, allowing for the identification of the variables and relationships established in the texts. 3) Comprehensive synthesis. Results: In Quebec, reflection focuses mostly on the invasive diagnosis of T21. In France, the debate extends to issues related to NIPD. In both countries, health practitioners, pregnant women and their partners are equally concerned with 1) the lack of information on NIPD and 2) the importance attributed to genetic counseling. Finally, health practitioners have questions relating to 3) standards of practices and 4) ethics; pregnant women have questions relating to 5) decision-making autonomy and 6) the attitude of society towards individuals living with T21. Conclusion: Parallel to the evolution of NIPD techniques, we continue to lack probative data that would allow us to establish health policies ensuring that this change in paradigm in the field of prenatal diagnosis does not diminish the quality of services and decision-making autonomy of women, and does not encourage discrimination against people with T21.

2863W

A single nucleotide polymorphism-based approach to non-invasive prenatal testing identifies lingering cell-free fetal DNA in pregnancies with vanishing twins. *S. McAdoo, M. Savage, M.P. Hall, B. Zimmermann, S. Sigurjonsson, Z. Demko, M. Rabinowitz.* Natera, Inc., San Carlos, CA.

Objective: Using single nucleotide polymorphism (SNP)-based non-invasive prenatal testing (NIPT) to analyze cell-free fetal DNA (cffDNA) allows for identification of samples with multiple maternal and/or paternal haplotypes; these additional parental haplotypes distinguish single from multiple gestations. Vanishing twins may confound fetal copy number calling in singleton pregnancies due to cffDNA from the vanishing twin placenta. A SNP-based approach allows detection of vanishing twins in reportedly singleton pregnancies, and may reduce the frequency of incorrect results. We present three such cases. **Methods:** Commercial samples were processed at a single reference laboratory. Isolated cfDNA was amplified using multiplex PCR targeting 19,488 SNPs covering chromosomes 13, 18, 21, X, and Y. Sequencing data was analyzed using Next-generation Aneuploidy Test Using SNPs (NATUS) algorithm that uses Bayesian statistics to analyze multiple copy number hypotheses and determine the Maximum Likelihood hypothesis, calculating chromosome-specific accuracies without requiring a reference chromosome. Follow-up clinical information was sought for cases in which NATUS was suggestive of a vanishing twin. **Results:** Case 1: vanishing twin reported at 6w0d, maternal blood drawn at 14w3d; Case 2: twin loss reported at 6w5d, maternal blood drawn at 13w5d. Neither case had evidence of a second sac at blood draw. Case 3: two sacs but one fetus reported at 12w6d, maternal blood drawn at 19w1d. **Conclusions:** This method detected cffDNA from a vanishing twin as many as 8 weeks after demise. The contribution of cffDNA from a vanishing twin increases the total cffDNA estimate, potentially resulting in analysis of samples with cffDNA levels from the surviving twin below that which is appropriate for accurate detection of aneuploidy. Additionally, residual cffDNA from a vanishing twin could be aneuploid (explaining the cause for the loss) whereas the surviving twin could be euploid. Together, this may result in incorrect results when using non-SNP-based methods, as these are classified as singleton pregnancies at the time of maternal blood draw. This SNP-based NIPT approach identifies these cases, potentially decreasing incorrect results.

2864T

A KEY ROLE FOR AMNIOTIC FLUID CITRULLINE ANALYSIS IN THE PRENATAL DIAGNOSIS OF CITRULLINEMIA TYPE I. *M.J. Miller, V.R. Sutton, Q. Sun, S.H. Elsea.* Biochemical Genetics Laboratories, Baylor College of Medicine, Houston, TX.

The urea cycle disorder citrullinemia type I results from argininosuccinate synthetase 1 (ASS1) deficiency and can be diagnosed prenatally using fetal enzymatic and/or molecular testing. Additional analysis of amino acid levels in amniotic fluid (AF-AA) has been suggested to aid in the prenatal diagnosis of multiple urea cycle disorders but in the case of citrullinemia the published experience is limited (PMID#6739433). To assess the clinical utility of AF-AA testing we reviewed the 30 prenatal citrullinemia cases diagnosed with both enzyme and AF-AA assays in our laboratory over the last decade, 16 of which had further molecular testing and all involved mothers with prior affected offspring. To provide additional support to our normal reference range we also quantified the AF-AA levels from 50 mothers receiving prenatal testing unrelated to citrullinemia and for whom no abnormalities were detected. Overall, AF-citrulline values strongly correlated with argininosuccinate synthetase enzyme activity in amniocytes. Consistent with this, AF-citrulline levels for affected fetuses (n=11) were significantly elevated above unaffected/carrier fetuses (n=69). Setting a cutoff of 50 μM AF-citrulline afforded a positive predictive value of 100% with 91% sensitivity. In two instances AF-citrulline confirmed a diagnosis in the face of ambiguous molecular and/or biochemical test results. Taken together our data demonstrate that amniotic citrulline levels serve as an important adjuvant in the interpretation of enzyme and DNA data in the prenatal diagnosis of citrullinemia.

2865W

Experience using a rapid assay for aneuploidy and microdeletion detection in over 2900 prenatal specimens. *S. Sulpizio, S.A. Morton, A.M. Bandholz, L.D. McDaniel, R.A. Schultz, B.S. Torchia, J.B. Ravnan, J.W. Ellison, P. Mowery-Rushton, J.A. Rosenfeld.* Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA.

While microarray testing can identify chromosomal abnormalities missed by karyotyping, its prenatal use is often avoided in low-risk pregnancies due to the possible identification of variants of uncertain significance (VOUS). We tested 2971 samples using a rapid, BACs-on-Beads^{TM}-based assay with probes for sex chromosomes, common autosomal aneuploidies, and a panel of 15 microdeletion syndromes and 5 reciprocal microduplication syndromes, designed as an alternative to microarray testing in low-risk pregnancies and an alternative to rapid aneuploidy FISH in high-risk pregnancies that also undergo microarray analysis. Interpretable results were obtained in 2942 cases (99.0%), with 89% receiving results the next business day after sample receipt. Aneuploidies were detected in 7.3%, with the rate dependent on indication for study ($p<0.0001$). Partial chromosome abnormalities were detected in 0.5% (n=14), including six low-risk cases referred for maternal age, abnormal maternal serum screen, or isolated soft ultrasound markers. One VOUS was obtained (0.03%), which was determined to be benign following further characterization by microarray. Abnormalities were confirmed through secondary testing; neither false negatives nor false positives were found, within limitations of the test. Female polyploidy cannot be detected, while polyploidies with Y chromosomes are suspected and confirmed through additional analysis. This assay allows for detection of clinically significant microdeletions and aneuploidy with rapid results. It tests for more conditions than rapid aneuploidy FISH, though cannot detect all polyploidy. When combined with karyotyping, this assay provides increased interrogation of specific chromosomal regions, while limiting the identification of VOUS.

2866T

Methods for fetal fraction quantification in circulating cell-free DNA sequencing libraries. *J. Tynan[1], G. Hogg[1], J. Fox[1], P. Iyer[1], M. Ehrich[2].* 1) Sequenom Center for Molecular Medicine, San Diego, CA; 2) Sequenom, Inc., San Diego, CA.

OBJECTIVE: Circulating cell-free (ccf) fetal DNA fraction in maternal plasma can be measured using chromosome Y markers from a male fetus, fetal specific methylation markers, or paternally inherited alleles absent in the maternal genome. Here, we compare the quantification of ccf fetal DNA fraction in maternal plasma sequencing libraries using single nucleotide polymorphism (SNP) allele frequencies, chromosome Y representation by sequencing, and the ratio of chromosome Y and total genomic copies by droplet digital PCR (ddPCR). METHOD: Maternal plasma DNA was isolated and plasma DNA libraries were prepared using TruSeq reagent kits (Illumina, San Diego California), clustered at 12-sample plexing and sequenced on the HiSeq 2000 (Illumina) for 36 cycles. Plasma DNA sequencing reads were aligned to the human genome (hg19) and the proportion of reads derived from chromosome Y were used to estimate fetal fraction. Maternal plasma DNA libraries were also used as template for a multiplexed single-tube PCR SNP panel and a duplex ddPCR assay. The multiplexed single-tube PCR assay amplified 67 high minor allele frequency SNPs and was designed such that a subsequent universal PCR incorporated indexed adapter sequences to allow multiplexed sequencing on the MiSeq (Illumina). The duplex ddPCR assay was designed to quantify SRY copies arising from the male fetal genome and TERT total genomic copies on the QX100 (Bio-Rad, Hercules California). RESULTS: By sequencing 84 plasma DNA libraries for chromosome Y representation, 38 samples passing assay QC metrics were derived from a donor carrying a male fetus. Sequencing based chromosome Y fetal fraction significantly correlated to ddPCR based fetal fraction ($r2 = 0.72$). ddPCR fetal fraction showed significant correlation to SNP based fetal fraction ($r2 = 0.63$). Sequencing based chromosome Y fetal fraction showed significant correlation to SNP based fetal fractions ($r2 = 0.81$). CONCLUSION: These data demonstrate concordance of ccf DNA fetal fraction measurement in sequencing libraries by multiple methods. SNP allele frequencies and chromosome Y representation showed the highest correlation. The ddPCR method likely is negatively impacted by sampling error of SRY copies leading to a higher variability of fetal fraction quantification. Of course, only the SNP allele frequency assay is able to quantify fetal fraction in pregnancies carrying a female fetus.

2867W

Prenatal MLPA screening and aCGH analysis detected cytogenomic abnormalities in four cases with fetal ultrasound anomalies. *J. Xie[1], Z. Xu[1, 2], Q. Geng[1], F. Xu[2], P. Li[2].* 1) Center for Prenatal Diagnosis, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, Guangdong, China; 2) Department of Genetics, Yale School of Medicine, New Haven, CT.

Ultrasound detected fetal anomalies have been the most important clinical indications for prenatal genetic diagnosis. However, the associations of specific ultrasound findings with most syndromic cytogenomic disorders remain elusive. We have applied rapid MLPA screening and aCGH analysis to high risk pregnancies with abnormal ultrasound findings. Cytogenomic abnormalities were characterized in four cases. The first fetus with ultrasound findings of growth retardation, bilateral cleft lip and palate, right-sided aortic arch and multiple cystic hypoplasia was detected with a 40.701 Mb duplication of 8q22.3-q24.3 and a 23.839 Mb deletion of 7q33-q36.3. Follow up parental study found the father to be a carrier of a balanced 7q33/8q22.3 translocation. The second fetus with single-ventricle heart, transposition of the great arteries and double superior vena cava was diagnosed as Jacobsen syndrome (OMIM#147791) by a de novo 13 Mb distal deletion of 11q24.1-q25. The third fetus with split-hand split-foot malformation and multiple umbilical cord cysts was diagnosed as type 1 split-hand/foot malformation (SHFM1, OMIM#183600) with a de novo 19.971 Mb interstitial deletion of 7q11.23-q21.3. The fourth fetus with pleural effusion, a hypoplastic cerebellum and a large ventricle showed a 699.8 Kb deletion including the TERT gene at 5p15.33 for the diagnosis of Cri du Chat syndrome (OMIM#123450). Post-test genetic counseling was performed with detailed genomic information and well characterized postnatal syndromic features. All parents made an informed decision to elect termination. Prenatal ultrasound findings reported in literature for Jacobsen syndrome, SHFM and Cri du Chat syndrome were reviewed and summarized. Our results demonstrated that comprehensive evaluation of abnormal prenatal ultrasound findings and their association with known cytogenomic disorders will lead to better practice of prenatal genetic testing and counseling.

2868T

Chinese Alliance of Translational Medicine for Maternal and Children's Health (CATMMACH): A unique resource for longitudinal cohort study of pregnancies. *J. Pan[1,2], Q-X. Shi[1,2], Y. Gu[1,2], X-C. Lou[1,2], M-F. Hua[1,2], N. Zhong[1,2,3,4].* 1) Lianyungang Maternal and Children's Hospital, Lianyungang, Jiangsu, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health (CATMMACH), China; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in DD, Staten Island, NY.

Globally, 536,000 women died from causes related to pregnancy and childbirth in 2005, and 3.7 million newborn infants died in 2004. Globally, an estimated 15 million infants are born preterm annually, and more than one million infants die annually from preterm birth complications, making these complications the most frequent cause of all deaths worldwide of children younger than 5 years of age. Recent translational research to elucidate the molecular mechanism underlying preterm births is focusing on two major categories: conventional research to explore biomarkers, which may be applied to early prevention of preterm birth, and an integrated systems biology approach with '-omic' technology, to uncover the genetic risk factors, including gene mutation, genetic predisposition or environmental-genetic interaction. However, the molecular mechanisms underlying preterm birth are not yet clearly understood because it is a complex disorder. Many environmental and genetic factors are involved, and gene-gene and gene-environmental interaction(s) likely play key roles. To study these roles, a national network in China, the Chinese Alliance of Translational Medicine for Maternal and Children's Health (CATMMACH), has been developed, comprising eight hospitals and two medical genetics centers. Collection of quantitative specimens from pregnancies, along with detailed clinical information, has become a critical resource to enable the biobanking program of this alliance to conduct a longitudinal cohort study of pregnancy outcomes. Availability and accessibility of the pregnancy specimens from the biobank of the alliance allow international and domestic investigators to perform translational research to uncover the molecular pathophysiology of pregnancy and its outcomes.

2869W

Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. *R. Moslehi[1], J.L. Mills[2], C. Signore[3], A. Kumar[1], X. Ambroggio[4], A. Dzutsev[5].* 1) Epidemiology and Biostatistics, Cancer Genomics Center, University at Albany, State University of New York, NY; 2) Epidemiology Branch, Eunice Kennedy Shriver National Institutes of Child Health and Human Development, NIH, Bethesda, MD; 3) Pregnancy and Perinatology Branch, Eunice Kennedy Shriver National Institutes of Child Health and Human Development, NIH, Bethesda, MD; 4) Bioinformatics and Computational Biosciences Branch, National Institutes of Allergy and Infectious Disease, NIH, Bethesda, MD; 5) Cancer Inflammation Program, Center for Cancer Research, National Cancer Institute, NIH, Frederick, MD.

Background Based on our clinical observations and subsequent genetic epidemiologic studies of DNA repair disorders, we previously identified associations between specific nucleotide excision repair (NER)/transcription gene mutations in the fetal genome and the risk of placental maldevelopment and preeclampsia, possibly due to impairment of Transcription Factor (TF)IIH-mediated functions in placenta. To identify the underlying mechanisms, we designed an integrative analysis of relevant transcriptome data sources containing gene expression arrays of fetal cell-derived tissue. This novel approach has enabled enrichment of causative mechanisms which underlie the previously-reported associations. **Methods** We conducted meta-analysis of gene expression patterns in placenta from four case-control studies of preeclampsia. A preeclampsia-specific gene list obtained from this meta-analysis was then interrogated in three relevant data sources [normal and time course placentas (i.e., placentas from first, second and third trimester pregnancies), hypoxic trophoblasts and XPD^{TTD} fibroblasts (i.e., cells predisposed to preeclampsia)]. **Results** Our meta-analysis of placental gene expression patterns revealed 419 differentially-regulated genes (136 downregulated and 283 upregulated) at false discovery rate(FDR)<0.05. Genes coding for TFIIH subunits and for components of RNA Pol-II complex were significantly-downregulated in preeclamptic placentas. Interrogation of the preeclampsia-specific gene signature through a filter constructed from the three relevant data sources identified *EGFR* and *ATF3* as key regulators of preeclampsia development. ATF3 upregulation was found as an upstream event to FLT1 and ENG induction. Integrative Gene Ontology(GO) and canonical pathway analyses identified EGF signaling as a key pathway involving the majority of differentially-regulated genes across all datasets. **Conclusion** Our overall integrative analysis suggested that, in the presence of hypoxia and oxidative stress, EGFR signaling deficiency, which can be caused by TFIIH impairment, results in ATF3 upregulation, inducing molecular mediators of clinical symptoms of preeclampsia such as FLT1 and ENG. EGFR- and ATF3-dependent pathways play prominent roles in cancer development. Thus, we propose that dysregulation of these cancer molecular pathways occurs in preeclampsia and delineate the relevance of TFIIH, providing etiologic clues which could eventually translate into a therapeutic approach.

2870T

TLR SNP T399I and early gestational age in a Wisconsin population of black newborn infants. *D. Pillers[1], M. Baker[1], S. Schrodi[2], L. Zyduck[1], J. DeValk[1], B. Pattnaik[1], S. Tokarz[1].* 1) Pediatrics, University of Wisconsin, Madison, WI, United States; 2) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI United States.

Background: Toll-like receptors (TLRs) are present in many cell types and serve as the first point of defense in the innate immune system by initiating the inflammatory cascade in response to infection. Single nucleotide polymorphisms (SNPs) are present in many TLR genes and have been associated with disorders of inflammation. TLR4 D299G and T399I SNPs are associated with increased susceptibility to infection from various pathogens. As inflammation can play a role in the development of preterm labor, TLR SNPs that alter the response to infection may have a different frequency in the preterm population. Objective: Screen a large, diverse population of Wisconsin newborns for TLR4 D299G and T399I SNPs and determine the genotype frequency in relation to gestational age (GA). Design/Methods: Anonymized DNA samples from 3095 infants were obtained in collaboration with the Wisconsin State Laboratory of Hygiene. Race was self-determined by patient as White (hispanic/non-hispanic), Black, Asian, and American Indian. TLR SNP assays (rs4986790, rs4986791) were purchased from Applied Biosystems Inc (ABI) and run on the ABI StepOnePlus™. Data were analyzed using StepOne™ software v2.2.2. Statistical analysis was carried out in collaboration with the Marshfield Clinic and the National Institute of Standards and Technology. Results: We did not find a significant correlation between the 299G and 399I alleles and early gestational age in the total population. However, when adjusting for racial background, we found that among Black infants the 399I allele frequency was 0% in term infants (37+ weeks GA), whereas it was 7.6% in preterm infants <37 weeks GA (P=0.039) and 7.7% in those <33 weeks GA (P=0.020). Conclusions: We show that in a population of Wisconsin infants the TLR4 SNP 299G did not associate with early gestational age. Our data are not consistent with current literature that show a correlation between the TLR4 SNP 299G allele and early gestational age. However, we did show an association between the 399I allele and early gestational age in a cohort of Black infants. Our data suggest that the racial background of an individual may affect the contribution of a specific TLR allele to preterm birth.

2871W

Copy number variation of RYR1 locus, which is involved in myomentrial contraction and relaxation, is associated with preterm births. *M. Liu[1], Y. Chen[1,2], N. Zhong[2,3].* 1) Center for Reproduction and Genetics, Suzhou Municipal Hospital Affiliated to Nanjing Medical University, Suzhou, Jiangsu, China; 2) Chinese Alliance of Translational Medicine for Maternal and Children's Health, China; 3) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

Preterm birth refers to a life birth before 37 gestational weeks. It is the leading cause of neonatal death. As well, it has been listed as one of the most common causes for under-5 children's mortality and morbidity. Although many risk factors have been identified to associate with preterm birth, genetic deficit is yet unknown. In this study, we have applied a genome wide study to assess the association of copy number variation (CNV) with preterm birth. A cohort of 26 preterm birth samples with gestational age of <32 weeks, for which, common known factors had been excluded, were analyzed with Agilent 1M microarray comparative genomic hybridization (aCGH), compared to the results of 49 normal samples. 639 CNV loci were found from preterm birth samples and 12 CNVs, mainly located on chromosome 1, 17, 19, were associated with preterm birth. Among these 12 loci, three were shown statistically significant. Variation within these loci includes duplication, homozygous and heterozygous deletion. Among 26 preterm birth samples, 3 (12%) at three loci and 8 (31%) at two loci had copy number variations simultaneously. A gene RYR1 located within the significant loci were noticed to be involved in calcium regulation in smooth muscle and most interestingly, in myomentrial contraction and relaxation that indicates this gene may contribute to the premature labor for preterm birth.

2872T

LncRNA Pathways Involved in Premature Preterm Rupture of Membrane (PPROM). *Q. Shi[1], X-C. Luo[1], Y. Gu[1], J. Pan[1], N. Zhong[1,2,3,4].* 1) Lianyungang Maternal and Children's Hospital, Lianyungang, China; 2) March of Dimes Global Network of Maternal and Infant Health, NY; 3) Peking University Center of Medical Genetics, Beijing, China; 4) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY.

Preterm birth (PTB) is a live birth delivered before 37 weeks of gestation (GW). About one-third of which result from the preterm premature rupture of membranes (PPROM). Presently, the pathogenic mechanism underlying PPROM is not yet clearly understood. In this study, we have investigated the differential expression of lncRNAs in placentas of PPROM, compared to controls, and their involvement in the pathogenic pathway of PPROM. A total number of 1954, 776, and 1050 lncRNAs were identified from placentas of PPROM (group A), which were compared to full term (FT) birth (group B), PTB (group C), and premature rupture of membrane (PROM) (group D), respectively. Instead of investigating the individual pathogenic function of each lncRNA involved in the pathogenic mechanism underlying PPROM, we have focused on investigating the metabolic pathways and their functions to explore what is the likely association and how they are possibly involved in the development of PPROM. Six groups, including up-regulation and down-regulation in the comparison of A vs. B, A vs. C, and A vs. D, of pathways were analyzed. Our results showed that 22 pathways were characterized as up-regulated 7 down-regulated in A vs. C, 18 up-regulated and 15 down-regulated in A vs. D, and 33 up-regulated and 7 down-regulated in A vs. B. Functional analysis showed pathways of infection and inflammatory response, ECM-receptor interactions, apoptosis, actin cytoskeleton, and smooth muscle contraction are the major pathogenic mechanisms involved in the development of PPROM, which opened a new avenue for further investigating of the regulation of lncRNAs in PPROM as well as PTB.

2873W

Functional validation of severe ciliopathies in a cohort of terminated pregnancies. *F.W. Verheijen, R. Oegema, L.M. van Unen, J.M. Verhagen, K.M. Diderich, M.L. van der Sterre, R.M. Hofstra, G.M. Mancini.* Clinical Genetics, Erasmus Univ MC, Rotterdam, Netherlands.

The wide application of prenatal ultrasound often leads to the detection of multiple congenital anomalies and consequently to termination of pregnancy (TOP) without a definitive diagnosis. Severe congenital anomalies at ultrasounds, such as (cranial) neural tube defects, ventriculomegaly, cerebellar and kidney dysplasia, skeletal abnormalities and heart-loop lateralization defects are frequent in defects of the primary cilium (ciliopathies), mostly recessively inherited, such as Meckel syndrome or hydrolethalus syndrome, but establishing a diagnosis in a fetus can be challenging. Ciliopathies have an incidence above 1 in 1,000 conceptuses and around 2500 genes are expected to be related to function and/or structure of the cilia. We therefore sought to develop a diagnostic screening assay and subsequent workflow for ciliopathies in fetuses of pregnancies, terminated on the basis of the ultrasound findings. Initially, immunostaining of the primary cilium and basal body in cultured skin fibroblasts proved to accurately diagnose patients with known ciliopathies (such as Joubert syndrome, Meckel syndrome, Rotatin mutation). Additionally, Shh-mediated Gli response was used as functional test for cilia-related pathways and confidently detected patients with known mutations, even with minor/no cilia structural anomalies. In a cohort of TOP fetal fibroblasts derived from fetuses with above mentioned malformations, we detected a high percentage of structural and functional cilia abnormalities. In fetuses with undefined ciliopathy, this approach complements the use of targeted NGS panels for ciliopathy genes, useful for both patient selection and confirmation of genomic variants. This approach also detects high recurrence risk pregnancies, offering the possibility of early prenatal diagnosis.

2874T

An Inverse Association Between Telomere Lengths and Gestational Age. *H. Naderi[1], K. Ryckman[2], J. Dagle[3], J. Murray[3].* 1) Carver College of Medicine, University of Iowa, Iowa City, IA; 2) Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA; 3) Department of Pediatrics, University of Iowa, Iowa City, IA.

Preterm birth (PTB; defined as birth before 37 weeks of completed gestation) is a complex disorder and is associated with significant neonatal morbidity and mortality rates. Genetic variations between preterm and term infants may provide insights into the pathways involved with PTB. One such genetic variation of interest is telomere length (TL). In humans, telomeres are non-coding TTAGGG repeats located at the end of chromosomes. With each cell division, telomeres shorten. Eventually shortened telomeres lead to cell senescence and serve as one marker of aging. Shorter TL has been associated with cardiovascular disease, cancer, dementia, arthritis, osteoporosis, and hypertension. The main determinants of TL are age, inheritance, and gender. Studies have shown conflicting results regarding a relationship between gestational age (GA) and TL. We tested the association between TL and GA in an initial 260 cord blood samples taken from a cohort of infants with GA between 24-41 weeks. Relative TL was measured using a qPCR method. Primers specific to telomeres (the target of interest) and to the 36B4 gene (a single copy gene on chromosome 12 used as a reference gene) were used to obtain a ratio of telomere to single copy gene length (T/S). This ratio was normalized to control for quantity of genomic DNA. Standard curves were produced as a means to check the efficiency of the assay. Relative quantification was used to analyze the data. The data was stratified into Group 1, GA less than 37 weeks (n=166), and Group 2, GA of 39 to 41 weeks (n=75). The average GA for the first group was 30.7 weeks while that of the second group was 40.0 weeks. The normalized T/S ratio of Group 1 is 3.69 while that of Group 2 is 2.18. This indicates that the TL of the preterm group (Group 1) is about 1.7 times longer than that of the term group (Group 2). A single variable regression of TL on GA yields a beta value of -0.093 (p-value = 0.033) indicating that a one unit increase in GA results in a TL loss of 0.093. The effect of TL on PTB outcome, as well as how TL changes with respect to gestational age, will now be examined with a larger sample size.

2875W

Case report of a pregnant woman with inherited thrombocytopenia associated with MYH9 mutation. *O. Samura[1,2], T. Mizunoe[2], S. Kunishima[3].* 1) Dept OB/GYN, Onomichi General Hospital, Onomichi, Hiroshima, Japan; 2) Dept OB/GYN, Kure Medical Center, Kure,Hiroshima,japan; 3) Dept of Advanced Diagnosis,clinical Research Center, Nagoya Medical Center, Nagoya,Japan.

[Introduction] MYH9 disorders are autosomal dominant macrothrombocytopenias with leukocyte inclusion bodies, characterized by giant platelets, thrombocytopenia, and Döhle body-like cytoplasmic inclusion bodies. May-Hegglin anomaly (MHA), caused by mutations of MYH9, is classified as a MYH9 disorder. Here, we report a rare case of a pregnant woman with familial congenital thrombocytopenia and confirmed MHA genetic diagnosis.[Case] A 31-year-old woman was diagnosed with thrombocytopenia when she was 20-year- old, but she did not receive treatment. Her familial history included low platelet levels in her paternal grandfather, father, and younger sister, and during her first pregnancy the platelet count was approximately 5.0 × 104/μL. However, she had a vaginal delivery without excessive bleeding. The patient was referred to our department for the management of her second pregnancy; however, due to her low platelet count (4.7 × 104/μL) and family history, congenital thrombocytopenia was suspected. A blood film revealed giant platelets, disparity in platelet size, and Döhle body-like cytoplasmic inclusion bodies, which were present in almost all segmented leukocytes. By immunofluorescent staining, abnormal myosin aggregation in granulocyte cytoplasm was suspected; therefore, we performed MYH9 testing with the consent of the patient and her father at gestational week 30. Genetic testing revealed an E1841K substitution in exon 38 of MYH9, and MHA was confirmed. Her platelet count ranged from 4-6 × 104/μL during her second pregnancy without the need for blood transfusion. She had a normal cephalic vaginal delivery of a 3148-g female with Apgar scores of 9 and 9 at 1 and 5 min, respectively, at gestational week 40. However, the second child also had thrombocytopenia and the same abnormal morphological findings as our patient. [Conclusion] In pregnancy, thrombocytopenia management in a patient with a family history should be carefully observed by peripheral blood smears because MYH9 disorders are not merely benign abnormalities but serious syndromic disorders that affect the kidneys, inner ears, and lens. Genetic diagnosis of MYH9 disorders is mandatory for accurate nonhematological complication prognosis and pregnancy management.

2876T

KIR2DL4 polymorphism and perinatal HIV-1 transmission. *C. Czarnecki[1], E. Rossnagel[2], S. Ramdahin[2], J. Embree[2], F. Plummer[1, 2], M. Luo[1, 2].* 1) National Microbiology Laboratory, Winnipeg, Canada; 2) University of Manitoba, Winnipeg, Canada.

Objectives: Killer cell immunoglobulin like receptor (KIR) 2DL4 is a framework gene located within the leukocyte receptor complex. It can be found in almost all individuals and is believed to play an important role during pregnancy. The only known ligand for KIR2DL4 is HLA-G, which is highly expressed at the maternal-fetal interface and also conducts an important role during pregnancy. We previously examined the effect of HLA-G polymorphism on perinatal HIV-1 transmission and observed HLA-G*01:03 to be associated with a decreased risk of transmission. In this study we analyzed the genetic polymorphism of KIR2DL4 and their epistatic interactions with HLA-G in perinatal HIV-1 transmission. Design: One hundred and ninety nine HIV-1 positive drug treatment naïve mothers enrolled in a mother-child HIV-1 transmission cohort in Nairobi, Kenya were genotyped for 2DL4 using a sequence-based typing method. Associations of major 2DL4 alleles and their epistatic interactions with HLA-G alleles were analyzed by statistical analysis using SPSS 13.0 for 274 mother-chair pairs. Results: A total of 20 2DL4 alleles were identified in this population, with 8 of them uncharacterized. Only 6 alleles had a phenotype frequency of greater than 5%. 2DL4*0010301 was the most common allele with an allele frequency of 42.71% and phenotype frequency of 62.81%. 2DL4*022 was associated with an increased perinatal HIV-1 transmission (p=0.009, OR:2.977 95% CI:1.276-6.942). Epistatic interactions between 2DL4*008 and HLA-G*01:01:02:01 increased the odds of perinatal transmission. The odds ratio for mothers with HLA-G*01:01:02:01 alone was 1.322 (p=0.28, 95% CI:0.826-2.115) and for mothers with 2DL4*008 alone was 1.157 (p=0.617, 95% CI:0.654-2.046). However, in mothers with both of these alleles the odds ratio of transmitting HIV-1 to their children increased to 2.011 (p= 0.035, 95% CI:1.044-3.876). Conclusions: KIR2DL4 polymorphism and its interaction with HLA-G can influence the risk of perinatal HIV-1 transmission.

2877W

Genome-Wide Association Study Identifies More Than 10 Loci Associated with Hypospadias. *F. Geller[1], B. Feenstra[1], L. van der Zanden[2], A. Nordenskjöld[3], L. Carstensen[1], I. van Rooij[2], I. Baranowska-Körberg[3], T. Schnack[1], M. Melbye[1].* 1) Statens Serum Institut, Copenhagen, Denmark; 2) Radboud University Medical Centre, Nijmegen, The Netherlands; 3) Karolinska Institutet, Stockholm, Sweden.

Hypospadias is a common birth defect in boys with a prevalence of about 4 cases in 1,000 births in Denmark. Severe cases require surgery and a previous study observed substantial familial aggregation with recurrence risk ratios of hypospadias for male twin pairs, first- and second-degree relatives of a hypospadias case, at 50.8, 11.6 and 3.27, respectively. This motivated us to perform a genome-wide association study based on 1,006 surgery confirmed cases identified from the National Hospital Registry and 5,468 controls, all genotyped on Illumina Omni chips. Twelve loci showed association at P<5/10-8, including the previously described locus on chromosome X close to DGKK. Combined, these genetic variants explain more than 10% of the variance in liability for hypospadias. Currently, independent replication study groups from Denmark, the Netherlands and Sweden are genotyped for these loci. Genes with general roles in development are overrepresented among the identified genomic regions.

2878T

Simultaneous uniparental disomy of chromosome 5 and 16; A case report. *H. Akar, D. Torun, Y. Tunca.* Medical Genetics, Gulhane Military Medical Faculty, Ankara, Turkey.

Uniparental disomy (UPD) describes the inheritance of both homologues of a pair of chromosomes from only one parent. Mechanisms of formation are trisomy rescue, gamete complementation, mitotic duplication, and post-fertilisation errors. Problems associated with UPD are placental or even fetal mosaicism mostly because of formation by trisomy rescue, homozygosity of autosomal recessively inherited mutations and aberrant genomic imprinting describing parent of origin dependent gene expression. A 29 years old healthy mother has come to our notice because of fetal loss. The family was followed up due to recurrent fetal losses before the last abortion and the mother revealed 46,XX, t(5;16)(q22;q24)mat,t(5;16)(q22;q24)pat karyotype. The mother stated as homozygous identical resiprocal translocation carrier. Unconsanguineous father had 46,XY normal constitutional karyotype. Parents of mother was consanguineous and carriying the same translocaticon as heterozygote. The last pregnancy was terminated at 7th week of gestation due to absence of fetal heartbeat. We have performed fetal cell culture to evaluate the cytogenetic abnormalities and karyotype analysis revealed 46, XX, t(5;16)(q22;q24),t(5;16)(q22;q24). We estimated that both of the reciprocal translocations inherited from the mother. This finding suggested that miscarriage resulted from maternal UPD of chromosome 5 and 16. As we know, this is the first report of simultaneous UPD of chromosome 5 and 16 that arise from homozygous identical resiprocal translocation carrier proband.

2879W

Proviral loads of human T-cell leukemia virus type 1 in the peripheral blood samples from carrier pregnant women. *N. Fuchi[1], K. Miura[1], T. Tsukiyama[1], D. Sasaki[2], N. Inokuchi[2], K. Yanagihara[2], S. Kamihira[2], H. Moriuchi[3], K. Yoshiura[4], H. Masuzaki[1].* 1) Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Medicine, Ngasaki, Japan; 2) Department of Laboratory Medicine, Nagasaki University Graduate School of Medicine, Ngasaki, Japan; 3) Department of Pediatrics, Nagasaki University Graduate School of Medicine, Ngasaki, Japan; 4) Department of Human Genetics, Nagasaki University Graduate School of Medicine, Ngasaki, Japan.

Objective: Human T-cell leukemia virus type 1 (HTLV-1) is causative virus of adult T-cell leukemia (ATL), which is still intractable against any medical treatments. As the proviral loads of HTLV-1 in blood samples is associated with the risk of ATL later, those in pregnant women may be also a risk factor of maternal-child transmitted infection. However, the information regarding proviral loads of HTLV-1 in carrier pregnant women is still unknown. Therefore, the aim of this study is to clarify the changes of proviral loads of HTLV-1 genome in the peripheral blood samples from carrier pregnant women before and after delivery. Material and Methods: When the pregnant woman was decided as positive or false positive by the first screening test by chemiluminescent enzyme immunoassay (CLEIA), blood sampling before delivery was performed at 28-34 weeks of gestation and the quantitative real-time PCR of HTLV-1 genome (30ng genomic DNA) was done as confirmation test. When HTLV-1 genome was detected in blood sample, pregnant woman was diagnosed as HTLV-1 carrier. After delivery, blood samples of HTLV-1 carrier pregnant women were obtained within 24 hours, and the proviral loads of HTLV-1 were measured by quantitative real-time PCR before and after delivery. PCR primers for HTLV-1genome were located at the pX region, while PCR primers for beta-globin as an internal control were on exon 2. HTLV-1 proviral load was calculated by the formula [(HTLV-1 pX copy number)/(beta-globin copy number/2)] ×10, 000 cells. Result: A total of 36 pregnant women were diagnosed as HTLV-1 carrier. Mean (minimum-Maximum) proviral loads were 83.3 (0.87-502.3)104 cells before delivery, while 34.27 (0-149.37)104 cells after delivery. HTLV-1 proviral loads before delivery was higher than those after delivery (Wilcoxon signed rank test, P<0.001). Conclusion: The proviral load of HTLV-1 genome in blood samples was decreased significantly after delivery, suggesting that the changes of HTLV-1 proviral load may reflect the changes of CD4-positive T cells in pregnant women.

2880T

Characterization of placenta-specific microRNAs in fetal growth restriction pregnancy. *A. Higashijima[1], K. Miura[1], H. Mishima[2], A. Kinoshita[2], O. Jo[1], S. Abe[1], Y. Hasegawa[1], S. Miura[1], K. Yamasaki[1], A. Yoshida[1], K. Yoshiura[2], H. Masuzaki[1].* 1) OB/GYN, Nagasaki Univ Hosp, Nagasaki, Japan; 2) Human Genetics Dept, Nagasaki Univ, Nagasaki, Japan.

Objective:The aim of this study was to characterize placenta-specific microRNAs in fetal growth restriction (FGR) pregnancy. Method:Placenta-specific miRNAs were identified by next-generation sequencing analysis. Subsequently, quantitative real-time RT-PCR was used to identify FGR placenta-specific miRNAs whose level of expression was significantly decreased in FGR placenta (n=45) compared with uncomplicated placenta (n=50). FGR pregnancy-associated, placenta-specific microRNAs were identified in maternal plasma after delivery at significantly decreased concentrations, and their circulating levels in maternal plasma was compared between FGR pregnancies (n=10) and uncomplicated pregnancies (n=10). Results: Out of the 10 placenta-specific microRNAs that we identified, seven placenta-specific microRNAs (hsa-miR-518b, hsa-miR-1323, hsa-miR-516b, hsa-miR-515-5p, hsa-miR-520h, hsa-miR-519d and hsa-miR-526b) from the chromosome 19 microRNA cluster were identified as FGR placenta-specific microRNAs. Four FGR placenta-specific microRNAs (hsa-miR-518b, hsa-miR-1323, hsa-miR-520h and hsa-miR-519d) were confirmed as FGR pregnancy-associated, placenta-specific miRNAs, but their circulating levels in maternal plasma showed no significant differences between FGR pregnancy and uncomplicated pregnancy. Conclusion: Our data suggest that reduced expression in placenta of certain FGR placenta-specific miRNAs is associated with FGR, and that the discrepancy between expression in FGR placenta and their circulating levels in maternal plasma will be crucial to understanding how placenta-specific microRNAs are released into the maternal circulation.

2881W

Heritability of cardiovascular diseases in a preeclampsia family cohort. *L.C.V. Thomsen[1,2], P.E. Melton[3], K. Tollaksen[4], I. Lyslo[4], P. Solberg[5], L.T. Roten[6,7], A.S. Gundersen[6,7], M.L. Odland[7], K.M. Strand[7], O.K. Nygård[2,8], C. Sun[1], A-C. Iversen[7], R. Austgulen[7], E.K. Moses[3], L. Bjørge[1,2].* 1) Department of Obstetrics and Gynecology, Haukeland University Hospital; 2) Department of Clinical Science, University of Bergen; 3) Centre of Genetic Origins of Health and Disease, University of Western Australia; 4) Department of Obstetrics and Gynecology, Stavanger University Hospital; 5) Department of Obstetrics and Gynecology, Levanger Hospital; 6) Department of Laboratory Medicine, Children's and Women's Health, Faculty of Medicine, Norwegian University of Science and Technology; 7) Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology; 8) Department of Heart Disease, Haukeland University Hospital.

Purpose: Preeclampsia (PE) is a complex genetic disease of pregnancy with potentially severe outcomes for mother and fetus. Women who develop PE have been observed to have increased risk (2-8x) of developing cardiovascular diseases (CVD) later in life. We aimed to estimate the heritability of CVD phenotypes in well-characterized families with an acknowledged predisposition to PE. Methods: Using information contained in the Norwegian Preeclampsia Family Biobank, a cohort composed of families with an increased occurrence of PE, we classified 496 participants from 138 families according to predefined CVD phenotypes; a) risk factors for CVD (treated hypertension and/or hypercholesterolemia), b) established atherosclerotic vascular disease (subtypes: angina, acute myocardial infartion, stroke, aneurism, thrombosis), or c) no diagnosis of CVD. Heritabilities were calculated using a variance components procedure implemented in the SOLAR software with the inclusion of covariate age, sex, and their interactions. Results: Average age of the participants was 46.9 (range 18-87) years. Heritability estimates (h2r) were statistically significant for PE (h2r=0.54±0.3, p=0.02), non-pregnancy related hypertension (h2r=0.62±0.1, p=2.1×10-3) and hypercholesterolemia (h2r=0.55±0.3, p=0.03). When only the presence or absence of any of the CVD phenotypes was assessed the heritability of CVD was highly significant (h2r=0.31±0.12, p=1.9×10-3). The trait heritability was not significant for the other CVD phenotypes examined. Conclusion: This study represents the first link between hereditary development of PE and specific CVD traits identified in a family cohort. Due to the relative young age of these participants most of the cohort would not yet have experienced manifest atherosclerotic disease in form of acute myocardial infarction or stroke, but rather the established risk factors of CVD. Further identification of genetic variants specific to the affected families will avail us to focus on certain genetic areas potentially regulating biologic functions implicated in development of both CVD and PE.

2882T

Transcriptional profiling analyses of RNA in uncultured amniotic fluid using GeneChip® PrimeView™ Human Gene Expression Array. *D. Cha[1,2], S.H. Sung[2], S.R. Sung[2], M. Chin[2], J. Park[2], S. Kang[2], K. Kang[2], S. Lyu[1].* 1) Dept OB/GYN, Kangnam CHA Hosp, Seoul, South Korea; 2) Genetics Laboratory of Fertility Center.

Amniotic fluid is the only body fluid in direct contact with the the fetal oropharynx, lungs, gastrointestinal tract, skin, and urinary system and amniotic fluid cell-free fetal RNA(AF cffRNA) can provide biological information on developing fetal organ system. The aim of this study is to clarify AF cffRNA profile of normal fetuses in three different gestational periods, 16~17 weeks (group 1), 25~26 weeks (group 2), and 36~38 weeks (group 3). The total 13 AF samples, 5 of group 1, 4 of group 2 and 4 of group 3, were collected. Total RNA was extracted from 10 mL of the amniotic fluid supernatant using the QIAamp Circulating Nucleic Acid Kit. The extracted RNA was hybridized to GeneChip® PrimeView™ Human Gene Expression Array to compare gene expression in amniotic fluid supernatant samples. The data were analyzed with Robust Multi-array Analysis using Affymetrix default analysis settings and global scaling as normalization method. The normalized, and log transformed intensity values were then analyzed using GeneSpring 12.5. In group 1, total 3289 transcripts were identified. We were confirmed 532 transcripts that have been known to be associated with fetal development through a search of pathway analysis software and Medline. We found 1,977 genes to be differentially expressed between group 1 and group 2. Among these genes, 27 genes increased and 1950 genes decreased in group 2. Also, we found 2,039 genes to be differentially expressed in group 3 compared to group 1. 48 out of 2,039 genes increased and 1475 genes decreased in group 3. Our result presented cffRNA profiles in amniotic fluids of pregnant women with normal fetal development in three different gestational periods. This study might contribute to the further studies for the discovery of biomarkers for abnormal fetal growth.

2883W

Recurrent enlarged nuchal translucency: prenatal presentation of a familial 15q26.2→qter deletion syndrome. *R. Reiss[1], D. Ahern[1], M. Sandstrom[2], L. Wilkins-Haug[1].* 1) Center for Fetal Medicine and Prenatal Genetics, Dept of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA; 2) Dept of Pathology, Brigham and Women's Hospital, Boston, MA.

Objective: To elucidate familial recurrence of enlarged nuchal translucency (NT) with chromosomal microarray and FISH techniques. Methods: A patient presented in each of 3 pregnancies with enlarged NT (4.7 mm, 4.0 mm, 7.1 mm). The first pregnancy was terminated when cardiac defect was suspected at 13.1 wks. Products of conception (POC) were karyotyped as 46,XX. In the next pregnancy, chorionic villus sampling at 12 wks showed karyotype 46,XX, and no deletion at 22q11.2. Ultrasound at 17 wks identified talipes. Fetal echocardiogram was normal at 19 wks. Pregnancy continued, complicated by IUGR. Multiple anomalies were identified after birth (micrognathia, malrotated kidney, talipes, brachydactyly). Postnatal growth lag prompted microarray testing at 5 months. In a 3rd pregnancy, septated NT at 11.7 wks again led to pregnancy termination. Chromosome analysis of POC showed 46,XY. Microarray was performed on fetus. Karyotype and FISH were performed on the mother. Results: 6.0 SNP microarray on the liveborn girl showed a 6.53Mb deletion at 15q26.2-q26.3, confirmed by 15q subtelomeric FISH. CGH microarray on fetal DNA from pregnancy 3 showed a 6.49Mb loss of 15q26.2→qter, consistent with the deletion seen in the affected child. FISH using a 15q subtelomere probe confirmed microarray findings. Subtelomeric 15q FISH analysis, performed retrospectively on a destained slide of cultured metaphases of POC from pregnancy 1, confirmed 15q26.2-26.3 deletion. Maternal blood chromosome analysis was normal, 46,XX. FISH using BAC probes from 15q26.2 and 15q26.3, with control probe to 15q11.2, showed no deletion or translocation of the 15q26.2 region. Studies on father and CVS from a recent 4th pregnancy are pending. Conclusions: Terminal microdeletion 15q26.2→qter, an emerging syndrome of pre- and postnatal growth lag and anomalies, can present in 1st trimester with large NT. Our case highlights the importance of offering chromosomal microarray as well as conventional karyotyping when large NT is identified. Retrospective performance of microarray or FISH on DNA/cells from fixed samples can help document a familial syndrome. Based on review of the literature, most cases of terminal 15q deletion are de novo. This is the 2nd report of a familial case. We postulate that the father may carry a chromosome rearrangement involving the 15q terminal region. Alternatively a small inversion in the region of the breakpoints, or germline mosaicism might be present in a parent.

2884T

Gene expression differences between preeclamptic and healthy placentas - an RNA sequencing study. *T. Kaartokallio[1], A. Cervera[2], S. Hautaniemi[2], J. Kere[1,3,4,5], H. Laivuori[1,6].* 1) Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland; 2) Systems Biology Laboratory, Institute of Biomedicine and Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 3) Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden; 4) Molecular Neurology Research Program, Research Programs Unit, University of Helsinki, Finland; 5) Folkhälsan Institute of Genetics, University of Helsinki, Helsinki, Finland; 6) Department of Obstetrics and Gynaecology, Helsinki University Central Hospital, Helsinki, Finland.

Objectives: Preeclampsia (PE), a common pregnancy disorder originating from impaired placental development, is responsible for morbidity and mortality to both mother and child. In the present study, we utilized RNA sequencing to identify genes that are differently expressed between preeclamptic and healthy placentas. **Methods:** Placental RNA of nine preeclamptic women and nine healthy controls was sequenced in pools of three with Illumina chemistry. The PE pools 1, 2 and 3 consisted of placental samples from 38-39, 34-36 and 33 weeks of gestation, respectively, and the control pools from 38-39 weeks of gestation. All women delivered by C-section without labor. The data was processed and analyzed with appropriate tools including Tophat, Cuffdiff and CummeRbund. The two groups were tested for differential gene expression by comparing all the case pools as well as subsets of the case pools against the control pools. The results will be validated by qPCR in a sample set of 20 cases and 19 controls including the samples utilized in the RNA sequencing. In this sample set, the cases and controls do not differ significantly for maternal age or body mass index, gestational age, mode of delivery or sex of the child. **Results:** Initially, a total of 58 genes were found to be differentially expressed between PE and control groups. After excluding genes with inconsistent expression pattern within the groups, hemoglobin genes, and genes located in Y chromosome, as the fetal sex distribution between the groups was not equal, we were left with 40 genes. Of these, 15 will be subsequently validated by qPCR. The genes selected for the validation are involved e.g. in immunological processes, placental development and function, invasion, angiogenesis and vasodilation, all of which are processes relevant for the development of PE. Some of these genes have also been previously associated with autoimmune or renal diseases. In addition to our main goal of identifying expression differences between preeclamptic and healthy placenta, we wish to emphasize the importance of taking account of gestational age when studying placental gene expression in pregnancy complications. Therefore, in the qPCR validation, we also aim to demonstrate the effect of gestational age on expression of certain genes. **Conclusions:** By comparing transcriptomes of preeclamptic and healthy placentas, we were able to identify genes that were differentially expressed between the two groups.

2885W

Towards a Mouse Model of Thrombocytopenia with Absent Radius (TAR) Syndrome. *V.L. Horner[1], A. Dodd[1], A. Long[1], C.L. Martin[2], T. Caspary[1].* 1) Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA.

TAR syndrome is characterized by low blood platelet counts (thrombocytopenia) and absence of the radius bone in each forearm. Thrombocytopenia can lead to easy bruising and/or hemorrhage in infancy, which may become less severe over time. The absence of radius bones results in shortened forearms. TAR syndrome is inherited in an autosomal recessive manner and is due to changes in the gene *RBM8A*. Compound inheritance of one *RBM8A* null allele (most often as a large deletion) and one *RBM8A* hypomorphic allele causes TAR syndrome. At present, two hypomorphic alleles have been described. Somewhat surprisingly, both alleles contain the minor allele of a single nucleotide polymorphism (SNP) in a regulatory region of *RBM8A*. These two regulatory SNPs lower *RBM8A* transcription *in vitro*. Further, the level of the *RBM8A* gene product, Y14, is reduced in individuals who carry the SNPs compared to wild-type controls. Together, these results lead to the hypothesis that there is a dose-effect phenomenon, in which a null allele together with a hypomorphic allele brings Y14 levels below some critical threshold to cause TAR syndrome. To determine where, when and how much *RBM8A* production is required during embryonic development, we are generating mouse alleles in which the endogenous *RBM8A* locus is under the control of a tetracycline-dependant transactivator. This system will enable us to control the amount of Y14 protein available during embryogenesis, by varying the amount of tetracycline fed to pregnant mothers. We are generating two strains of mice: the first strain will have the transactivator (TetOn) under the control of the endogenous *RBM8A* promoter. The second strain will have the tetracycline-responsive promoter (tetO) driving *RBM8A* expression. The presence of tetracycline allows the transactivator to bind and drive *RBM8A* transcription. There is a linear relationship between the amount of tetracycline and the amount of transcription, enabling us to define the critical dosage threshold of Y14 that causes TAR syndrome phenotypes as well as the dosage that causes lethality. Ultimately, this system will also allow us to examine the temporal and spatial requirements for *RBM8A* during development, by providing tetracycline to pregnant mothers at various stages of embryogenesis and by driving the transactivator with promoters specific for various tissues. These experiments are the first step to defining the therapeutic level of Y14 for TAR syndrome treatment.

2886T

A single mutant *HTT* allele is sufficient to elicit early alterations to the brain in a knock-in mouse model of Huntington's disease. *M. Kovalenko[1], A.J. Milnerwood[2], S. Tappan[3], J. St.Claire[1], J.R. Guide[1], R.C. Switzer III[4], L.A. Raymond[2], J.-M. Lee[1], V.C. Wheeler[1].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 2) Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada; 3) Microbrightfield laboratories, MBF Labs, Williston VT, USA; 4) NeuroScience Associates, Knoxville, TN, USA.

Huntington's disease (HD) is a dominantly inherited neurodegenerative disease caused by the expansion of a CAG repeat in the *HTT* gene encoding huntingtin (Htt). To increase the chance of identifying early responses to mutant Htt that are relevant to the disease process in patients we are using accurate genetic knock-in mouse models of the HD mutation to delineate phenotypes that result from the expression of a single allele's worth of full-length mutant Htt as occurs in the majority of HD patients.

Here, as part of our ongoing phenotyping effort we report the results of neuropathology and electrophysiology studies in heterozygous $Hdh^{Q111/+}$ mice. We assessed neuropathology in multiple brain regions in $Hdh^{Q111/+}$ mice at 10 and 18 months of age. To probe potential underlying disease mechanisms we also assessed a cohort of 10 month mice that had been fed a high-fat diet. Immunohistochemistry revealed reduced levels of DARPP-32 in $Hdh^{Q111/+}$ striata at 10 months of age, independent of diet, and widespread increased myelin basic protein (MBP) staining in $Hdh^{Q111/+}$ at 18 months. Stereology revealed subtly reduced volumes of striatum, caudate-putamen, cerebrum, cortex and nucleus accumbens and subtly reduced cortical thickness at 10 months of age. Both cortical volume and thickness were decreased to a greater extent in $Hdh^{Q111/+}$ mice on a high-fat diet than on a control diet suggesting that a high-fat diet provides a sensitized background to these measures. Medium-spiny striatal neurons (MSNs) are particularly vulnerable to the effects of the *HTT* CAG mutation. To explore a potential functional deficit at the level of the MSN, as indicated by reduced DARPP-32, we used acute slice electrophysiology to probe very early alterations in glutamatergic transmission from the cortex to the MSNs. As early as 1.5 months of age $Hdh^{Q111/+}$ mice exhibited a slowed decay of evoked NMDA receptor-mediated postsynaptic currents compared to wild-type controls. At 5 months of age $Hdh^{Q111/+}$ mice exhibited elevated extrasynaptic NMDA currents as well as altered AMPA receptor-mediated spontaneous postsynaptic currents.

Together, these data strongly indicate that the expression of endogenous levels of full-length mutant Htt expressed from a single expanded allele is sufficient to elicit phenotypes consistent with neuronal dysfunction at an early age.

2887F

From Chromosomal Translocation to Functional Analysis, Identification of ATG4C in Cleft Palate Malformation. *L.J. Rochard, T. Hyos, M. Talkowski, J. Gusella, C. Morton, E.C. Liao.* Massachusetts General hospital, BWH, Harvard Medical School.

Orofacial defects such as cleft lip and palate (CLP) are the most common structural congenital anomalies, affecting 1 out of 700 births. Elucidating the genetic basis of these orofacial clefts is essential to identify risk loci and develop new therapeutic and preventive measures. In an ongoing gene discovery effort, the Developmental Genome Anatomy project (DGAP) has developed whole genome sequencing strategies to characterize genes contributing to such human congenital anomalies. Through this methodology, we discovered a new candidate gene for CLP, ATG4C that encodes an autophagy related cysteine protease. The isolated deletion of ATG4C was identified in a 12 year old Caucasian boy presenting a CLP. Previous studies in our lab have demonstrated conservation of the fundamental molecular pathways and homology of primary palate formation between mammals and zebrafish. Therefore we utilized the model organism zebrafish to study and validate the role of ATG4C in palate development. By whole mount in situ hybridization, we determined that the spatiotemporal expression of atg4c is compatible with a role in craniofacial development. Additionally, knockdown studies performed with morpholino injections resulted in both a shortened and cleft palate, suggesting a strong role for atg4c as the causal gene in the human case of CLP. Our current work is focusing on a mechanistic analysis of ATG4C in craniofacial morphogenesis and autophagy. Our future studies will track migrating cranial neural crest cells and help determine if the failure of upper jaw morphogenesis in atg4c mutants is due to a defect in cell migration, proliferation or convergence extension mechanisms. We also plan to assess effects of atg4c loss of function in a GFP-LC3 (a component of autophagosome) transgenic line. To this end, we are creating mutants for this gene by using the targeted mutagenesis method, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). This study goes beyond functional annotation of ATG4C to include detailed biological analysis of gene function, and highlights the utility of the zebrafish model in high throughput and reverse genetics approaches that are important in functional genomics projects.

2888W

***Trim37*-deficient mice recapitulate several features of the multi-organ disorder Mulibrey nanism.** *K.M. Kettunen[1, 2, 3], R. Karikoski[4], R.H. Hämäläinen[2], T-T. Toivonen[1, 2, 3], H. Jalanko[5], A-E. Lehesjoki[1, 2, 3].* 1) Folkhälsan Institute of Genetics, Helsinki, Finland; 2) Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Finland; 3) Neuroscience Center, University of Helsinki, Finland; 4) Central Hospital of Tavastia, Hämeenlinna, Finland; 5) Hospital for Children and Adolescents, Helsinki University Central Hospital, Finland.

Objective: Mulibrey nanism (MUL) is a rare autosomal recessive multi-organ disorder caused by mutations in the *TRIM37* gene. The main symptoms include growth restriction, cardiopathy, infertility, increased risk for tumors, metabolic syndrome and type 2 diabetes. The disease mechanisms and the physiological function of TRIM37 are unknown. We characterized thoroughly a *Trim37* knock-out mouse ($Trim37^{-/-}$) to gain understanding of the pathogenesis of MUL. *Materials and Methods:* A congenic C57BL/6JOlaHsd $Trim37^{-/-}$ mouse strain was produced by use of BayGenomics gene-trap ES cells. *Trim37* knock-out mice and wild-type littermates were monitored for fertility and weight development. CT scans were performed to measure the skeletal size of adult mice. Samples of brain, heart, lung, liver, pancreas, spleen, intestine, kidneys, adrenal glands, ovaries / testes, rib, femur and white adipose tissue were collected at the age points 1, 4, 6 and 18 months. Histological samples were stained with hematoxylin and eosin, Oil red O for lipids and specific antibodies when appropriate. Basic laboratory measurements were determined from serum samples collected after a 24h fast in six-month-old mice. *Results and discussion:* $Trim37^{-/-}$ mice were found to have significantly smaller skull size compared to wild-type mice, despite of similar weight development in both genotypes. $Trim37^{-/-}$ mice were infertile due to degeneration of testes and ovaries. On histological level both male and female gonads showed germ cell aplasia and Leydig cell hyperplasia. Common pathological findings in 1.5 year old mice included cardiomyopathy, hepatomegaly and fatty liver, splenomegaly and various tumors. Fat accumulation in liver started already at the age of 16 weeks and increased with age. Adult six-month-old $Trim37^{-/-}$ mice had elevated fasting blood glucose values (4,71 ± 0,7mmol/l vs. 3,79 ± 0,5 mmol/l, p = 0,0047) and low fasting serum insulin values (0,37 ± 0,1ng/ml vs. 0,59 ± 0,3ng/ml, p = 0,065), compared to wild-type mice. Thus $Trim37^{-/-}$ mice have disturbances in glucose metabolism and most likely impaired glucose tolerance or type 2 diabetes. *Conclusion:* $Trim37^{-/-}$ mice recapitulate several features of the human disease phenotype, including growth restriction, infertility, risk of various tumors, cardiomyopathy, disturbances in glucose and lipid metabolism and fatty liver and thus provide a good model to study disease pathogenesis related to TRIM37 deficiency.

2889T

An in vivo mouse model to study the phosphorylation of FMRP. *M. Santoro, S.T. Warren.* Department of Human Genetics Emory University Atlanta, GA.

Fragile X syndrome, the most common form of inherited intellectual disability, results from a lack of Fragile X Mental Retardation Protein (FMRP). FMRP is a selective mRNA-binding protein and plays a key role in synaptic plasticity. In neurons, FMRP represses translation of its target mRNAs but allows translation to proceed upon receipt of specific synaptic signals. In the hypothalamus, activation of mGluR receptors causes FMRP to derepress translation of its target mRNAs, resulting in a burst of local protein synthesis, internalization of AMPA receptors, and long term depression. In vitro experiments indicate that FMRP's activity is regulated by the phosphorylation status of a specific serine (S499 in mice; S500 in humans). In order to more fully investigate the translational regulation of FMRP we have developed a knock-in mouse model with a Ser499Ala mutation. This results in an amino acid replacement that structurally resembles unphosphorylated FMRP. Indeed, using an antibody raised against phosphor-FMRP, no signal is observed on Western blots from Ser499Ala mouse tissue, although normal levels of FMRP are seen with the conventional antibody. Phenotypic studies are underway but preliminary data suggest that the Ser499Ala mouse may not be a complete phenocopy of the FMRP null mouse.

2890F

A systematic genome-wide knockout generation and analysis of zebrafish protein-coding gene function. *C.M. Dooley[1], E.M. Busch-Nentwich[1], R.N.W. Kettleborough[1], C. Scahill[1], I. Sealy[1], R.J. White[1], J.C. Collins[1], N. Wali[1], C. Herd[1], R. Gibbons[1], S. Carruthers[1], A. Hall[1], R.C. Clark[1], Z. Pusztai[1], M. Niemi[1], F. van Eeden[2], J.C. Barrett[1], D.L. Stemple[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) MRC-CDBG/Department of Biomedical Science, The University of Sheffield, United Kingdom.

Although the human genome sequence was completed a decade ago less than half of the identified vertebrate genes have been assigned a function. Detailed investigations involving model organisms have played a fundamental role in connecting genotype to phenotype and will continue to do so. The ability to thoroughly investigate the role of a gene in various biological processes greatly depends on loss of function analysis which has traditionally been carried out, in vertebrates, on a gene-by-gene basis. The zebrafish is a genetically tractable organism with an external, transparent and rapidly developing embryo offering an in toto model of the complex cellular processes required for the formation of all major vertebrate organs. In recent years it has moved beyond being a model of early development as 70% of all human genes are represented by at least one zebrafish ortholog. This has helped to make zebrafish a first stop choice for human geneticists aiming to assign function to human orthologs or clinicians hoping to develop a model for a particular disease. Enabled by a completed and annotated zebrafish reference genome sequence, high-throughput sequencing and efficient chemical mutagenesis, we describe the active Zebrafish Mutation Project (ZMP) at the Wellcome Trust Sanger Institute which aims to identify and phenotype disruptive mutations in every zebrafish protein-coding gene. Thus far we have identified potentially disruptive mutations in more than 48% of all of the 26,000 known zebrafish protein coding genes. We have developed a multi-allelic phenotyping scheme to efficiently assess the effects of each allele during embryogenesis and have analyzed the phenotypic consequences of more than 1600 alleles. Our phenotyping scheme is also adaptable to phenotypic analysis beyond embryogenesis and we are currently expanding our analysis to encompass a wider array of phenotypes as we set forth to fully functionally annotate the zebrafish genome. To accompany our chemical mutagenesis we have also developed the Cas9/CRISPR system to allow for single as well as multiplexed targeted bespoke knockouts. A further goal is to use our rich collection of knockouts and knockout generation capabilities to model human disease and to work towards the complete functional annotation of the entire zebrafish genome. To this end, all knockout alleles and data are immediately made available to the community via our website http://www.sanger.ac.uk/cgi-bin/Projects/D_rerio/zmp.

2891W

Accessing mouse model data for human disease research. *J.T. Eppig, S.M. Bello, H. Dene, A. Anagnostospoulos, H. Onda, R. Babiuk, M.N. Knowlton, M. Tomczuk, C.L. Smith.* Jackson Laboratory, Bar Harbor, ME.

The laboratory mouse is the primary model for human disease based on its extensive genetics, fully-sequenced genome, and large-scale mutagenesis programs that have created ENU (point mutations) and knock-out (null) mutations covering most of its genome. In addition, programs to develop genetically-defined variant populations mimicking human population variation provide new means to study quantitative traits and complexly inherited syndromes. This genomic knowledge, coupled with increasing phenotyping precision and micro-technologies, add exquisite detail to knowledge about morphological and physiological variation in mouse.

The Mouse Genome Informatics (MGI, www.informatics.jax.org) database and resources provide a roadmap to discovery for biologists and clinicians seeking to use the extensive genetic, phenotypic, and disease model data available to correlate mouse phenotypes and human disease. MGI holds data on more than 750,000 mutant allelic variants in 21,400 genes. More than 250,000 phenotypic annotations to genotypes are curated using the Mammalian Phenotype Ontology. Mouse models are annotated to over 1,240 unique human diseases in OMIM.

MGI is developing new interfaces to human disease information to foster use by clinical and human genetics researchers. These interfaces will provide easy access to mouse models via human genes, human genome locations, human disease symptoms, and human anatomy terms. Our aim is to reduce the barriers to use of mouse data and increase the synergy derived from comparative genomics and comparative phenotyping. This will facilitate mining data for common pathways, disease variants, and potential therapeutic interventions. We describe these new human-centric entry points to MGI data and present examples of insights derived from examining human diseases in light of mouse models. We demonstrate the ease of obtaining mouse disease models from repositories, the path to reference collections, and the search for collaborators. Supported by NIH grant HG000330.

2892T

Reducing variability in variation data using a common genotype-phenotype model. *N.L. Washington[1], M. Brush[2], A. Bandrowski[3], C. Borromeo[4], K. Eilbeck[7], J. Espino[4], J. Grethe[2], A. Gupta[3], H. Hochheiser[4], S. Hoffmann[2], S.E. Lewis[1], L. Lui[3], M. Martone[3], C.J. Mungall[1], P. Robinson[5], D. Smedley[6], C. Torniai[2], N. Vasilevsky[2], M.A. Haendel[2].* 1) Genomics Division, Lawrence Berkeley National Lab, Berkeley, CA; 2) Library and Dept. of Medical Informatics and Epidemiology, Oregon Health Sciences University, Portland, CA; 3) Center for Research in Biological Systems, University of California San Diego, La Jolla, CA; 4) Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA; 5) Institut für Medizinische Genetik und Humangenetik, Charité - Universitätsmedizin Berlin, Berlin, Germany; 6) Wellcome Trust Sanger Institute, Cambridge, UK; 7) University of Utah, UT.

Investigating the basis of disease requires correlating and integrating genotype, phenotype, and environmental factors. Ideally, we could leverage the depth of knowledge available from model systems. However, choosing an appropriate and sufficient model organism (e.g. genotype) for the study of a disease is a challenge, as it often requires an understanding of the organism-specific vocabularies, nomenclature guidelines, and a whole host of nuances attributed to laboratory animal husbandry, breeding, and maintenance. Though the volume of curated genotype-to-phenotype data is steadily growing, our ability to query across different organisms and data resources is limited. This is due, in part, to a lack of a common genotype and phenotype reporting format, and results from the variety of ways in which this data is reported, including free text, organism-specific vocabularies, nomenclatures, syntaxes. The data is further complicated because phenotypes may be linked to different aspects of a genotype, such as an allele, a QTL, or the complete genome. Despite efforts to standardize human variation through proposed standards such as HGVS many alternatives still exist, leading to difficulties when integrating human phenotype data from different database resources. A first step toward the integration of organism genotypes (including human and model systems) is to translate them to a common genotype model. As part of the Monarch Initiative, we have created a pan-species genotype representation that allows users to access human variation and model organism genotype-to-phenotype data in a uniform way. This permits cross-species querying by genotype components without needing to know about a specific resources' nomenclature, and is the first step in being able to aggregate and attribute phenotype data to different components of a genotype. This standardization enables translational genotype-phenotype hypotheses and analysis of treatment options, and will make model system data available for rare and undiagnosed disease diagnosis and clinical decision support tools.

2893F

MYBPC1 mutations impair skeletal muscle function in zebrafish models of arthrogryposis. *D.M. Alvarado[1], K. Ha[1], J.G. Buchan[2], K. McCall[1], A. Vydyanath[3], P.K. Luther[3], M.I. Goldsmith[4], M.B. Dobbs[1,5], C.A. Gurnett[1,2,4].* 1) Orthopaedic Surgery, Washington Univ, St Louis, MO; 2) Neurology, Washington Univ, St. Louis, MO; 3) Heart and Lung Institute, Imperial College, London, UK; 4) Pediatrics, Washington Univ, St. Louis, MO; 5) St. Louis Shriners Hospital for Children, St. Louis, MO.

Myosin binding protein C1 (MYBPC1) is an abundant skeletal muscle protein that is expressed predominantly in slow twitch muscle fibers. Human MYBPC1 mutations are associated with distal arthrogryposis type 1 and lethal congenital contracture syndrome (LCCS4). Because the function of MYBPC1 is incompletely understood, the mechanism by which human mutations result in contractures is unknown. Here, we demonstrate that *mybpc1* is required for embryonic motor activity and survival in a zebrafish model of arthrogryposis using antisense morpholino knockdown. *Mybpc1* morphant embryos have severe body curvature, cardiac edema, impaired motor excitation and are delayed in hatching. Myofibril organization is selectively impaired in slow skeletal muscle and sarcomere numbers are greatly reduced in *mybpc1* knockdown embryos. To evaluate the effects of human distal arthrogryposis mutations, *mybpc1* mRNAs containing the corresponding human W236R and Y856H MYBPC1 mutations were injected into embryos. Dominant-negative effects of theses mutations were suggested by the resultant mild bent body curvature, decreased motor activity, myofibril disorganization, as well as impaired overall survival compared to overexpression of wild-type RNA. These results demonstrate a critical role for *mybpc1* in slow skeletal muscle development and establish zebrafish as a tractable model of human distal arthrogryposis.

2894W

Scaling up: Integrating high throughput mouse phenotyping data with additional genomic resources for gene discovery. *C.L. Smith, H. Dene, R. Balderelli, S. Gianatto, K. Forthover, J. Kadin, J. Richardson, J.T. Eppig, Mouse Genome Informatics Staff.* Mouse Genome Informatics, Jackson Laboratory, Bar Harbor, ME.

The mouse is the most commonly studied model organism for understanding gene function in human disease and development. Comparative genetics and genomics can assist in prioritization of candidate genes for human disease, and model organisms can be used for confirmation or validation of gene function. However, comprehensive functional and phenotypic annotation of mouse genes is incomplete in many instances and mouse phenotype data are available for only 7770 of of the estimated 33,735 protein coding and RNA only coding genes.

The International Mouse Phenotyping Consortium (IMPC) is gearing up to collect phenotype data on mice made from the large collection of knockout alleles in embryonic stem cells available from the International Knockout Mouse Consortium. The goal is to provide standardized phenotypic data derived from the same battery of tests for every gene in the mouse genome. Mouse Genome Informatics (MGI: www.informatics.jax.org) has already established pipelines to the WTSI and Europhenome pilot phenotyping projects and has developed the infrastructure to integrate and display these initial sets of data with all of the all other genetic and functional information available about these genes in mouse.

Phenotypic representation of genetically engineered mice integrated with other biological data at MGI will aid in model building, understanding biochemical pathways, and determining the underlying mechanisms of human genetic disease. We will show web based query results utilizing query forms designed to search phenotypic, biochemical function and process, subcellular location, expression, pathways and sequence data that may assist in finding novel candidate genes for human disease. Supported by NIH grant HG000330.

2895T

Translational Modeling of Calpain-5 Vitreoretinopathy Mechanisms in Mice. *K.J. Wert[1, 2], J.M. Skeie[3, 4], S.H. Tsang[1, 2], V.B. Mahajan[3, 4].* 1) Department of Ophthalmology, Columbia University Medical Center, New York, NY; 2) Institute of Human Nutrition, College of Physicians and Surgeons, Columbia University, New York, NY; 3) Omics Laboratory, University of Iowa, Iowa City, IA; 4) Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA.

A wide range of human diseases including cancer, multiple sclerosis, Alzheimer's disease, cataract, diabetes, and muscular dystrophy have implicated a role for calpains in their pathogenesis. We previously identified *CAPN5* (calpain-5) as the causative gene for retinal degeneration and autoimmune uveitis in human patients with Autosomal Dominant Neovascular Inflammatory Vitreoretinopathy (ADNIV). To understand the molecular and genetic mechanisms of ADNIV, a human *CAPN5* disease allele was expressed in mouse retinas using a lentiviral vector. Mouse *Capn5* DNA sequence, RNA expression, and protein expression was compared to *CAPN5* in the human retina. Lentiviral vectors were created to express either the wild-type human (*h*) *CAPN5* or the ADNIV mutant *hCAPN5-R243L* allele under a rhodopsin promoter with tandem green fluorescent protein (GFP) expression. Vectors were injected into the subretinal space of perinatal mice, and expression was determined by autofluorescent imaging. Mouse phenotypes were analyzed using electroretinography, histology, and inflammatory gene expression profiling. Mouse calpain-5 shows high homology to its human ortholog with over 98 percent sequence identity that includes the ADNIV mutant residue. In correlation with human expression, *Capn5* RNA was detected in mouse retina. Calpain-5 protein was expressed in the inner and outer segments of both rod and cone photoreceptors and in the inner plexiform layer. Live imaging of the lentiviral GFP reporter showed good uptake and distribution throughout the retina. Electroretinography, retinal histology, and expression of inflammatory genes indicate the *CAPN5-R243L* allele elicits an ADNIV-like disease in mice. Taken together, our studies suggest that ADNIV is due to a *CAPN5* gain-of-function, and retinal expression may be sufficient to generate an autoimmune response. Moreover, genetic models of ADNIV can be developed in the mouse and serve as pre-clinical models for therapeutic testing. Further elucidation of the pathogenic mechanism of *CAPN5* in ADNIV, by examining the effects of *hCAPN5* compared with *hCAPN5* including the single amino acid change R243L, will provide important new insight into causes of irreversible human blindness.

2896F

A primary cell culture model to study an inherited macular degeneration, DHDR/ML. *R. Fernandez-Godino, D. Garland, E. Pierce.* Ophthalmology, Ocular Genomics Institute- MEEI-Harvard Medical School, Boston, MA.

The p.Arg345Trp mutation in the *EFEMP1* gene causes the inherited macular degeneration Doyne Honeycomb Retinal Dystrophy/ Malattia Leventinese (DHRD/ML). The hallmark of DHRD/ML is the formation of drusen at an early stage. Gene targeted *Efemp1*$^{R345W/R345W}$ knockin mice develop extensive basal deposits which are considered precursors to drusen. In previous studies using *Efemp1*$^{R345W/R345W}$:complement *C3*$^{-/-}$ double mutant mice we demonstrated a critical role for complement system in the formation of the basal deposits. The purpose of this study was to use a primary RPE cell culture system to investigate the mechanisms involved in the pathogenesis of basal deposit formation in the *Efemp1*$^{R345W/R345W}$ mutant mice and the role of the complement system in this process. Primary RPE cells from wild type, *Efemp1*$^{R345W/R345W}$ knockin, and *Efemp1*$^{R345W/R345W}$:*C3*$^{-/-}$ double mutant mice were grown on a permeable support under polarizing conditions. Cell phenotype was characterized by electron and fluorescent microscopy. Expression levels of complement genes and *Efemp1* were quantified by qRT-PCR. Protein expression in the RPE cultures was determined by western blot analysis and immunofluorescence using confocal microscopy. As demonstrated by results of qRT-PCR experiments RPE cells from each strain express the following complement components and factors: *C1q, C2, C3, C4, Cfb, Cfd* and *Cfh*. Moreover, the expression of *C3* was significantly higher in *Efemp1*$^{R345W/R345W}$ knockin cells (ANOVA, p=0.0218) than in wild type. Electron micrographs showed that RPE cells from *Efemp1*$^{R345W/R345W}$ knockin but not from wild type or double mutant *Efemp1*$^{R345W/R345W}$:*C3*$^{-/-}$ mice secrete deposit-like material *in vitro*. Immunostaining of the deposits revealed that EFEMP1, C3 and other complement components colocalize in the deposits. The primary RPE cell culture system recapitulates *in vivo* observations. The primary RPE cell culture model is capable of reproducing deposit formation *in vitro*, which provides an excellent approach to investigate the very early effects of the *Efemp1* mutation on RPE dysfunction. The results suggest that complement expressed locally by the RPE is important in basal deposit formation in *Efemp1*$^{R345W/R345W}$ mutant mice. This model may be a useful system to test drugs, such as complement inhibitors, to prevent basal deposit formation in inherited macular degeneration as well as the more common age-related macular degeneration AMD.

2897W

Inactivation of the *miR-183/96/182* cluster gene results in syndromic retinal degeneration. *S. Xu[1,2,3], S. Lumayag[1,2,3], C.E. Haldin[1,2,3], N.J. Corbett[3], K.J. Wahlin[4], C. Cowan[1,2,3], S. Turturro[1,2,3], P.E. Larsen[5], P.D. Witmer[6], D. Valle[6], D.J. Zack[4], D.A. Nicholson[3].* 1) Pharmacology; 2) Ophthalmology; 3) Neurological Sci, Rush Univ Med Ctr, Chicago, IL; 4) Wilmer Eye Institute, The Johns Hopkins University School of Medicine, Baltimore, MD; 5) Biosciences Division, Argonne National Laboratory, Lemont, IL; 6) McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD.

The *miR-183/96/182* cluster is highly expressed in the retina and other sensory organs. To uncover its in vivo functions in the retina, we generated a knockout mouse model, designated as *miR-183C*$^{GT/GT}$, using a gene-trap embryonic stem cell clone. Our results showed that inactivation of the cluster resulted in early-onset and progressive synaptic defects of the photoreceptors, leading to abnormalities of scotopic and photopic ERGs with decreased *b*-wave amplitude as the primary defect and progressive retinal degeneration. In addition, inactivation of the *miR-183/96/182* cluster resulted in global gene-expression changes in the retina, with enrichment of genes important for synaptogenesis, synaptic transmission, photoreceptor morphogenesis and phototransduction, suggesting that the *miR-183/96/182* cluster plays important roles in postnatal functional differentiation and synaptic connectivity of photoreceptors. Our data suggest that the *miR-183/96/182* cluster is a new candidate gene for inherited retinal degeneration and susceptibility gene for age-related retinal dysfunction/degeneration. Although the phenotypes and molecular changes in *miR-183C*$^{GT/GT}$ mice are not a perfect match for any known human disease, mutation-negative Usher syndrome-like patients would be good candidates for *miR-183/96/182* loss-of-function mutations. Studies on genetic variation or polymorphisms around the *miR-183/96/182* gene are warranted in patients with age-related, progressive retinal and/or multiple sensory defects, and other neurological conditions.

2898T

The Dhtkd1 Tyr486* knock-in mouse model recapitulates some phenotypes of Charcot-Marie-Tooth disease type 2Q. *M. Gu[1], W. Guo[1], C. Luan[1], Z. Yu[1], Y. Chen[1], S. Dang[1], Y. Kuang[2], Z. Wang[1,2,3].* 1) Department of Medical Genetics, Shanghai Jiao Tong University School of Medicine (SJTUSM), Shanghai 200025, China; 2) Research Center for Experimental Medicine, Rui-Jin Hospital at SJTUSM, Shanghai 200025, China; 3) Shanghai Research Centre for Model Organisms, Shanghai 201203, China.

Charcot-Marie-Tooth disease (CMT) or peroneal muscular atrophy is one of the most common inherited neurological disorders with a prevalence estimated at 1/2500. On the basis of clinical manifestation and electrophysiological properties, CMT has been divided into four main types: CMT1, CMT2, CMT4 and CMTX. CMT2 is an axonal peripheral neuropathy characterized by distal muscle weakness and atrophy, normal or near-normal nerve conduction velocities. Our previous study demonstrated that the nonsense mutation [c.1455T→G (p.Tyr485*)] in exon 8 of DHTKD1 (dehydrogenase E1 and transketolase domain containing 1) was one of disease-causing genes in CMT2 subtype which had been named as CMT2Q (MIM 615025), implicating a significant role for DHTKD1 in the neurological development and mitochondrial energy production (AJHG, 2012, 91: 1088-1094). To identify the molecular mechanisms of mutant DHTKD1 causing CMT2Q, we generated a knock-in mouse model for the Tyr486* mutation in the Dhtkd1 gene. At first, we evaluated the gene expression, energy production and apoptosis of the mutant Dhtkd1 gene in the positive ES cell clones. We found that the gene expression and ATP production were significantly decreased in the heterozygous ES cells. Moreover, mutant cells became easier to apoptosis than wild-type cells under hydrogen peroxide (H_2O_2)-induced oxidative stress. Then, the phenotypes were observed among wild-type mice, heterozygous mice and homozygous mice from two heterozygous parental mice. The results showed the homozygous knock-in mice did not appear embryonic lethality or developmental abnormalities. The sciatic nerve and peroneal muscular of mutant mice at eight month of age also did not appear significant pathological changes. However, we found the ATP production in homozygous mice liver was obviously lower than that of the heterozygous and wild-type mice. The ratio of $NADP^+$/NADPH in homozygous and heterozygous mice was obviously higher than that of wild-type mice. These data demonstrated that nonsense mutation of Dhtkd1 gene led to decline in metabolic capacity in mice. We also found the response for hot in homozygous mice significantly slower than that of in heterozygous and wild-type mice. Taken together, we indicated that the Dhtkd1 Tyr486* knock-in mouse model recapitulated some phenotypes of CMT2Q. Further in-depth studies are underway.

2899F

Comprehensive characterization of a zebrafish model for pseudoxanthoma elasticum reveals a role for the abcc6 transporter in cardiovascular development. *M.J. Hosen[1,2], O. Vanakker[1], A. Willaert[1], P. Coucke[1], A. De Paepe[1].* 1) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 2) Dept. of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, Bangladesh.

Pseudoxanthoma elasticum (PXE) is characterized by elastic fiber calcification and fragmentation, resulting from ABCC6 mutations. Recently, up-regulation of the pro-osteogenic BMP2-RUNX2 pathway was demonstrated in patients. Also, recent reports on vascular malformations in patients suggested a possible developmental role for ABCC6. We used 4 zebrafish (ZF) models, targeting 3 abcc6 isoforms and combined knock-down, to assess the developmental phenotype, mineralization and mineralization-associated pathways. In-situ hybridization demonstrated all isoforms to be mainly expressed in the pronephric ducts. Abcc6 knock-down was obtained by injecting translational blocking or splice junction morpholinos (MOs) of abcc6 in 1-4 cell stage embryos. To avoid non-target-related phenotypes due to apoptosis, co-injection of an anti-p53 MO along with the experimental MO was performed. Morphants exhibited a delay in gastrulation. At 3 days post fertilization (dfp), curving and shortening of the tail, variable in severity, pericardial edema, decreased mobility, total body length and body pigmentation as well as underdevelopment of head and eyes were observed. Cardiovascular evaluation, using MO injection on Fli:gfp embryos demonstrated the development of abnormal rudimentary tubular heart with decreased heart beat and blood flow compared to controls, contributing to early demise of the morphants. Further, underdevelopment and delayed sprouting of multiple vessels was noted. Calcein staining of the morphants (5dpf) showed advanced skeletal mineralization compared to controls. QPCR analysis revealed up-regulation of bmp2a, runx2a and msxc. We illustrate a distinct phenotype, affecting longitudinal growth, eye and cardiovascular development. The cardiovascular underdevelopment in the morphants corroborates the hypothesis that ABCC6 may also have an embryological role. Advanced mineralization and confirmation of BMP2-RUNX2 involvement suggest that ZF is a useful model organism for further PXE cell signaling research.

2900W

Impaired viability of muscle precursor cells in muscular dystrophy with glycosylation defects and amelioration of its severe phenotype by limited gene expression. *K. Kobayashi[1], M. Kanagawa[1], C.C. Yu[1], C. Ito[1], S.I. Fukada[2], T. Chiyo[3], T. Okada[3], S. Takeda[3], T. Toda[1].* 1) Neurology/Molecular Brain Science, Kobe Univ Grad Sch Medicine, Kobe, Japan; 2) Molecular and Cellular Physiology, Osaka Univ Grad Sch Pharmaceutical Sciences, Osaka, Japan; 3) Molecular Therapy, National Institute of Neuroscience, NCNP, Tokyo, Japan.

A group of muscular dystrophies, dystroglycanopathy, is caused by abnormalities in post-translational modifications of dystroglycan (DG). To better understand the pathophysiological roles of DG modification and to establish effective treatment for dystroglycanopathy, we generated 2 distinct conditional knock-out (cKO) mice for *fukutin*, the first dystroglycanopathy gene identified for Fukuyama congenital muscular dystrophy. The first dystroglycanopathy model~myofiber-selective fukutin-cKO (MCK-fukutin-cKO) mice~showed mild muscular dystrophy. Forced exercise experiments in presymptomatic MCK-fukutin-cKO mice revealed that myofiber membrane fragility triggered disease manifestation. The second dystroglycanopathy model~muscle precursor cell (MPC)-selective cKO (Myf5-fukutin-cKO) mice~exhibited more severe phenotypes of muscular dystrophy. Using an isolated MPC culture system, we demonstrated that defects in the fukutin-dependent modification of DG lead to impairment of MPC proliferation, differentiation, and muscle regeneration. These results suggest that impaired MPC viability contributes to the pathology of dystroglycanopathy. Since our data suggested that frequent cycles of myofiber degeneration/regeneration accelerate substantial and/or functional loss of MPC, we expected that protection from disease-triggering myofiber degeneration provides therapeutic effects even in mouse models with MPC defects; therefore, we restored fukutin expression in myofibers. Adeno-associated virus (AAV)-mediated rescue of fukutin expression that was limited in myofibers successfully ameliorated the severe pathology even after disease progression. In addition, compared to other gene therapy studies, considerably low AAV titers were associated with therapeutic effects. Our findings indicated that fukutin-deficient dystroglycanopathy is a regeneration-defective disorder, and gene therapy is a feasible treatment for the wide range of dystroglycanopathy even after disease progression.

2901T

Mutations in Sonic hedgehog signaling pathway predispose to fatty liver. *A.F. Martinez[1], R.J. Lipinski[2], M. Guillen-Sacoto[1], S.K. Hong[1], J.L. Everson[3], K.K. Sulik[3], N.S. Trivedi[4], A.G. Elkahloun[5], B.D. Solomon[1], M. Muenke[1].* 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Dept. of Comparative Biosciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI; 3) The Bowles Center for Alcohol Studies, University of North Carolina, Chapel Hill, NC; 4) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Holoprosencephaly (HPE) is a severe genetic disorder that affects forebrain and craniofacial development. Several developmental factors implicated in HPE belong to or interact with Sonic hedgehog (Shh) signaling as the central pathway affected in this disorder. In our clinical studies we observed a significantly higher prevalence of fatty liver in HPE patients and unaffected carrier relatives (pediatric: p=0.0023, adult: p=0.0135) compared to the general population. While typically quiescent in the adult healthy liver, the Shh pathway becomes active in alcoholic and non-alcoholic fatty liver disease. The gene-environment studies described here explored the effect of Shh pathway mutations on liver metabolism using a mouse model of fatty liver disease. Along with their wild-type (WT) littermates, heterozygous knockout mice for *Gli2* (a proximal Shh signaling mediator) were fed a high-fat (HF) or a control diet for 12 weeks and their livers removed for gene expression profiling. All animals fed the HF diet showed an increase in body mass and liver steatosis, but $Gli2^{+/-}$ mice gained more weight and showed higher steatosis scores compared to the WT mice. $Gli2^{+/-}$ mice fed the control diet showed a differential expression of genes involved in carbohydrate and lipid metabolism, immune function, oxidative stress and cell cycle regulation, which suggests these mutants have an inherent metabolic defect impacting liver function. Top molecules include *Ctse, Taf1d, Gadd45g, Snora74a, Lcn2, Hsph1, Sult2a1, Sult2a2, Cyp3a5, Cyp2c9, Rnf186, Pdk4 and Rsad2*. Shh pathway genes did not show significant differences in gene expression. Though both $Gli2^{+/-}$ and WT mice fed the HF diet showed changes in genes involved in the same cellular processes, the $Gli2^{+/-}$ mice revealed a higher number of genes and genes associated with more advanced stages of liver disease. These results have key implications for the management of HPE patients and their families. A simple abdominal ultrasound may help identify possible mutation carriers in families with history of HPE. Future studies will involve complete metabolic/endocrinological profiling and histological evaluation of the hypothalamic/pituitary axis of the $Gli2^{+/-}$ mice, and the study of additional animal models (currently, $Shh^{+/-}$ mice). Our ultimate goal is to identify genetic risk factors for idiopathic fatty liver in the general population, a longstanding clinical problem posing a heavy burden on the healthcare system worldwide.

2902F

Imaging disease states using mouse models. *S. Rockwood, M. Osborne, M. Lessard, M. Sasner.* Genetic Resource Science, The Jackson Laboratory, Bar Harbor, ME.

The mouse is a widely used model organism well-suited for the study of mammalian development and physiology, and for elucidating details of disease mechanisms. Recent advances in genetic engineering have greatly expanded the options available to researchers to precisely label specific cellular subsets. The Jackson Laboratory (JAX) Mouse Repository, as part of its mission, facilitates the access of these valuable mouse strains to the research community. The approaches used in designing these genetically engineered mice incorporate both constitutive and inducible strategies, affording researchers a great deal of flexibility in controlling spatial and temporal expression. Numerous models employ a promoter with a well-characterized expression pattern, driving transcription of a reporter molecule. Among this growing set are GFP-expressing reporter lines developed by the Allen Institute for Brain Science (AIBS), various lacZ-expressing lines, and the latest generation of the well-known multi-colored 'Brainbow' models. The 'Confetti' strains are similar in concept to 'Brainbow' lines with the difference being that they can be expressed in any tissue and cell type. A large number of strains that express cre in the brain are available, including sets generated by the NIH Neuroscience Blueprint Cre Driver Network and the AIBS. The cre activity patterns for many of these strains have been characterized in a comprehensive manner at AIBS and JAX. The cre portal database (www.creportal.org) facilitates identification of an appropriate cre-expressing line by providing tools to search by promoter gene or by the anatomical site of expression. In combination with the wide array of cre reporter lines available, these resources enable the user to label the tissue/cell subsets of choice with the preferred fluorescent or other reporter molecule by simply mating the appropriate cre-expressing and cre-reporter lines. These alleles can be used in conjunction with existing mouse models of human disease to address translational questions by providing a way to better image tissues involved in disease pathologies.

Donating a strain to the Jax Mouse Repository is an easy way to fulfill the NIH's requirements for sharing mice. Researchers wishing to have strains considered for inclusion in the Repository are encouraged to submit their strains at: **www.jax.org/donate-a-mouse**.

2903W

The cDNA expression pattern in the brain of Long Evans Cinnamon rat, the animal model of Wilson disease. *G. Kim[1], B. Lee[1,2], J. Kim[1], J. Kim[1], S. Heo[1], M. Kang[1], J. Choi[2], H. Yoo[1].* 1) Med Gen Clinic & Lab, Asan Med Ctr, Seoul, South Korea; 2) Dept. Pediatrics, Asan Med Ctr, Seoul, South Korea.

Wilson's disease (WD; OMIM #277900) is one of the most common inborn metabolic disorders, which is caused by a deficiency in the P-type Cu-transporting ATPase (ATP7B). The pathogenic process of WD begins in the liver because ATP7B is primarily expressed in the liver, where it plays a central role in the regulation of intracellular Cu metabolism. As the pathogenic process progresses, the injuries widely extend to extrahepatic tissues, particularly the brain, and the neurological manifestations usually follow those of the liver in an age-dependent manner. The brain damage in WD is expected to be related to diverse molecular processes, including Cu-related toxicity, but the majority of the underlying mechanisms remain elusive. The study investigated the gene expression profiles using RatRef-12 bead array with the brains of WD animal model Long-Evans Cinnamon (LEC) rat at 24 weeks of age in order to identify the important early molecular changes that underlie the development of neurological symptoms in WD. Biological ontology-based analysis revealed diverse altered expressions of the genes related to copper accumulation. Of particular interest, we found altered expression of genes connected to mitochondrial respiration (Sdhaf2 and Ndufb7), calcineurin-mediated cellular processes (Ppp3ca, Ppp3cb, and Camk2a), amyloid precursor protein (Anks1b and A2m) and alpha-synuclein (Snca). In addition to copper-related changes, compensatory upregulations of Cp and Hamp reflect iron-mediated neurotoxicity. Of note, reciprocal expression of Asmt and Bhmt is an important clue that altered S-adenosylhomocysteine metabolism might underlie brain injury in WD, which is directly correlated to the decreased expression of S-adenosylhomocysteine hydrolase in hepatic tissue in LEC rats. This study indicates that diverse molecular changes, both variable and complex, underlie the development of neurological manifestations in WD. Copper-related injuries were found to be the principal pathogenic process, but iron (Fe)- or adenosylhomocysteine-related injuries were also implicated.

2904T

Dube3a expression levels affect axonal propagation and resting potential in fly models of Angelman syndrome and Duplication 15q autism. *L. Reiter[1,2], C. Valdez[1], R. Scroggs[2].* 1) Dept Neurology, UTHSC, Memphis, TN; 2) Dept of Anatomy and Neurobiology, UTHSC, Memphis, TN.

Changes in UBE3A expression levels in neurons can cause neurogenetic disorders ranging from Angelman syndrome (AS) (decreased levels) to 15q duplication autism (increased levels). Here we investigated the effects on neuronal function of varying UBE3A levels using the *Drosophila* neuromuscular junction as a model for both of these neurogenetic disorders. Stimulations that evoked excitatory junction potentials (EJPs) at 1 Hz intermittently failed to evoke EJPs at 15 Hz in a significantly higher proportion of Dube3a over-expressors (*C155*-GAL4>UAS-*Dube3a*) relative to controls (*C155*-GAL4 alone). However, in the Dube3a over-expressing larval neurons with no failures at 15 Hz, there was no difference in EJP amplitude at the beginning of the train, or the rate of decrease in EJP amplitude over the course of the train compared to controls. In the absence of tetrodotoxin (TTX), spontaneous EJPs were observed in significantly more *C155*-GAL4>UAS-*Dube3a* larva compared to controls. In the presence of TTX, spontaneous and evoked EJPs were completely blocked and mEJP amplitude and frequency did not differ among genotypes. These data suggest that over-expression of wild type Dube3a, but not a ubiquitination defective *Dube3a*-C/A construct, compromises the ability of motor neuron axons to support closely spaced trains of action potentials, while at the same time increasing excitability. EJPs evoked at 15 Hz in the absence of Dube3a ($Dube3a^{15b}$ homozygous mutant larvae) decayed more rapidly over the course of 30 stimulations compared to w^{1118} controls, and $Dube3a^{15b}$ larval muscles had significantly more negative resting membrane potentials (RMP). These data suggest that reduced *UBE3A* expression levels may influence factors involved in maintenance of muscle and nerve RMP and neurotransmitter release from motorneurons. Similar affects of under- and over-expression of UBE3A on membrane potential and synaptic transmission may underlie the synaptic plasticity defects observed in both AS and autism.

2905F

Dysregulation of inflammatory pathways in a Familial Hemiplegic Migraine 1 mouse model after the induction of cortical spreading depression. *E. Eising[1], B. de Vries[1], R. Shyti[1], L.S. Vijfhuizen[1], L.A.M. Broos[1], N.A. Datson[1], E.A. Tolner[1], P.A.C. 't Hoen[1], M.D. Ferrari[2], A.M.J.M. van den Maagdenberg[1,2].* 1) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 2) Department of Neurology, Leiden University Medical Centre, Leiden, The Netherlands.

Familial Hemiplegic Migraine type 1 (FHM1) is a rare monogenic subtype of migraine with aura caused by mutations in the CACNA1A gene. CACNA1A encodes a subunit of voltage-gated CaV2.1 calcium channels that regulate neurotransmitter release in brain synapses. In FHM1 knock-in mouse models these mutations enhance cortical glutamatergic signaling and increase the susceptibility for cortical spreading depression (CSD): a transient wave of neuronal and glial depolarization followed by a prolonged depression of activity that underlies the migraine aura. We used CSD induction as a migraine related trigger to measure gene expression profiles in our FHM1 mouse model under triggered conditions. We induced seven CSDs in the caudal cortex by KCl application onto the dura, which were recorded in the frontal cortex. In sham surgeries NaCl was used instead as it does not induces CSDs. Some 24 hours after surgery, the middle part of the cortex, through which the CSD waves travelled, was isolated and subjected to deep-Serial Analysis of Gene Expression (SAGE) sequencing (a tag-based Next Generation sequencing method for gene expression profiling). Sequencing data was aligned to the mouse genome with Bowtie and analyzed for differential expression using the Bioconductor package EdgeR. Gene expression differences were validated using qPCR. We identified genes differentially expressed; 1) between FHM1 and wild-type mice (genotype effect), 2) between CSD and sham groups (CSD effect), and 3) due to an interaction between genotype and CSD effects (interaction effect). Only few genes showed a genotype effect. The CSD effect comprised more genes, and a gene ontology term analysis identified an enrichment of immune related functions. The largest number of differentially expressed genes showed an interaction effect. Using k-means cluster analysis, we identified a set of genes that showed a CSD effect only in the FHM1 mice. Interestingly, this set is predominantly involved in inflammation and includes a network of interferon-related genes. Whereas the cortical expression profile of our FHM1 mice was previously shown to be relatively stable under basal (untriggered) conditions, numerous gene expression differences were identified between FHM1 and wild-type mice after CSD induction. Our results point towards a stronger immune reaction to CSD in FHM1 mice than in wild-type mice, which highlights a role for disturbed immune function in the pathophysiology of migraine.

2906W

Analysis of the role of GAA repeat expansion instability in Friedreich ataxia pathology in a humanized mouse model. *J. Sarsero[1, 2].* 1) Cell and Gene Therapy, Murdoch Childrens Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia; 2) Department of Paediatrics, The University of Melbourne, Royal Children's Hospital, Parkville, Victoria, Australia.

There is evidence that age-dependent and tissue-specific somatic instability of the GAA repeat expansion may be a determinant of the progressive pathology of Friedreich ataxia (FRDA), and is evident in FRDA patients, FRDA iPS cells and a YAC-based GAA repeat expansion mouse model. Interruptions in GAA repeat sequences can alleviate transcription inhibition and reduce genetic instabilities. We explored the role of instability of the GAA repeat expansion on FRDA pathology using a humanized mouse model containing an interrupted GAA repeat expansion. An interrupted GAA repeat expansion was introduced into the first intron of the human *FXN* gene present on a BAC clone by recombineering. The genomic insert was used to generate humanized transgenic/KO mice. The presence of the introduced interrupted GAA repeat expansion resulted in markedly decreased levels of human *FXN* transcript and frataxin protein in humanized mouse tissues. The region immediately upstream of the interrupted GAA repeat expansion region was found to be almost completely methylated. Somatic instability of the interrupted GAA repeat expansion was not detected using the small pool PCR technique. The assessment of phenotypic symptoms of FRDA by a series of behavioral, neurological, biochemical and histological tests did not reveal any significant phenotypic differences between humanized and wild type mice. The interruption of the GAA sequence contributes to the somatic stability of the repetitive element, which in turn results in the mice lacking an obvious phenotype despite the low levels of *FXN* mRNA and frataxin protein and repressive epigenetic changes.

2907T

Beta-glucosidase 2 (GBA2), which is mutated in inherited spastic paraplegia and cerebellar ataxia, is sensitive to inhibition by conduritol B epoxide. *A.C. van der Spoel[1], C.M. Ridley[1], K.E. Thur[1], N.B. Thillaiappan[2], A.A. Rahim[3], S.N. Waddington[3].* 1) Departments of Pediatrics and Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada; 2) Department of Pharmacology, University of Oxford, Oxford OX1 3QT, UK; 3) Gene Transfer Technology Group, Institute of Women's Health, University College London, London WC1E 6HX, UK.

Genetic diseases caused by defects in the lysosomal degradation of glycosphingolipids are well characterized and their pathophysiologies are increasingly understood. Examples of these lysosomal storage disorders are GM1-gangliosidosis, Sandhoff, Tay-Sachs, Fabry, and Gaucher disease. The latter condition results from deficiencies in glucocerebrosidase, which is an endolysosomal glycohydrolase that cleaves glucosylceramide into glucose and ceramide. Recently, two inherited conditions were found to be caused by defects in the extralysosomal degradation of glucosylceramide. Individuals affected with hereditary spastic paraplegia 46 (SPG46) and autosomal-recessive cerebellar ataxia carried mutations in the β-glucosidase 2 gene (*GBA2*). Similar to glucocerebrosidase, GBA2 cleaves glucosylceramide, but GBA2 is located in the ER and/or at the plasma membrane. GBA2 activity was reported to be drastically reduced in a lymphoblastoid cell live derived from mononuclear white blood cells from an SPG46 patient. Assessment of enzymatic activity of GBA2 is thus clinically relevant. This raises the challenge to distinguish the GBA2 activity from that of glucocerebrosidase, as both β-glucosidases can degrade the same natural and artificial substrates under similar conditions. The first approach proposed to specifically measure GBA2 activity was to use conduritol B epoxide (CBE), which is an irreversible inhibitor of glucocerebrosidase. CBE was thought not to affect GBA2. Nevertheless, we found that CBE irreversibly inactivated GBA2. The extent of GBA2 inhibition depended on the CBE concentration and time of exposure. We propose an alternative strategy to measuring GBA2 activity using *N*-butyldeoxygalactonojirimycin (*N*B-DGJ), which inhibits GBA2, but leaves glucocerebrosidase fully active. Therefore, we redefine GBA2 activity as the *N*B-DGJ-sensitive β-glucosidase. In mouse tissues and cultured cells, the *N*B-DGJ-sensitive β-glucosidase activity was up to 10-fold higher than the CBE-resistant β-glucosidase activity. In fact, instead of representing less than 10% of the total β-glucosidase activity, GBA2 was responsible for the majority of the detergent-independent β-glucosidase activity of most tissues and cell lines. Our results enable a more accurate assessment of GBA2 activity, which will be an asset in evaluating the roles of GBA2 in hereditary disorders, and may be helpful in clinical practice.

2908F

A biological role for impaired BDNF transcription in Familial Dysautonomia. *M. Nilbratt[1,2], G. Lee[3], E. Morini[1,2], S.J. Haggarty[1,2], L. Studer[4], S.A. Slaugenhaupt[1,2].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Institute for Cell Engineering, John Hopkins University, Baltimore, MD; 4) Center for Stem Cell Biology, Memorial Sloan-Kettering Cancer Center, New York, NY.

Familial Dysautonomia (FD), or Riley-Day syndrome, is an autosomal recessive disorder with extensive sensory and autonomic nervous system involvement present at birth. FD is characterized by recurrent episodes of hypertension with tachycardia, nausea and vomiting, impaired thermoregulation, and deceased sensitivity to pain and temperature. The clinical features of FD are due to a genetic mutation in the IKBKAP gene that results in a tissue-specific reduction of IKAP protein. Neuropathological studies show a marked reduction of neurons in the sympathetic and sensory ganglia. All FD patients have an intronic splice site mutation in the IKBKAP gene, the scaffolding member of the Elongator protein complex involved in transcriptional elongation. This mutation results in tissue-specific skipping of exon 20 in the mRNA with aberrant splicing most pronounced in neuronal tissues. The alternative splicing defect leads to reduced production of normal IKAP protein in FD patients. In addition, complete embryonic loss of Ikbkap in mice causes early embryonic lethality, suggesting that this gene is required for effective transcription of genes involved in very early neural development. Neurotrophic factors are implicated in survival and differentiation of several neuronal populations and in the regulation of neurogenesis, synaptic plasticity, learning and memory. Although there is evidence for a role of neurotrophins in various human neuropathies, their relevance to the disease process is not fully understood. Previous studies have indicated reduced neurotrophic activity mediated by nerve growth factor (NGF) in fibroblasts from FD patients. Here we show a functional role of reduced brain-derived neurotrophic factor (BDNF) transcription associated with neurotrophic activity in our induced pluripotent stem (iPS) cell model system of human neural development. We found impaired neurotrophic support from patient fibroblast cells and by inhibition of BDNF activity from normal fibroblasts on the development of iPS cell-derived neurons. Interestingly, the reduced biological activity from patient cells was rescued by increasing the IKAP expression using small chemical compounds. Our data indicate a mechanism underlying FD that may also provide insights for potential new therapeutic interventions.

2909W

Species-specific expression of acidic mammalian chitinase in stomach tissues. *M. Ohno[1], Y. Togashi[1], K. Tsuda[1], K. Okawa[1], M. Kamaya[2], M. Sakaguchi[1], Y. Sugahara[1], F. Oyama[1].* 1) Biotechnology Laboratory, Dept Applied Chemistry, Kogakuin Univ; 2) Enviroment Analytical Chemistry Laboratory, Department of Environmental and Energy Chemistry,Kogakuin Univ.

Two active chitinolytic enzymes, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase), have been identified in mouse and human, although mammals do not have chitin synthase. Since these chitinolytic enzymes are significantly increased in Gaucher disease and mouse models of asthma, both chitinases have been regarded primarily as a host-defense mechanism against chitin-containing pathogens and/or play important roles in the pathophysiology. Little is known, however, about the mutual regulation of both chitinases between human and mouse. We quantified expression levels of the chitinases to compare the mRNA levels between human and mouse tissues on the same scale using a human-mouse hybrid standard DNA. Our analysis showed that Chit1 mRNA is expressed at similar levels in normal human and mouse lung. On the other hand, AMCase is predominantly overexpressed in mouse but not human stomach. To see whether these mRNA changes are reflecting actual differences in the levels of protein expression, we measured chitinolytic activity at pH 2.0 and pH 5.0 using the synthetic substrate of 4-methylumbelliferyl β-D-N, N'-diacetylchitobiose (4MU-chitobiose) and detected robust chitinolytic activity in mouse stomach extract at pH 2.0 and strong activity at pH 5.0. In contrast, no activity was detected in that of human at pH 2.0 and very low activity was observed at pH 5.0. Furthermore, the anti-AMCase antibodies recognized a robust single protein band in extract from mouse stomach but not from human. Thus, mRNA differences between human and mouse stomach tissues were reflecting differences in the chitinolytic activities and levels of protein expression. Thus, the expression level of the AMCase in the stomach is species-specific.

2910T

Quantitative expression analysis of mammalian chitinases in human and mouse tissues. *F. Oyama, M. Ohno, Y. Togashi, K. Tsuda, K. Okawa, M. Sakaguchi, Y. Sugahara.* Dept Applied Chemistry, Kogakuin Univ, Hachioji, Tokyo, Japan.

Chitinase hydrolyzes chitin, which is an N-acetyl-D-glucosamine polymer that is present in a wide range of organisms, including insects, parasites and fungi. Although mammals do not have chitin synthase, genes encoding chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase) and their translation products have been found in both human and mouse. Recent studies have shown an association between the mammalian chitinases and inflammatory diseases. For instance, Chit1 levels were elevated in the plasma of Gaucher disease and the bronchoalveolar lavage fluid of smokers with chronic obstructive pulmonary disease (COPD). AMCase expression and activity is upregulated in allergic airway responses in mouse models of asthma. Little is known, however, about mutual regulation of gene expression between Chit1 and AMCase in vivo. We established a quantitative PCR system using a single standard DNA and found that both chitinases mRNAs are predominantly expressed in mouse stomach. In stomach, AMCase mRNA is expressed 7-10 fold higher than those of housekeeping genes and expressed comparative level with pepsinogen C (progastricsin) mRNA, a major component of the gastric mucosa. Furthermore, we applied this methodology to the quantification of chitinase mRNAs in human tissues. We found that both chitinase mRNAs were widely expressed in normal human tissues. Chit1 mRNA was highly expressed in the human lung, whereas AMCase mRNA was not overexpressed in normal human stomach tissues. The levels of these mRNAs in human tissues were significantly lower than the levels of housekeeping genes. These results indicate that the Chit1 and AMCase expression levels are relatively low in the human tissues examined and that the AMCase expression level in the stomach differs significantly between human and mouse.

2911F

Molecular basis of a new form of hyperekplexia. *J. Capo-chichi[1,2], S. Boissel[3], E. Brustein[3], F.F. Hamdan[1,4], M. Samuels[1], G.A. Rouleau[4], P. Drapeau[3], J.I. Michaud[1,2].* 1) Centre de Recherche du CHU Sainte-Justine, Montréal, Québec, Canada; 2) Département de biochimie , Université de Montréal,Québec,Canada; 3) Département de pathologie et biologie cellulaire,Université de Montréal,Québec,Canada; 4) Centre de recherche du CHUM, Montréal,Québec, Canada.

Hyperekplexia (HK) is characterized by increased tone with startle reactions inducible upon tactile or auditory stimulations. HK is a genetic condition caused by disruption of glycine signaling in the CNS. Here we report on a consanguineous family from Cambodian origin composed of 4 affected female siblings who, in addition to severe HK, showed small birth weight, microcephaly with simplified gyral structure, and burst suppression on the EEG. They could not breath by themselves and died a few days after birth. We performed homozygosity mapping on all 4 affected siblings followed by exome sequencing on 2 of them and identified, in the shared regions of homozygosity, 1 rare predicted-damaging homozygous mutation in CLPB (p.Ile562fs) that segregated with the disease. CLPB belongs to a subfamily of AAA+ proteins that includes clpb (bacteria) and hsp104/hsp78 (yeast), both of which function in protein disaggregation and resolubilization. Members of this subfamily form homo-hexameric ring. p.I562Ifs truncates the protein upstream of a well-conserved segment that has been shown to be required for hsp104/clpb protein oligomerization. It is thus likely that p.Ile562fs abolishes CLPB function. To investigate the functional consequence of p.I562Ifs, we interfered with the zebrafish ortholog Clpb mRNA translation using a splice-blocking antisense morpholino (AMO). The specificity of the morpholino was confirmed by RT-PCR and western blot. The Clpb KD zebrafish larvae were well formed but presented some morphological deficits (smaller eyes, shorter head and body length) compared to wild-type, non-injected larvae. They also exhibited a motor phenotype characterized by an abnormal touch-evoked response with increased maximum swim velocity and faster tail beat frequency. All larvae co-injected with wild-type human CLPB mRNA and Clpb AMO exhibited restoration of the morphological and motor phenotypes. Conversely, larvae co-injected with mutated human CLPB mRNA and Clpb AMO showed similar morphological and motor deficits to those observed in Clpb KD larvae. These phenotypes due to Clpb KD are reminiscent of the microcephaly, small birth weight and hyperexcitabilty observed in our patients. Taken together, our results suggest that the disruption of CLPB is responsible for the observed HK phenotype.

2912W

A homozygous missense mutation in HSPA9 causes epiphyseal-vertebral-ear (EVE) dysplasia. *J. Amiel[1,2], M. Oufadem[2], A. Linglart[3], D. Lehalle[1], C. Bole[2], P. Nithske[2], A. Munnich[1,2], S. Lyonnet[1,2], C. Gordon[2].* 1) Dept Genetics, Necker Hosp, AP-HP, Paris, France; 2) IMAGINE Institute, INSERM, Paris, France; 3) Pediatric Endocrinology, Bicêtre Hosp, AP-HP, Kremlin Bicêtre, France.

Epiphyseal-vertebral-ear (EVE) dysplasia associates distinctive dysmorphic features and ossification defects. A single family has been reported, with two affected sisters born to healthy, consanguineous parents. Key features were midface hypoplasia, depressed nasal bridge, anteverted nares, symmetrical dysplasia of the helices, epiphyseal dysplasia, mid-coronal clefting of vertebral bodies, odontoid hypoplasia, elbow dislocation and intrauterine growth restriction with post-natal catch-up growth. Similar craniofacial and skeletal features have been noted in cerebral, ocular, dental, auricular, skeletal anomalies (CODAS) syndrome. The genetic causes of EVE and CODAS syndromes have not been reported. We performed exome sequencing of the two affected EVE dysplasia patients, and after filtering for published and in-house SNPs and selecting for deleterious variants, the only deleterious homozygous variant shared between both sisters was the missense mutation p.Thr362Ile in heat shock 70kDa protein 9 (HSPA9). The catalytic activity of HSPA9 is encoded by the ATPase domain, a structure conserved in chaperones from vertebrates to bacteria. Thr362 is absolutely conserved in known orthologues of HSPA9, falls within the ATPase domain and structural modelling of the mutation Thr362Ile indicates disruption of hydrogen bonding, predicted to lead to protein instability. Published transcriptomic data indicates that HSPA9 is highly expressed in human fetal cartilage. This work identifies HSPA9 as a novel regulator of craniofacial and skeletal development, with its mutation causing EVE dysplasia. These findings add specific roles during development to the ones already ascribed to the HSP70 family of chaperones. Whether CODAS syndrome is caused by mutations in HSPA9, a related gene or pathway remains to be explored.

2913T

Autosomal Recessive Congenital Ichthyosis (ARCI) and related disorders: Mutation in *CERS3*, coding for ceramide synthase 3, reveal major puzzle peaces for the understanding of epidermal barrier formation. *K.M. Eckl[1,2], R. Tidhar[3], H. Thiele[4], V. Oji[5], I. Hausser[6], S. Brodesser[7,8], M.L. Preil[9], A. Önal-Akan[2], F. Stock[10], D. Müller[10], K. Becker[2], R. Casper[2], G. Nürnberg[4], J. Altmüller[4], P. Nürnberg[4,8], H. Traupe[5], A.H. Futerman[3], H.C. Hennies[1,2,8].* 1) Center for Dermatogenetics, Innsbruck Medical University, Innsbruck, Austria; 2) Center for Dermatogenetics, Cologne Center for Genomics, University of Cologne, Cologne, Germany; 3) Dept. of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel; 4) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 5) Dept. of Dermatology, University Hospital of Münster, Münster, Germany; 6) Dept. of Dermatology, University Hospital of Heidelberg, Heidelberg, Germany; 7) Inst. of Med. Microbiology, Immunology and Hygiene, University of Cologne, Cologne, Germany; 8) Cluster of Excellence on Cellular Stress Responses in Aging-associated Diseases, University of Cologne, Cologne, Germany; 9) Practice for Dermatology Dres. Krnjaic, Merk, Preil, Schäfer, Ansbach, Germany; 10) Inst. of Human Genetics, University Hospital of Leipzig, Leipzig, Germany.

The barrier function of the human epidermis is supposed to be governed by lipid composition and organization in the stratum corneum. We report on a large German family with multiple consanguinity. The pedigree consists of several branches each with affected and unaffected family members. Using a combined approach of SNP-based autozygosity mapping and exome sequencing in autosomal recessive congenital ichthyosis we identified a homozygous interval of 3.4 Mb on chromosome 15 containing homozygous missense mutations in two genes. Cosegregation analysis pinpointed mutations in *CERS3*, coding for ceramide synthase 3 (CerS3), as underlying the phenotype. Using LC-ESI-MS/MS we demonstrate that the mutation impairs the activity of CerS3, which synthesizes very long chain ceramides in the skin. We show a specific loss of free and bound ceramides with chain lengths from C26 up to C34 in terminally differentiating patient keratinocytes. Reconstructed skin from either healthy control cells or patient fibroblasts and keratinocytes reveals premature epidermal differentiation and a moderate impairment of barrier function in patient skin models. Our findings demonstrate that specific synthesis of very long chain ceramides by CerS3 is a crucial early step for the formation of the epidermal barrier and indicate that disorders characterized by ichthyosis can be attributed to defects in the epidermal metabolism of ceramides and other components of the cornified lipid envelope. With the identification of mutations in *CERS3* a missing link between several genes implicated in congenital ichthyosis has been identified, and we propose a crucial epidermal pathway of the lipid metabolism involving *CERS3* and the ARCI genes *TGM1*, *ALOX12B*, *ALOXE3*, *ABCA12*, *CYP4F22*, but also for *LIPN*, mutated in late onset autosomal ichthyosis; as well as for *FALDH*, *ELOVL4*, *FATP4*, *ABHD5*, which are mutated in syndromic ichthyosis disorders.

2914F

The Centriolar satellite protein AZI1 interacts with BBS4 and is involved in ciliary trafficking of the BBSome. *X. Chamling[1], S. Seo[2], C.C. Searby[4], G.H. Kim[1], D.C. Slusarski[3], V.C. Sheffield[1,2,4].* 1) Pediatrics, University of Iowa, Iowa City, IA; 2) Department of Ophthalmology and Visual Sciences, University of Iowa Carver College of Medicine, Iowa City, IA; 3) Department of Biology, University of Iowa, Iowa City, IA; 4) Department of Pediatrics, Howard Hughes Medical Institute, IA.

Bardet-Biedl syndrome (BBS) is a ciliopathy with mutations reported in 17 different genes. Most of the protein products of the BBS genes localize at or near the primary cilium and the centrosome. Near the centrosome, BBS proteins interact with centriolar satellite proteins, and the BBSome (a complex of seven BBS proteins) is believed to play a role in transport of ciliary membrane proteins. However, the precise mechanism by which BBSome ciliary trafficking activity is regulated is not fully understood. We show that a centriolar satellite protein, AZI1 (also known as Cep131), interacts with the BBSome and regulates BBSome ciliary trafficking activity. Furthermore, we show that AZI1 interacts with the BBSome through BBS4. AZI1 is not involved in BBSome assembly, but accumulation of the BBSome in cilia is enhanced upon AZI1 depletion. Under conditions in which the BBSome does not normally enter cilia, such as in BBS3 or BBS5 depleted cells, knock down of AZI1 with siRNA restores BBSome trafficking to cilia. Finally, we show that azi1 knockdown in zebrafish embryos results in typical BBS phenotypes including Kupffer's vesicle abnormalities and melanosome transport delay. These findings associate azi1 with the BBS pathway. Our findings provide further insight into regulation of BBSome ciliary trafficking and identify AZI1 as a novel ciliopathy candidate gene.

2915W

NPHP10 (SDCCAG8) interacts with components of the multi-aminoacyl-tRNA synthetase complex. *K. Weihbrecht[1,2], M. Humbert[1,2], V. Sheffield[1,3], S. Seo[2].* 1) Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA; 2) Dept. of Ophthalmology and Visual Sciences, Univ. of Iowa, Iowa City, IA; 3) Howard Hughes Medical Institute, Univ. of Iowa, Iowa City, IA.

Nephronophthisis (NPHP) is a recessive kidney disorder that is the leading cause of early onset, end-stage renal failure. Many proteins mutated in cystic kidney disease have been shown to localize to the primary cilia and centrosomes, providing a coalescing mechanism for NPHP-related ciliopathies (NPHP-RC). Aside from renal failure and kidney cysts, retinal degeneration and dysplasia or degeneration of the cerebellum are also seen in many NPHP-RCs. SDCCAG8 is a nephronophthisis gene (NPHP10), with patients exhibiting retinal and renal abnormalities, obesity, and learning disabilities. Mutations in SDCCAG8 were also found in several BBS patients, making SDCCAG8 the 16th BBS gene (BBS16). However, little is known about the molecular functions of NPHP10 and how loss of NPHP10 function leads to the observed phenotypes. Our goal was to gain insight into the function of NPHP10 by determining its interactors. Here, we show that NPHP10 interacts with components of the multi-aminoacyl tRNA synthetase complex (MSC), including 8 out of 9 aminoacyl tRNA synthetases (ARS) as well as aminoacyl-tRNA-synthetase-complex interacting multifunctional protein 2 (AIMP2). We performed tandem affinity purification using stably transfected HEK293T cells expressing FLAG- and S-tagged NPHP10 and isolated its associated proteins. We further determined that among the MSC components, NPHP10 directly interacts with AIMP2. Co-immunoprecipitation was used to determine interacting domains of these two proteins. Sucrose gradient ultracentrifugation of WT mouse tissues showed that NPHP10 is associated with the MSC in the kidney, brain, and eye, all affected tissues in NPHP10 patients. Lastly, we show that disruption of the MSC through AIMP2 knockdown significantly reduces ciliogenesis. Altogether, these findings suggest that NPHP10 functions through an MSC-mediated mechanism or vice versa, and that the MSC is an important component for ciliary function.

2916T

Fragile X Mental Retardation Protein (FMRP) in cell differentiation: The MEG-01 as a new study model. *M. Mc Coy, F. Corbin.* Biochemistry Dept, University of Sherbrooke, Sherbrooke, Quebec, Canada.

Introduction: Fragile X Syndrome (FXS) is the most common form of inherited mental retardation. It is triggered by the expansion of CGG repeats in the 5'UTR of the FMR1 gene, as well as from the methylation of its CpG Island. The latter causes transcriptional silencing of the encoded Fragile X Mental Retardation Protein (FMRP). FMRP is an RNA-binding protein commonly associated with polyribosomes and is thought to play a critical role in the maturation of the synaptic network by repressing translation. Consequently, in the absence of FMRP, deregulated protein synthesis results in the abnormal maturation of neuronal dendritic spines. To investigate the more universal role of FMRP in architecture development during cell differentiation, a new model was established using a megakaryoblastic cell line: MEG-01. **Methods:** MEG-01 cells were activated with PMA and left to differentiate until complete maturation, with the release of platelet-like particles (PLPs). Complementary techniques were employed to study the behavior of FMRP during those changes. In fractionation experiments, cells were lyzed and cellular constituents were separated either by differential ultracentrifugation or on sucrose gradient. Protein content was then revealed by Western Blot analysis. In immunofluorescence experiments, protein expression was analyzed by FACS, while confocal microscopy was performed for subcellular localization and colocalization assays. **Results:** Cell fractionation experiments revealed the redistribution of FMRP and other RNA-binding proteins throughout the cell upon activation. By microscopy, the reorganization of the cytoskeleton and the elongation of filopodial extensions, releasing PLPs was observed. FMRP and other RNA-binding proteins were shown to colocalize with microtubules in the cytoplasmic extensions. **Conclusion:** The regulation of local synthesis appears to be important in the maturation of the cytoplasmic extensions, considering that FMRP and other proteins of the translational machinery accompany mRNPs towards regions of the cell undergoing development. The reorganization of the cytoskeleton and colocalization of mRNPs with microtubules suggests that translational complexes are carried from the core of the cell to filopodial extensions for local synthesis by means of microtubular transport, as is the case in neuronal maturation. FMRP is therefore necessary as part of a complex for regulating the process of maturation of elaborate cellular networks.

2917F

Hypervitaminosis D due to 1,25-(OH)2D-24 hydroxylase (CYP24A1) deficiency causing nephrocalcinosis and nephrolithiasis. *G. Nesterova[1], M. Malicdan[1], T. Sakaki[2], M. Collins[3], D. Adams[4,5], C. Boerkoel[4], W. Gahl[4,5].* 1) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD, USA,; 2) Department of Biotechnology, Faculty of Engineering, Toyama Prefectural University, Toyama, Japan; 3) 3Skeletal Clinical Studies Unit, Craniofacial and Skeletal Diseases Branch, National Institute of Dental and Craniofacial Research, NIH, Bethesda, MD, USA; 4) NIH Undiagnosed Diseases Program, NIH Office of Rare Diseases Research and Human Genome Research Institute, Bethesda, MD, USA; 5) Office of the Clinical Director, National Human Genome Research Institute, NIH, Bethesda, MD, USA.

Background/Objectives: An excess of an active form of Vitamin D1, or 1α,25(OH)2D3, can result in nephrocalcinosis and nephrolithiasis. Nephrolithiasis represents a global health problem with prevalence at 10%; we evaluated the cause of increased 1α,25(OH)2D3 levels in the development of those disorders. Patients and Methods: We measured serum and urine calcium and phosphate, vitamin D metabolites and performed mutation analysis in CYP24A1 gene encoding 1,25$(OH)_2$D-24-hydroxylase in two patients with hypervitaminosis D, nephrocalcinosis or nephrolithiasis enrolled in the National Institutes of Health Undiagnosed Diseases Program. CYP24A1 pathological variants and mutations were evaluated using the dbSNP database and the pathogenicity prediction programs. Results: Both patients exhibited hypercalciuria, low level of 24,25(OH)2D3, elevated 1α,25(OH)2D3, and undetectable activity of 1,25$(OH)_2$D-24-hydroxylase that inactivates 1α,25(OH)2D3. They had biallelic mutations in the CYP24A1 leading to the loss of function of this enzyme. Based upon dbSNP data, the frequency of deleterious biallelic CYP24A1 mutations and variants is estimated to be 1960 per 100,000 in the general population. That would be a significant number contributing to the prevalence of nephrolithiasis . Conclusion/Discussion: We found that CYP24A1 loss of the function mutations is associated with nephrocalcinosis and nephrolithiasis. Our assessment of CYP24A1 gene variants predicted that pathogenic defects in CYP24A1 may account for a significant fraction of nephrolithiasis. Recognition of CYP24A1 deficiency could prompt the recommendation that 1α,25(OH)2D3 levels be determined in such patients and promotes therapy for hypervitaminosis D.

2918W

ARID1B inhibits WNT signaling through interaction with BRG1 and β-catenin. *G. Vasileiou[1], A.B. Ekici[1], S. Uebe[1], J. Behrens[2], A. Reis[1], M.V. Hadjihannas[2].* 1) Inst. of Human Genetics, Universität Erlangen-Nürnberg, Erlangen, Germany; 2) Nikolaus Fiebiger Center, Universität Erlangen-Nürnberg, Erlangen, Germany.

The SWI/SNF chromatin remodeling complex regulates transcription by remodeling chromatin and through interactions with transcriptional regulators. Recently, deleterious mutations in subunits of this complex, which carry enzymatic activity and confer functional specificity, have been found in a diverse spectrum of human cancers, with an overall frequency approaching that of mutations in the p53 oncogene. Likewise, haploinsufficiency due to mutations in various components of the SWI/SNF complex was shown to be a frequent cause of the Coffin-Siris syndrome and intellectual disability. The mechanisms by which mutations in SWI/SNF subunits contribute to cancer and intellectual disability remain poorly understood. Here we show that ARID1B suppresses Wnt/β-catenin signaling and that Wnt/β-catenin target genes are upregulated in intellectual disability patients carrying mutations in ARID1B as compared to normal subjects. We found that ARID1B associates with β-catenin in nuclear puncta through a mechanism involving the SWI/SNF subunit BRG1, a binding partner of both β-catenin and ARID1B. Mutation of ARID1B as seen in an intellectual disability patient leads to partial truncation of its BRG1 binding domain, which reduces the ability of ARID1B to localize β-catenin to nuclear puncta and interferes with the suppressive effect on β-catenin-driven transcription. Aberrant activation of Wnt/β-catenin signaling is a causative feature of many human cancers and has been previously linked to brain development. Our results offer an explanation as to how the tumor suppressive effect of chromatin remodeling complexes is mediated and specifically how mutations in ARID1B lead to some of the clinicopathological features observed in syndromic and non syndromic intellectual disability. Finally, our results indicate that while chromatin remodeling can have both activating and repressive effect on transcription, its repressive function could play an important role in human development and cancer through modulation of developmental/oncogenic pathways.

2919T

Cardiovascular manifestations in a family with GLA nonsense mutation (W162X). *E. Severin[1], C. Dragomir[2], A. Stan[2], G. Sarca[3].* 1) Dept Human Genetics, Carol Davila Univ Med Pharm, Bucharest, Romania; 2) Genetic Lab, Bucharest, Romania; 3) C. Angelescu Hospital, Bucharest, Romania.

Background: Previous studies reported that different mutations of GLA gene are associated with multiorgan damage. A three-generation Caucasian family with alpha-galactosidase A gene mutation is described. Complete or partial deficiency of alpha-galactosidase A results in the accumulation of globotriaosylceramide in cells throughout the body, leading to severe and progressive cardiovascular manifestations. Despite being recessive X-linked, carrier females are affected by Fabry disease and can express the same cardiovascular symptoms as males. Objectives: To describe and compare cardiovascular findings, to confirm the diagnosis of Fabry disease in male and female family members and to identify the mutation causing variable symptoms of cardiovascular phenotype. Patients and Methods: 2 males and 5 females of the family were enrolled and clinically assessed (including echocardiography); enzyme activity levels were also evaluated. Genomic DNA was isolated from blood samples of all family members and analyzed for GLA gene mutation. Results: Both male patients were found hemizygotes having one copy of GLA gene mutation (c.485G>A) and no enzyme alpha GAL activity. Also, all five females were found carrier of the same mutation and a broader range of residual alpha GAL activity. Despite having the same mutation, they exhibited different cardiovascular phenotypes. One male patient (31-year-old) had no cardiac involvement but his maternal uncle (56-year-old) presented left ventricular hypertrophy, cardiomyophaty and cerebrovascular manifestations. Mild left ventricular hypertrophy and moderate interventricular septum thickening were noticed in females by the third decade of life. None of these females received enzyme replacement therapy (ERT) prior to the study. Conclusions: Cardiovascular abnormalities were described in both male and female family members. Differences between genders were noticed but cardiovascular abnormalities increase with age and alpha GAL deficiency. Early genetic testing should be considered in younger female with a positive family history of Fabry disease. ERT should be instituted at an earlier age, particularly in carrier females with low alpha GAL activity, in order to limit severe expression of Fabry disease.

2920F

Unused Program Number

2921W

Analysis of *GH-releasing hormone (GHRH), GHRHR, GH1, PROP1, HESX1* and *GLI2* reveal the etiology of congenital growth hormone (GH) deficiency. *I.J.P. ARNHOLD, M.M. França, M.G.F. Osorio, S. Marui, M.Y. Nishi, L.R. Carvalho, A.P. Otto, F.A. Correa, E.F. Costalonga, A.A.L. Jorge, B.B. Mendonca.* Unidade de Endocrinologia do Desenvolvimento, Laboratorio de Hormonios e Genetica Molecular, LIM42, Divisao de Endocrinologia, Hospital das Clinicas da Faculdade de Medicina da Universidade de Sao Paulo, SAO PAULO, SP, Brazil.

Background: Mutations in transcription factors involved in pituitary organogenesis and in the GHRH-GH axis genes are involved in the etiology of congenital GH deficiency. **Objective:** to diagnose the etiology of congenital GH deficiency by selecting candidate genes based on a) the deficiency of GH alone (isolated GH deficiency, IGHD) or associated to other pituitary hormones (combined pituitary hormone deficiency, CPHD), and b) on magnetic resonance imaging (MRI) studies of the pituitary gland. **Patients and Methods:** Retrospective analysis of genetic results in 207 patients with congenital GH deficiency. GH deficiency was diagnosed by failure to respond to two GH-stimulation tests: clonidine test followed by combined insulin, TRH and GnRH test, which also evaluated other pituitary hormone deficiencies. *GH1, GHRH* and *GHRHR* were studied only in patients with IGHD. *PROP1* was studied in CPHD patients with topic posterior pituitary lobe (PPL). *HESX1* and *GLI2* were studied in patients with IGHD and CPHD irrespective of MRI findings. DNA was extracted from peripheral leukocytes and the coding regions and intronic boundaries of the respective genes were amplified by PCR and sequenced by Sanger method. **Results:** Among 57 patients with IGHD, 8 had homozygous *GH1* mutations, 7 (5 families) homozygous *GHRHR* mutations, all with a PPL in the normal position, but none *GHRH* mutations. Among the 140 patients with CPHD, 11 (9 families) had homozygous *PROP1* mutations, 1 patient a homozygous *HESX1* mutation and 27 patients heterozygous *GLI2* mutations (3 frameshift/stop codon and 24 non-synonymous; 26 CPHD and 18 ectopic PPL). One *GLI2* mutation was associated to GH deficiency in 4 additional family members. In one family *GLI2* mutation was associated to polydactyly and in another to midline facial defects. Holoprosencephaly and septo-optic dysplasia were not present. The pattern of inheritance was autosomal recessive and segregation with phenotype was complete for all genes except for *GLI2* which was autosomal dominant with incomplete penetrance and variable expression. **Conclusion:** Homozygous mutations in *GH1, GHRHR* and *PROP1* explained the etiology of congenital GH deficiency in 13% of patients. Heterozygous *GLI2* variants were the most frequent genetic abnormality (13%) but penetrance was variable. Hormonal testing and MRI of the pituitary are helpful to select candidate genes for the etiology of congenital GH deficiency.

2922T

Somatic mutations in *ATP1A1* and *ATP2B3* lead to aldosterone-producing adenomas and secondary hypertension. *T. Wieland[1], F. Beuschlein[2], S. Boulkroun[3,4], A. Osswald[2], H.N. Nielsen[5], U.D. Lichtenauer[2], D. Penton[6], V.R. Schack[5], L. Amar[3,4,7], E. Fischer[2], A. Walther[1], P. Tauber[6], T. Schwarzmayr[1], S. Diener[1], E. Graf[1], B. Allolio[8], B. Samson-Couterie[3,4], A. Benecke[9], M. Quinkler[10], F. Fallo[11], P-F. Plouin[3,4,7], F. Mantero[12], T. Meitinger[1,13,14], P. Mulatero[15], X. Jeunemaitre[3,4,7], R. Warth[6], B. Vilsen[5], M-C. Zennaro[3,4,7], T.M. Strom[1,13], M. Reincke[2].* 1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Medizinische Klinik und Poliklinik IV, Ludwig-Maximilians-Universität München, Munich, Germany; 3) Institut National de la Santé et de la Recherche Médicale (INSERM), Unité Mixte de Recherche Scientifique (UMRS) 970, Paris Cardiovascular Research Center, Paris, France; 4) Université Paris Descartes, Sorbonne Paris Cité, Paris, France; 5) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 6) Medizinische Zellbiologie, Universität Regensburg, Regensburg, Germany; 7) Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Paris, France; 8) Department of Medicine I, Endocrine and Diabetes Unit, University Hospital Würzburg, Würzburg, Germany; 9) Centre National de la Recherche Scientifique (CNRS), Institut des Hautes Etudes Scientifiques, Bures sur Yvette, France; 10) Clinical Endocrinology, Campus Mitte, University Hospital Charité, Berlin, Germany; 11) Department of Medicine, University of Padova, Padova, Italy; 12) Endocrine Unit, Department of Medicine, University of Padova, Padova, Italy; 13) Institute of Human Genetics, Technische Universität München, Munich, Germany; 14) DZHK (German Centre for Cardiovascular Research), partner site, Munich Heart Alliance, Munich, Germany; 15) Department of Medical Sciences, Division of Internal Medicine and Hypertension, University of Torino, Turin, Italy.

We started to investigate sporadic endocrine adenomas by sequencing tumor and matched control tissue. Here, we report our investigations in aldosterone-producing adenomas. Primary aldosteronism is the most prevalent form of secondary hypertension and is present in up to 7% of hypertensive individuals. Up to 60% of the primary aldosteronism cases are caused by unilateral aldosterone-producing adenomas (APAs). Recent reports showed that 30-40% of these APAs are caused by mutations in the potassium channel gene *KCNJ5*. To identify further genetic determinants of primary aldosteronism, we performed exome sequencing in APA-tumor and matched control tissue from nine males without *KCNJ5* mutation. We identified multiple somatic variants in two members of the P-type ATPase gene family, *ATP1A1* (encoding an Na^+/K^+ ATPase) and *ATP2B3* (encoding the plasma membrane Ca^{2+} ATPase). We then performed targeted resequencing in 299 additional APAs. The complete collection of APA samples (n=308) contained 16 (5.2%) somatic mutations in *ATP1A1* and 5 (1.6%) in *ATP2B3.* ATPase alterations were predominantly found in males (81%). Mutation-positive cases had significantly higher preoperative aldosterone concentrations (P=0.02) and lower potassium concentrations (P=0.013) compared to cases without mutation. All five identified alterations are either located in the transmembraneous α helix M1 or the juxtaposed α helix M4 in both *ATP1A1* and *ATP2B3.* The recurrence of mutations affecting these highly conserved regions involved in interaction with the transported cations in two paralogs is suggestive of a gain-of-function effect. Functional in vitro studies of *ATP1A1* mutants showed loss of pump activity and strongly reduced affinity for potassium. Electrophysiological ex vivo studies on primary adrenal adenoma cells provided evidence for inappropriate depolarization of cells with ATPase alterations, suggesting cation influx. In summary, gain-of-function alterations in two members of the ATPase gene family result in autonomous aldosterone secretion in roughly 7% of aldosterone-producing adenomas. The same mutational mechanism might be causative for other endocrine tumors. We will present our investigations in further endocrine tumor entities.

2923F

A novel homozygous mutation IVS6+5G>T of CYP11B1 gene of patient with Vietnamese congenital adrenal hyperplasia. *M.T.P. Nguyen[1,2], T.H Nguyen[2], N.D Ngo[1], C.D Dung[1], V.H Nong[2], T.L Nguyen[1], H.H Nguyen[2].* 1) Human Genetics, National Hospital of Pediatrics, Hanoi, Hanoi, Viet Nam; 2) Institute of Genome Research, Hanoi, Vietnam.

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disease which is characterized by a deficiency of one of the enzymes involved in the synthesis of cortisol from cholesterol by the adrenal cortex. More than 90% of CAH cases are due to 21-hydroxylase deficiency (21OHD), whereas 5-8% arise from 11β-hydroxylase deficiency. Mutations in the CYP11B1 gene are the cause of 11β-hydroxylase deficiency. This study performed on a patient with congenital adrenal hyperplasia, with premature development such as enlarged penis, muscle development, high blood pressure, bone age equivalent of 5 years old. Biochemical tests for steroids confirmed the diagnosis of CAH. We used PCR and sequencing to screen for mutations in CYP11B1 gene. Results showed that the patient has a novel homozygous mutation of guanine (G) to thymine (T) in intron 6 (IVS6+5G>T). We analysed this mutation on a MaxEntScan boundary software which showed this mutation could affect the gene splicing during transcription. This CYP11B1 genotype testing is the basis for genetic counseling and prenatal diagnosis in the future.

2924W

Alterations in the PAX8 promoter region cause thyroid dysgenesis.
P. Hermanns[1], M. Morlot[2], M. Donaldson[3], J. Jones[3], P. Pohlenz[1]. 1) Molecular Endocrinology, University Medical Center Mainz, Mainz, Germany; 2) Endokrinologikum Hannover, Centre for Hormone and Metabolic Diseases, Hannover, Germany; 3) 3Department of Child Health, Royal Hospital for Sick Children, Glasgow, Ireland.

Thyroid dysgenesis (TD) is the cause of approximately 80% of patients diagnosed to have congenital hypothyroidism. So far, mutations in five candidate genes (*PAX8*, *TSHR*, *TTF1*, *TTF2* and *NKX2.5*, respectively) have been identified to cause TD. We screened 190 patients with TD for mutations in these genes. Besides previously described and characterized mutations, we detected new so far unreported mutations in the coding regions of *PAX8*, *TSHR*, *TTF1*, *TTF2* and *NKX2.5*. 100 normal individuals did not harbour those mutations. Interestingly, in addition to mutations in the coding region of the known candidate genes we also detected four different base pair exchanges in the *PAX8* promoter region. Very recently, we have found one of these base pair changes in a patient diagnosed with thyroid dysgenesis who was heterozygous for a mutation in the *NKX2.5* gene and heterozygous for a base pair change in the *PAX8* promoter. *In vitro* studies were performed to unravel the mechanisms by which these newly identified promoter base pair exchanges might be causative. Electromobility shift assay (EMSA) studies suggest no specific protein or protein complex binding to the altered promoter elements. Transient transfection studies in different cell lines showed that at least one of these base pair changes leads to a significantly decreased promoter activity and thus to an impaired *PAX8* gene expression. In summary, we identified a new group of *PAX8* promoter sequence alterations that might cause TD. Further studies are needed to prove this hypothesis.

2925T

Shortening the diagnostic odyssey of patients with very early onset inflammatory bowel disease.
S. Drury[1], J. Kammermeier[2,3], L. Jenkins[1], M. Elawad[2], K. Gilmour[4], N. Lench[1], N. Shah[2]. 1) NE Thames Regional Genetics Service, Great Ormond Street Hospital, London, London, United Kingdom; 2) Department of Gastroenterology, Great Ormond Street Hospital for Children, London, WC1N 3JH; 3) Centre for Translational Genomics-GOSgene, UCL Institute of Child Health, London, WC1N 1EH, United Kingdom; 4) Department of Immunology, Great Ormond Street Hospital for Children, London, WC1N 3JH.

Very early onset inflammatory bowel disease (VEO - IBD) results in chronic intractable and therapy resistant inflammation of the gut and severe growth failure. The pervasive nature of the disorder often requires lengthy hospital admission. Despite underlying genetic heterogeneity, clinical presentation is often homogeneous. The differential diagnosis of VEO-IBD includes monogenic conditions with defects in epithelial barrier function, T/B lymphocyte selection and activation and innate and adaptive immune response as well as conditions leading to neutropenia and defective phagocyte killing and hyper- and autoinflammation. We have implemented a next generation sequencing approach to expedite the diagnosis of VEO-IBD patients, by screening 40 genes associated with the disorder. Coverage of the panel overall is 99% at 30x or greater. Thus far 5 positive controls and 6 VEO-IBD patients have been screened. All control mutations were detected and the pathogenic mutation in 1 previously undiagnosed patient has been identified. The subject presented with severe intractable inflammation of the large and small bowel and was diagnosed with autoimmune enterocolitis. Despite lack of diagnosis and due to the severity of the condition the subject has undergone experimental haematopoietic stem cell transplant (HCST), which has a mortality rate of 20%. Subsequent molecular analysis, showed the patient to be a compound heterozygote for mutations in *TTC37* (c.2018G>A p.Gly673Asp and c.2808G>A p.Trp936*), associated with tricho-hepato-enteric syndrome (THES), a very rare disease (1/1,000,000 births). The phenotype of this patient is unlike the characteristic features reported in other cases; thus NGS has enabled a diagnosis of a) a rare disorder not characterised by phenotype alone and b) a genetic test which until now has not been offered clinically. Due to the rarity of THES, it is currently unknown whether HCST is optimal treatment for the condition. Identifying the molecular basis of VEO-IBD has the potential to profoundly affect patient management; e.g. patients with mutations in *IL10* can be successfully treated with HCST. In addition 42 primary immune deficiency genes are also captured with the same panel and analysis can be extended to these genes, offering a comprehensive molecular approach to this complex group of patients. We aim to introduce this extensive genetic screening early on in patient management, to guide medical intervention at the earliest opportunity.

2926F

Patient derived somatic and induced pluripotent stem cells as a model for functional assessment of mutations identified by exome sequencing in congenital diarrheal disorders.
M. Yourshaw[1], A. Vega-Crespo[2], R. Solorzano-Vargas[2], J. Wang[2], S. Stanford[2], C. Sosa[2], S. Nelson[2], J. Byrne[2], M. Martín[2]. 1) Human Genetics, UCLA Geffen School of Medicine, Los Angeles, CA; 2) UCLA Geffen School of Medicine, Los Angeles CA 90024.

Background: Children presenting with intractable diarrhea often have a poor prognosis; after a sometimes fruitless diagnostic odyssey they can end up with prolonged intravenous nutrition and/or intestinal transplant. Morbidity and mortality are high, and medical care costs overwhelming. Our pilot study of the use of whole exome sequencing to diagnose difficult cases of congenital diarrhea found convincing genetic evidence of mutations in genes that are known to cause malabsorptive diarrhea or other conditions, including ADAM17 (inflammatory skin and malabsorptive diarrhea), EPCAM (Tufting enteropathy), MYO5B (diarrhea with microvillus atrophy), NEUROG3 (generalized malabsorptive diarrhea), PCSK1 (PC1/3 deficiency), SI (sucrase isomaltase deficiency), and SLC5A1 (glucose/galactose malabsorption), as well as a number of novel candidate genes. It is now possible to model these disorders in gut epithelium generated from somatic intestinal stem cells and induced pluripotent stem cells (iPSCs). **Methods:** We performed exome sequencing on cases with congenital malabsorptive, secretory, or IBD-like diarrhea having no convincing diagnosis after a conventional medical and genetic workup. To model these disorders we took two approaches: 1) isolate somatic intestinal stem cells from patients' endoscopic biopsies or surgical samples from normal controls, or 2) derive induced pluripotent stem cells (iPSCs) from dermal fibroblasts obtained by skin biopsies, and subsequently direct the iPSCs towards intestinal epithelial fate. We used methods including RNAseq, immunofluorescence, and other assays to assess the functional effect of mutations identified by exome sequencing in patient cell lines, and normal cell lines in which selective genes are depleted using shRNA methods. **Results:** We developed disease-specific somatic intestinal stem cell and pluripotent cells lines, characterized them for reprogramming by assessing their expression of pluripotency markers, their karyotype, and their ability to form representatives of all three germ layers. We have differentiated disease-specific iPSC lines into intestinal epithelium for disease modeling purposes and identified phenotypes caused by mutations in a number of the genes. **Conclusion:** The development of somatic intestinal stem cell and iPSC lines has allowed us to model severe diarrheal disorders within a dish and provides a potential foundation for future personalized cellular therapeutics for digestive system disorders.

2927W

Exome sequencing of a familial trio with a suspected autosomal dominant idiopathic immune deficient syndrome identifies novel candidate mutations in the complement system and two other genes.
R. Golhar[1], J. Liang[2], D. Li[1], C. Kao[1], Y. Guo[1], W. Chen[2], L. Tian[1], F. Wang[1], J. Synder[1], N. Abdel-Magid[1], L. Vazquez[1], B. Keating[1,3,4], J. Zhang[2], H. Hakonarson[1,3,4]. 1) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

We sequenced the exomes of a familial trio with an affected mother and child showing signs of an autosomal dominant idiopathic immune deficient syndrome to determine the potential underlying genetic cause. The affected child suffers from chronic sinusitis, conjunctivitis, constipation, and recurrent skin lesions. The mother suffers from chronic respiratory infections, recurrent bacterial and viral infections. Both probands have a history of bacteremia and maintain a low white-blood cell count during infections without fever. We suspected an autosomal dominant inherited deleterious mutation in one of the Toll-like receptors or within the TLR signaling pathway. The Toll-like receptors detect pathogens by recognizing specific microbial components. Exome sequencing revealed a splice acceptor site mutation in NEMO at chrX:153,792,177 (T>C). The child appeared homozygote, while the mother and father both appeared heterozygote. Knowing the father cannot be heterozygote, we determined the second half of NEMO is duplicated approximately 100kb downstream on the 3' strand. Follow-up Sanger sequencing of NEMO and the NEMO-homolog using long-range PCR show the child has the mutation in both NEMO and the NEMO-homolog, while the father has the mutation in NEMO but not in the NEMO-homolog. As the father is unaffected, this rules out NEMO as a candidate. We performed a more exhaustive variant reduction analysis revealing a novel heterozygote deletion in C2 (chr6:31902193-4, AG) and 2 heterozygote mutations, in LILRB1 (chr19:55143563, A/G) and KIR3DL3 (chr19:55246741, A/T). C2 is part of the classical complement system which interacts with the adaptive immune system to clear pathogens. We suspect the adaptive immune system is attempting to recruit the complement system, however this 2bp deletion renders C2 unable to function resulting in a lack of immune response. LILRB1 is part of the leukocyte immunoglobulin-like receptor family, and is expressed on immune cells. LILRB1 controls inflammatory responses to focus immune response and limit autoreactivity. KIR3DL3 is killer cell immunoglobulin-like receptor expressed by natural killer cells, subsets of T cells and is thought to play a role in regulation of immune response. Both LILRB1 and KIR3DL3 exist in highly variable regions and the genetic variation observed may be normal. All three mutations need further functional validation before we can conclusively attribute them to this idiopathic immune deficient syndrome.

2928T

SNPs associated with cerebrovascular accident in a Brazilian cohort of sickle cell anemia patients.
P.R.S. Cruz[1], G. Ananina[1], F. Menaa[1,2], M.A.C. Bezerra[3], A.S. Araujo[3], G.P. Gil[1], W.M. Avelar[4], F. Cendes[4], F.F. Costa[2], M.B. Melo[1,2]. 1) Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas, Campinas, São Paulo, Brazil; 2) Hematology and Hemotherapy Center/HEMOCENTRO, University of Campinas, Campinas, São Paulo, Brazil; 3) Hematology and Hemotherapy Center of Pernambuco (HEMOPE), Recife, Pernambuco, Brazil; 4) Neuroimaging Laboratory, Department of Neurology, UNICAMP, Campinas, São Paulo, Brazil.

Although sickle cell anemia (SCA) results from the homozygosity for a single mutation in amino acid 6 of the β-hemoglobin *locus*, this disease presents high phenotypic heterogeneity. Virtually all organs may be affected by vascular occlusion, with emphasis on the Central Nervous System, where are observed transient ischemic attacks, stroke and cerebral hemorrhage, which affect approximately 25% of patients with SCA. Early identification of patients with sickle cell anemia susceptible to to cerebrovascular accident (CVA) could reduce the risk, possibly preventing the recurrence of infarcts and potentially reducing their incidence. Thus, we proposed to investigate the frequency of Single Nucleotide Polymorphisms (SNPs), using high-density microarrays (Genome-Wide Human SNP Array 6.0) in order to identify genomic regions potentially involved in the increased risk of CVA in sickle cell patients. SCA patients with and without CVA were recruited at two Brazilian centers of hematology: Hemocentro (Campinas, SP) and HEMOPE (Recife, PE). Patients' recruitment and blood collection methods were approved by the Ethics Committee of FCM-UNICAMP (Campinas-SP, Brazil). The genotyping protocol was carried out using Affymetrix® Genome-Wide Human SNP 6.0 Array (Affymetrix Inc., CA, USA) according to manufacturer's recommendations. SNP calling was performed with Affymetrix Genotyping Console software. We filtered SNP data for probe contrast quality (QC call rate > 0.40), excluded individuals with more than 5% of missing genotypes, as well as removed SNPs with call rate less than 95%, minor allele frequency below 5% and/or presenting deviation from Hardy-Weinberg equilibrium. Population structure was assessed by Principal Component Analysis (controlling for linkage disequilibrium) and did not suggest difference in subjects' ancestry. After filtering, the dataset consisted of 26 cases and 25 controls, which together underwent Fisher's exact test for 662,267 autosomal variants. Three intergenic SNPs were associated with CVA (located on chromosome 2, p-value: 4.8e-06; and chromosome 6, p-values: 6.2e-06 and 7.2e-07); chromosome 3 harbored two associated SNPs (on *SLC9A9* gene, p-value: 8.8e-07; and on a lincRNA, p-value: 8.8e-06), while chromosome 14 had a SNP on a serine/threonine kinase (*PRKD1*), suggestively associated to CVA (p-value: 5.3e-07). None of the variants was previously associated with CVA, possibly reflecting the singularity of the Brazilian population.

2929F

Clinical, Immunological, and Molecular Characterization of JAK3 Deficiency causing Severe Combined Immunodeficiency Disease in Saudi Arabia.
A. Hawwari[1], H. Al-Shammari[2], S. Al-Hisi[1], O. A-Smadi[1], H. Al-Dhekri[2], A. Al-Ghonaium[2], S. Al-Muhsen[2], B. AL-Saud[2], R. Arnaout[2], H. Al-Mousa[2,3]. 1) Genetics, Research Center, MBC: 03, King Faisal Specialist Hospital & Research Center, PO Box 3354, Riyadh 11211, Saudi Arabia,; 2) Department of Pediatrics, King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia; 3) Al-Faisal University, Riyadh, Saudi Arabia.

Molecular defects in IL2Rg gene are the most common genetic cause of T-B+NK- SCID worldwide followed by JAK3. However, no report has been published to describe the genetic defects of T-B+NK- SCID in Saudi Arabia. Due to high consanguinity rate in Saudi society, most cases that we observe in our patient population are mainly due to autosomal recessive pattern of inheritance. Hence, patients with T-B+NK- SCID were screened for mutations in JAK3 gene. Eight patients from 5 families diagnosed with T-B+NK- were screened for mutations in JAK3 gene. The diagnosis of T-B+NK- SCID is based on the profound, characteristic lymphopenia and early clinical presentation of immunodeficiency often with strong family history indicating an autosomal recessive inheritance. Five novel homozygous JAK3 mutations were identified. Three closely related patients (two siblings and a cousin) carry the nonsense mutation (Q305X) ; one patient was found to have a deletion mutation (A338LfsX14) leading to a stop codon 14 amino acid downstream; and two brother patients carry the R403C missense mutation. The remaining two patients carry either E1019K or R103H missense mutations. Contrary to the available data of X-linked IL-2 common gamma chain gene defect of T-B+NK- SCID as the most common, we demonstrate here that in the Kingdom of Saudi Arabia the most common form of T-B+NK- SCID is the autosomal recessive mode of inheritance associated with JAK3 deficiency.

2930W

Likely pathogenic hypomorphic mutation in the perforin 1 gene causing adult-onset familial hemophagocytic lymphohistiocytosis.
L. Massingham[1], J. Walsh[1], N. Shur[1], C. Benson[1], P. Rintels[2], N. Berliner[3], D. Treaba[4], J. Li[4], C. Phornphutkul[1]. 1) Department of Pediatrics, Division of Genetics, Hasbro Children's Hospital and Rhode Island Hospital, Providence, RI; 2) Division of Hematology, Rhode Island Hospital, Providence, RI; 3) Divsion of Hematology, Brigham and Women's Hospital, Boston, MA; 4) Department of Pathology, Rhode Island Hospital, Providence, RI.

Familial hemophagocytic lymphohistiocytosis (FHL) is characterized by fever and hepatosplenomegaly with cytopenias and hyperactivated macrophages and T-lymphocytes. FHL is inherited in an autosomal recessive manner and currently there are five known molecular subtypes (FHL1-5) with overlapping phenotypes. The age of onset is generally in infancy or the first few years of life, but more recently adult-onset FHL has been reported. Distinguishing between FHL and secondary (acquired) hemophagocytic lymphohistiocytosis (HLH) is difficult and treatment recommendations vary considerably. Mutations in the perforin 1 (PRF1) gene account for 80% of patients with FHL. Case: A 30 year old previously healthy man presented with a 2 week history of cough and fever. Initial labs revealed pancytopenia, elevated AST and ALT, low fibrinogen and elevated ferritin. Bone marrow biopsy revealed hypercellularity with megakaryocyte hyperplasia. Flow cytometry did not reveal any T-cell or B-cell clonal populations. IHC staining revealed large clusters of T cells. An extensive infectious evaluation was negative. CT scan revealed splenomegaly and a diffuse pulmonary interstitial infiltrate. He was presumed to have hemophagocytic lymphohistiocytosis and chemotherapy was initiated. A persistently abnormal coagulation profile prompted a repeat bone marrow biopsy, which showed a hypocellular marrow with no clonal populations by T-cell gene rearrangement or B-cell clonality studies. Upon readmission for fever, epistaxis, and pancytopenia, genetics was consulted and FHL testing was sent, which revealed a homozygous change for c.272C>T (p.Ala91Val) in the PRF1 of unknown significance. Conclusion: Our patient meets clinical criteria for HLH and the molecular results support this diagnosis. This missense mutation has been classified by some as a neutral polymorphism, but other functional studies have suggested that the p.Ala91Val variant may be hypomorphic. Our patient's late presentation is consistent with the hypothesis that missense mutations in PRF1 are hypomorphic, whereas, individuals with nonsense mutations typically present much earlier. Studies by the Filipovich group have identified hypomorphic alleles in approximately 15% of adult onset HLH. Molecular results in this patient supports a genetically inherited predisposition to HLH. Unlike acquired HLH, genetic counseling recommendations for patients with adult onset familial hypomorphic HLH is very different.

2931T

Ancestry and genetic admixture among sickle cell disease patients in North America.
Z. Wang[1], L. Diaw[1], M. Barr[1], M. Quinn[1], D. Diggs[1], A. Oguhebe[1], D. Darbari[1,2], A. Hutchinson[3], C. Hoppe[4], J.G. Taylor[1]. 1) Genomic Medicine Section, Hematology Branch, NHLBI, NIH, Bethesda, MD, USA; 2) Center for Cancer and Blood Diseases, Children's National Medical Center, Washington, DC, USA; 3) Cancer Genetics Research Laboratory, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA; 4) Oakland Children's Hospital, Oakland, CA, USA.

Genome wide association studies (GWAS) are an important tool for identifying complex human disease loci. An important variable for GWAS is the degree of admixture occurring within study populations. If undetected, admixture can lead to spurious associations, although it may also be exploited as a strategy for mapping by admixture linkage disequilibrium (MALD). As part of our efforts to identify genetic modifiers of sickle cell disease (SCD), we have examined ancestry and genetic admixture in SCD patients living in North America. We first examined ancestry from family histories for 649 adults with SCD recruited to the Bethesda Sickle Cell Cohort Study in the eastern US, where 61% of subjects are African American, 22% African, 12% Caribbean or South American and 1.8% of other origins, suggesting a high degree of population diversity. We further examined these observations with genetic markers by identifying 364 ancestry informative markers (AIMs) from over 4 million SNPs typed in Europeans (CEU) and Yorubans (YRI, Nigeria) available from the Human Haplotype Map. AIMs were typed in 445 of these SCD cohort subjects. Principle components analysis showed African American or Caribbean/South American subjects had the highest degree and a wider range of admixture compared to Africans with SCD or HapMap populations (CEU, YRI and CHB). STRUCTURE analysis showed similar results. We also performed a replication study using 471 anonymous DNA samples from SCD subjects in a western US newborn screen cohort. Similar to the analysis of population ancestry in the Bethesda cohort, the SCD newborns are highly admixed, plotting between YRI and CEU populations. However, the degree of admixture is different when comparing these 2 SCD populations from eastern and western regions of the US. Overall, SCD patients in North America have a high degree of genetic admixture that must be accounted for in candidate gene or GWA studies. Furthermore, the degree of admixture adjustment could vary for SCD subjects from different geographic regions of the US. Finally, these SCD populations may be ideal for mapping by MALD to identify genetic modifiers of this monogenic disease.

2932F

Erythropoiesis failure and ribosomal dysfunction in zebrafish model of Diamond-Blackfan anemia.
N. Kenmochi[1], T. Uechi[1], Y. Nakajima[1], G. Yadav[1], T. Sawada[1], M. Ikeda[2]. 1) Frontier Sci Res Ctr, Univ Miyazaki, Miyazaki, Japan; 2) Veterinary Pharmacology Dept, Univ Miyazaki, Miyazaki, Japan.

Ribosomes, the molecular factories that carry out protein synthesis, are essential for every living cell and ribosomal proteins (RPs) play important roles in the formation of a functional ribosome. Defects in ribosome biogenesis have been linked to many human diseases called ribosomopathies, a rare collection of genetic disorders that are associated with increased cancer susceptibility. Diamond- Blackfan anemia (DBA) represents the first and the most extensively studied human disease caused by defects in ribosomal proteins. RPS19 is most commonly mutated in DBA, although some patients show mutations in several other RP genes.

To investigate the molecular pathogenesis of DBA, we have developed a zebrafish model of DBA by knocking down the *Italic rps19* gene in the embryos using a morpholino oligonucleotide (MO). The knockdown embryos displayed a drastic reduction of red blood cells, whereas differentiation of other myeloid cells and endothelial cells seemed to be normal. The anemia phenotype was almost completely rescued by injection of synthesized *rps19* mRNA, but not by mutated mRNAs with patient-type mutations. The DBA model also showed developmental abnormalities in the head and tail regions due to increased cellular apoptosis. Simultaneous inhibition of *tp53* rescued the morphological abnormalities but did not alleviate the erythroid aplasia, suggesting that a TP53-independent but RPS19-dependent pathway could be responsible for defective erythropoiesis in DBA. To evaluate the impact of RPS19 deficiency on mRNA translation, we carried out polysome analyses. The polysome patterns were similar between rps19 morphants and control embryos but the amount of heavier fractions was less in the morphants. Finally, we carried out an RNA-Seq analysis of polysomal mRNAs that were purified from the embryos. We found that translational efficiencies of about 300 genes were significantly changed in rps19 morphants. These data suggest that selected mRNA translation may play an important role for the pathogenesis of DBA.

2933W

A family based exome sequence analysis identifies a rare AID deficiency causing mutation enriched in Finland. *L. Trotta[1], H. Almusa[1], M. Lepisto[1], P. Ellonen[1], S. Hannula[1], A. Palotie[1,2], K. Porkka[3], M. Seppanen[4], J. Saarela[1].* 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 3) Hematology Research Unit Helsinki, Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 4) Immunodeficiency Unit, Division of Infectious Diseases, Helsinki University Central Hospital, Hospital District of Helsinki and Uusimaa. Aurora Hospital, Helsinki, Finland.

Primary immune deficiencies (PIDs) are a large and heterogeneous group of congenital disorders of the immune function due to genetic defects impairing essential branches of the immune system. PIDs predispose affected individuals to increased susceptibility to recurrent and persistent infections and autoimmunity. PIDs are mainly inherited as rare monogenic disorders with more than 200 different genetic etiologies (5670 known mutations in genes with immune function) occurring in ~1:500-10,000 live births, with differences among countries and a much higher prevalence of some disorders in cases with high consanguinity rates or among genetically isolated populations. The molecular mechanisms behind many forms are not yet known and the significant clinical and immunological heterogeneity often delay the diagnosis and make treatment of PIDs challenging. Several cases of PIDs have been described in Finland, and because its unique genetic architecture the Finnish population has been defined as a model population for human genetic studies, especially for monogenic disorders. A founder effect, drifts in subsequent subisolates and rapid regional expansions shaped the gene pool of the population leading to the peculiar enrichment of rare variant responsible for the occurrence of the 'Finnish disease heritage' (i.e. more than 40 rare genetic disorders enriched in Finland). An exome sequencing analysis of a Finnish PID family with 5 affected siblings, characterized by lymphatic hyperplasia, autoimmunity and recurrent respiratory tract infections, as well as missing IgA, IgG, IgE levels and switched memory B cells, led us to identify the p.Met139Thr mutation in the AICDA gene. The gene, previously reported to cause AID deficiency, encodes for a single-stranded DNA-specific cytidine deaminase required for several crucial steps of B-cell terminal differentiation necessary for efficient antibody responses. The identified mutation is a rare variant, not reported in the 1000 Genomes database. We further monitored the frequency of the mutation in ~4000 Finnish individuals with sequence level data (kindly provided by the SISu Consortium) observing it as very rare (0,152%), however 12 fold higher in Finland than in the general European population (NHLBI Exome Sequencing Project data). This suggests an enrichment of the variant in the Finnish population and further studies may provide evidence to merit the disease to be included in the list of 'Finnish disease heritage'.

2934T

Exome sequencing identifies *NFKB2* mutations as a cause of autosomal dominant early-onset common variable immunodeficiency. *K. Chen[1], E. Coonrod[2], A. Kumanovics[2,5], Z. Franks[7], J. Durtschi[2], R. Margraf[2], W. Wu[6], N. Augustine[5], P. Ridge[2], H. Hill[2,3,4,5], L. Jorde[6], A. Weyrich[7], G. Zimmerman[3], J. Bohnsack[1], K. Voelkerding[2].* 1) Dept. of Pediatrics, Allergy & Immunology, University of Utah School of Medicine, Salt Lake City, UT; 2) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 3) Dept. of Medicine, University of Utah School of Medicine, Salt Lake City, UT; 4) Dept. of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT; 5) Dept. of Pathology, University of Utah School of Medicine, Salt Lake City, UT; 6) Dept. of Human Genetics, University of Utah, Salt Lake City, UT; 7) The Molecular Medicine Program, University of Utah, Salt Lake City, UT.

Common variable immunodeficiency (CVID) is a heterogeneous disorder involving low antibody levels, poor response to antigens, and immune dysregulation. CVID is one of the most common primary immune deficiencies, occurring in approximately 1:10,000 to 1:25,000 people. The genetic defects responsible for CVID have been identified in less than 15% of cases. To investigate the molecular cause of CVID, we studied a family affected by an autosomal dominant form of early-onset CVID. Exome sequencing analysis of four members of this family identified a heterozygous frameshift deletion (c.2563delA; p.Lys855SerfsX7) in the gene that encodes nuclear factor kappa-B2 (*NFKB2*; p100/p52, also known as p100/p49). This mutation is predicted to both truncate the protein and prevent nuclear translocation due to impairment of phosphorylation and proteosomal processing of the C-terminus into the active protein. Using long-range PCR and next-generation sequencing, we then screened for *NFKB2* mutations in a cohort of 33 additional CVID patients and identified a second heterozygous nonsense mutation at the same C-terminal locus (c.2557*C>T; p.Arg853X) in an unrelated CVID patient. All four patients from the two families with *NFKB2* mutations presented with an unusual combination of childhood-onset CVID and adrenal insufficiency. Variant Annotation, Analysis, and Search Tool (VAAST) analysis of the exome sequence data from the index family and an unrelated CVID patient ranked *NFKB2* highest and reached genome-wide significance ($p< 1.56\times10^{-11}$). Western blotting and immunocytochemical analysis of EBV-immortalized B cells from affected subjects demonstrated the presence of mutant NF-κB2 protein and decreased nuclear translocation. NF-κB2 is the principal protein involved in the noncanonical NF-κB pathway, is evolutionarily conserved, and functions in peripheral lymphoid organ development, B cell development, and antibody production. The identification of *NFKB2* mutations will further elucidate the pathogenic mechanisms of CVID and can provide potential targets for treatment. These findings describe the first primary immunodeficiency syndrome caused by a germline mutation in the noncanonical NF-κB signaling pathway.

2935F

Lobar holoprosencephaly (HPE) associated with additional clinical anomalies in two daughters of a consanguineous couple. Comparison of SNP-array analysis results as an attempt to search for a potential causative candidate gene. *P.M. Kroisel[1], B. Plecko[2], M. Brunner-Krainz[3], M.R. Speicher[1], C. Windpassinger[1].* 1) Human Genetics, Medical University of Graz, Graz, Styria, Austria; 2) Department of Pediatrics, University of Zurich, Zurich, Switzerland; 3) Department of Pediatrics, Medical University of Graz, Graz, Austria.

A consanguineous family (both parents are first degree cousins) has five children, three boys and two girls. A lobar holoprosencephaly (HPE) was already diagnosed prenatally in both female pregnancies. Male offsprings are healthy, however one of them shows a mild form of UDP-galactose-4-epimerase deficiency. In the first affected girl by prenatal sonographic investigation HPE was suspected in the 23. gestational week and was confirmed by MRI in the 26. gestational week. Cytogenetic analysis and array CGH performed after birth at term showed normal results but distinct facial features like hypertelorism, a depressed nasal bridge, low set simplified ears and a general increased skin hair density and length were recognized. Beside brain anomalies like the mentioned lobar HPE, a ventriculomegaly and a partial pachygyria was found. Psychomotor development was strongly reduced, and seizures, which became more frequent and severe, finally caused a fatal outcome at an age of 3 years. Her sister, who was born two years later shows a nearly identical clinical phenotype with lobar HPE, very similar facial and skin hair anomalies but also a hypoplastic vermis. Seizures are however clearly less frequent and antiepileptic treatment is more successful. However progress in her neurological development is of course related to her brain anomalies very severely delayed. Cytogenetic and array CGH analysis using a 60k Agilent oligonucleotide array also did not reveal a genomic unbalance or aberration. Based on these findings a monogenic form of HPE appears to be likely in our two patients. At least 14 genes have been implicated in HPE, with 4 major genes including SSH and 10 minor genes. We still had sufficient DNA of the deceased patient and her severely affected sister, who up to now seems to have a slightly better prognosis. Therefore we performed a 250k Affymetrixs SNP-array analysis of both DNA-samples in an attempt by homozygosity mapping and comparison of the results in our two patients to narrow down potential HPE candidate gene regions. By this approach we were able to reduce the number and size of homozygous chromosomal segments shared by both sisters to two chromosomal areas of less than about 2-3 Mb at 10p11 and 11q24. Only to the latter locus a possible minor candidate gene (CDON) has already been mapped. If analysis of this gene would fail to show a homozygous mutation, NGS analysis could be considered to identify a possible new causative gene for HPE.

2936W

Infantile-onset ascending hereditary spastic paraplegia with bulbar involvement due to the novel ALS2 mutation c.2761C>T. *S. Majid[1], R. Abuthuraya[1], K. Ramzan[1], S. Hagos[1], H. Al-Dossari[1], R. Al-Omar[1], H. Murad[2], A. Chedrawi[2], Z. Al-Hassnan[3], S. Bohlega[2].* 1) Department of Genetics, KFSHRC, RIYADH, Saudi Arabia; 2) Department of Neurosciences, KFSHRC, RIYADH, Saudi Arabia; 3) Department of Department of Medical Genetics, KFSHRC, RIYADH, Saudi Arabia.

Mutations in the alsin gene cause three clinically distinct motor neuron diseases, infantile-onset ascending hereditary spastic paraplegia (IAHSP), juvenile primary lateral sclerosis (JPLS), and juvenile amyotrophic lateral sclerosis (ALS2). A total of 23 different ALS2 mutations have been described for the three disorders so far. Most of these mutations are predicted to result in a frameshift leading to premature truncation of the alsin protein. We report the novel ALS2 truncating mutation c.2761C>T detected by homozygosity mapping and sequencing in two infants affected by IAHSP with bulbar involvement. The mutation c.2761C>T resides in the pleckstrin domain, a characteristic segment of guanine nucleotide exchange factors (GEFs) of the Rho GTPase family, which is involved in the overall neuronal development or maintenance. This study highlights the importance of using homozygosity mapping combined with candidate gene analysis to identify the underlying genetic defect as in this Saudi consanguineous family. This study suggests genetic screening and consideration of preimplantation genetic diagnosis in IAHSP families.

2937T

A Mutation in the SCN10A Voltage-gated Sodium Channel Genes Causes a Novel Autosomal Recessive Progressive Neuromuscular Disease and Epilepsy-related Syndrome. *M. Kambouris[1,2], T. Ben-Omran[3,4], Y. Al-Sarraj[1], Y. Bejaoui[1], M. Almuriekhi[3], H. Boulos[1], H. El-Shanti[1,5].* 1) Shafallah Medical Genetics Center, Doha, Qatar; 2) Yale University School of Medicine, Genetics, New Haven CT, USA; 3) Clinical & Metabolic Genetics, Pediatrics, Hamad Medical Corporation, Doha, Qatar; 4) Weill Cornell Medical College, Doha, Qatar; 5) University of Iowa, Pediatrics, Iowa City, IA, USA.

A consanguineous family of Sudanese ethnic origin affected by a putative novel autosomal recessive progressive neuromuscular disease associated with epilepsy was studied by homozygosity mapping and whole exome sequencing of a single affected individual to identify the responsible gene and defect. Clinical characteristics of the two affected siblings include moderate to severe intellectual disability; muscle weakness due to upper motor neuron lesion; -both siblings are non-ambulatory- anhydrosis; recurrent seizures and distinctive facial appearance with prominent forehead, long face & broad nasal bridge. The responsible gene was mapped to four possible intervals genome linkage [[3p, 11p/q, 11q, 14q], as the family structure did not allow the identification of a single interval with a significant LOD score. Whole exome target enrichment sequencing was performed on ABI SOLiD4 for a single-affected individual. A c.C4514T / p.T1505M homozygous mutation affecting a highly conserved amino acid residue was identified in the SCN10A gene -a member of the voltage-gated sodium channel gene family at the 3p22 linkage interval- with damaging effects according to PolyPhen and SIFT protein-modeling software. The variant validated by Sanger sequencing, co-segregates with the disease phenotype within the family and it is absent in 2000 ethnically matched control chromosomes. The affected amino acid residue [1505] is positioned between helices S1 and S2 of domain IV (DIV) of the sodium channel. The mutation is spatially located between the two helices in a putative extracellular position. Since epilepsy is also observed alongside mental retardation and peripheral neuropathy in the mutant SCN10A gene, the mutation probably interferes with normal functioning of this particular sodium channel, resulting in aberrant transmission of nerve impulses. Sodium channels are responsible for the instigation and proliferation of action potentials both in the central and peripheral nervous systems. Mutations, mostly autosomal dominant, affecting sodium channel genes are responsible for a range of epileptic or seizure-related disorders. This is the first autosomal recessive, epilepsy related disease involving a sodium channel gene.

2938F

Duplication of EYA1 causes Branchiootic Syndrome in a Brazilian family. *R. Mingroni-Netto[1], V.G.L. Dantas[1], E.L. Freitas[1], A.M.S.M. Moraes[2], M.C.C. Braga[1], S.B. Ramos[3], C. Rosenberg[1], V.A. Della-Rosa[4].* 1) Centro de Estudos do Genoma Humano, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; 2) Departamento de Medicina, Universidade Estadual de Maringá, Paraná, Brazil. Brazil; 3) Centro Universitário de Maringá, Maringá, Paraná, Brazil; 4) Departamento de Biotecnologia, Genética e Biologia Celular, Universidade Estadual de Maringá, Paraná, Brazil.

Branchiootic syndrome 1 (BOS1; OMIM#602588) is an autosomal dominant disorder characterized by malformations of the outer, middle, and inner ear associated with conductive, sensorioneural, or mixed hearing impairment, branchial fistulae and renal cysts. The penetrance is high, but incomplete and the phenotype extremely variable. A large pedigree with 12 affected individuals presenting with hearing loss and ear malformations was ascertained in the state of Paraná in southern Brazil. After clinical exams and computer tomography (five individuals), malformations of outer, middle and inner ear, auricular and cervical fistulae were detected, albeit without renal abnormalities. Genomic scanning was performed on DNA from 11 affected individuals using 50k microarray technology from Affymetrix (GeneChip® Human Mapping K Array Xba 240). Multipoint lod scores were calculated under an autosomal dominant model with a penetrance estimated to be 90%. Using MERLIN, a peak lod score of 2.6 was obtained for markers in the EYA1 gene region on chromosome 8. Sequencing of EYA1 did not reveal pathogenic mutations. Array-CGH (Oxford Gene Technology, 180k) was then performed with DNA samples from 2 affected individuals and a duplication including the EYA1 gene was detected (hg18 8:72,104,245-72,756,772). No other gene was included in the duplicated segment. In order to confirm this alteration and to investigate the segregation of the duplication, real time PCR was performed. The duplication was present in all 11 affected individuals investigated and absent in the 7 unaffected individuals. Our findings implicate this EYA1 duplication in the BOS1 phenotype observed in this pedigree.

2939W

DNA Copy-Number Variations in Prune Belly Syndrome. *S. Harrison[1], L. Baker[1,2].* 1) Department of Urology, University of Texas Southwestern Medical Center, Dallas, TX; 2) McDermott Center for Human Growth and Development, University of Texas Southwestern, Dallas, TX.

INTRODUCTION: Prune belly syndrome (PBS) is a rare complex of abnormal abdominal wall musculature, urinary tract anomalies, and bilateral cryptorchidism, and is associated with significant morbidity. The cause of PBS is unknown, however a male predominance (95%) and identified families with multiple affected individuals suggests PBS may be an X-linked or sex-limited genetic disease. Disease-causing DNA changes can vary from chromosomal alterations detectable by karyotyping to point mutations detected by DNA sequencing. More recently discovered, copy-number variations (CNVs) are small deletions or duplications that affect the number of copies of a gene. CNVs are missed by DNA sequencing and karyotyping but have been shown to account for previously unexplained genetic diseases by identifying candidate genes or regions involved in various disorders. The purpose of this study was to assess if novel CNVs are present in PBS patients. MATERIALS & METHODS: From 2008-current, blood samples from PBS patients were prospectively tested by whole genome comparative genomic hybridization (CGH) using the array version current (V8.1 or V8.3) at the time of ascertainment (Baylor Medical Genetics Laboratory, Houston TX). When possible, in cases wherein novel genomic CNVs were detected, parental testing (via FISH probes) was performed to identify whether CNVs are de novo or inherited. RESULTS: CGH testing on 12 PBS cases identified 5 (42%) patients with novel CNVs. Two patients have duplication CNVs: one patient has a maternally inherited 0.360Mb duplication on 2q11.2 and one patient has a 0.324Mb duplication on Xq23 of unknown genetic origin. Two patients have deletion CNVs: one patient has a paternally inherited 0.271Mb deletion on 7q31.1 and one patient has a maternally inherited 0.113Mb deletion on Xq22.1. Additionally, one adopted patient has two 0.066-0.151Mb deletions on 6q23.2 and 7q33.2 of unknown genetic origin. CONCLUSIONS: Prune belly syndrome (PBS) is a rare but morbid birth defect and novel CNVs are common in these males (42%). The finding that 2 of the 5 identified CNVs are located on chromosome X supports the hypothesis that PBS is an X-linked genetic disease. Further investigation of these genomic rearrangements may lead to the identification of genetic causes of PBS, thereby aiding prenatal diagnosis and genetic counseling.

2940T

A Mutation at the H2B Histone Family, member W [H2BFWT] gene causes a novel X-linked mental retardation with abnormal head shape syndrome. *V. Chini[1], R. Ali[2], N. Khattab[1], T. Bin Omran[2], Y. Al-Sarraj[1], M. Kambouris[1,3], H. El-Shanti[1,4].* 1) Molecular Genetics, Shafallah Medical Genetics Center, Doha, Qatar; 2) Pediatric Department, Hamad Medical Corporation, Doha, Qatar; 3) Yale University School of MEdicine, Genetics, New Haven CT, USA; 4) University of Iowa, Pediatrics, Iowa City, IA, USA.

A non-consanguineous Arabic family affected by a putative novel seemingly X-linked disease characterized by mental retardation and abnormal head shape was studied by gene mapping, candidate gene mutation screening and whole X chromosome Exome sequencing of a single affected member to identify the responsible gene defect. Clinical presentation includes mental retardation, hyperactivity, hypotonia, turricephaly (in one affected), dolichocephaly (in the other affected), narrow face, downward slanted palberal fissures, large ears, open mouth appearance, long and slender fingers. Neurologic and metabolic evaluations including urine organic acid, lysosomal enzyme analysis, gaunidinocompound, CDG and Fragile-X syndrome were negative. SNP genotyping with the HumanOmniExpress bead chip [Illumina, USA]; analyzed with the GeneMapper software mapped the responsible gene to four possible X-chromosome intervals [p22.33-22.2, p22.2-21.1, p31.1-q22.3, q26.3-28] as the family structure did not allow identification of a single interval with a significant LOD score. Based on the clinical presentation one candidate gene (FGF16) was screened but no pathogenic mutations were identified. Whole Exome target enrichment Next Generation Sequencing for the X chromosome was performed on the ABI SOLiD4 platform for a single affected individual. Three variants were identified within the linkage intervals: AMMECR1 [c.C208T/p.L70F] was excluded as it did not co-segregate with the disease phenotype. A second variant (c.C11T/p.P4L) in the Rab40 GTP-binding protein gene (RAB40AL). Pathogenic mutations within this gene have been associated with Martin-Probst X-linked mental retardation syndrome (MRXSMP) which presents with a different phenotype characterized by hearing loss. The last variant (c.G227A/p.C76Y) at the H2BFWT histone family gene has damaging effects according to PolyPhen and SIFT protein-modeling software, co-segregates to the disease phenotype and is absent in 752 ethnically matched control chromosomes. No known diseases have been associated with mutations in H2BFWT apart from male infertility. Tissue specific H2BFWT expression studies show expression in human fetal brain, making this mutation the likely developmental defect. At present, mutation analyses in additional healthy normal maternal male relatives are underway for the possible exclusion of H2BFWT as the offending gene.

2941F

The Solute Carrier SLC26A9 Accounts for Variability in Biomarkers of Cystic Fibrosis-Related Prenatal Exocrine Pancreatic Damage. *M.R. Miller[1], D. Soave[1,2], W. Li[1,2], T. Chiang[1], J. Gong[1], H. Levy[3,4], L. Sun[2], J.M. Rommens[1,5], F. Accurso[6,7], P. Durie[1,8], M.K. Sontag[7,9], L.J. Strug[1,2].* 1) Research Institute, The Hospital for Sick Children, Toronto, ON, Canada; 2) Biostatistics, University of Toronto, Toronto, ON, Canada; 3) Pediatrics, Medical College of Wisconsin, Milwaukee, WI; 4) Clinical Research Institute, Children's Hospital of Wisconsin, Milwaukee, WI; 5) Molecular Genetics, University of Toronto, Toronto, ON, Canada; 6) Pediatrics, University of Colorado Denver School of Medicine, Aurora, CO; 7) Pediatrics, Children's Hospital Colorado, Aurora, CO; 8) Pediatrics, University of Toronto, Toronto, ON, Canada; 9) Epidemiology, Colorado School of Public Health, Aurora, CO.

In Cystic Fibrosis (CF), exocrine pancreatic damage begins *in utero*. Serum levels of the pancreatic enzyme trypsinogen (IRT) are elevated at birth in CF individuals and form the basis of newborn screening. Postnatally, IRT levels decline rapidly, and reflect pancreatic reserve. IRT levels at birth act as a biomarker of prenatal exocrine pancreatic damage, and are heritable and variable, even in individuals with the same CFTR genotype, indicating a role for modifier genes. On average, IRT is lower in CF individuals with the intestinal obstruction Meconium Ileus (MI). In ~6300 CF individuals from the International CF Gene Modifier Consortium (the Consortium), eight SNPs in three solute carriers, *SLC26A9*, *SLC6A14* and *SLC9A3*, were associated with MI, as was a prioritized gene set corresponding to apical plasma membrane constituents that reside alongside CFTR. We hypothesized that these genetic contributors to MI would be associated with exocrine pancreatic damage, as measured by newborn screened (NBS) levels of IRT. Genome wide data and NBS IRT measures were available on 147 subjects from Colorado and Wisconsin, who were part of the Consortium, but not the MI gene modifier study. Linear regression was used to determine whether any of the eight SNPs in *SLC26A9*, *SLC6A14* and *SLC9A3* were individually associated with IRT. We then used the hypothesis-driven GWAS (GWAS-HD) method (Sun et al. 2012, *Nat Genet 44*:562-9), which uses biologic basis to prioritize genes in GWAS, to determine whether, like MI, constituents of the apical membrane were associated with IRT. SNPs in *SLC6A14* or *SLC9A3* did not provide evidence of association with IRT ($p>0.05$), but all four tested SNPs in *SLC26A9* were significantly associated with IRT after multiple test correction (SNP rs7512462, adjusted $p=3.82 \times 10^{-4}$). Each MI risk allele is associated with a ~53 ng/ml reduction in IRT (95% CI: -81, -28); this SNP explains 11% of the variation in IRT. While not significant at the genome-wide level, the GWAS-HD identified rs7512462 in *SLC26A9* as the top-ranked SNP. The GWAS-HD suggested *SLC26A9* plays the greatest role in exocrine pancreatic damage among all tested constituents of the apical membrane, demonstrating that biologically prioritizing genome-wide data can be a powerful tool in association studies. Together with the previous associations with MI, CF-related diabetes, and reported interaction with CFTR, we propose *SLC26A9* is a therapeutic target to ameliorate CF disease severity.

2942W

Unused Program Number

2943T

ALDH1A3 mutations cause recessive anophthalmia and microphthalmia. *L. FARES TAIE[1], N. CHASSAING[2,3], S. GERBER[1], G. CLAYTON-SMITH[4], S. HANEIN[1], E. SILVA[5], M. SEREY[1], V. SERRE[1,6], X. GERARD[1], C. BAUMANN[7], G. PLESSIS[8], B. DEMEER[9], L. BRETILLON[10], C. BOLE[11], P. NITSCHKE[12], A. MUNNICH[1], S. LYONNET[1], P. CALVAS[2,3], J. KAPLAN[1], N. RAGGE[13], J.M. ROZET[1].* 1) GENETICS, INSERM U781 - NECKER HOSP, PARIS 75015, France; 2) CHU Toulouse, Service de Génétique Médicale, Hôpital Purpan, 31059, Toulouse, France; 3) Université Paul-Sabatier Toulouse III, EA-4555, 31000, Toulouse, France; 4) Genetic Medicine, Manchester Academic Health Sciences Centre, St Mary's Hospital, Manchester M13 9WL, UK; 5) Department of Ophthalmology, Coimbra University Hospital, 3000-548 Coimbra, Portugal; 6) Université Paris Diderot, 75205 Paris cedex 13, France; 7) Department of Genetics, CHU Robert Debré, 75019 Paris, France; 8) Department of Medical Genetics, CHU de Caen, Hôpital de la Côte de Nacre, 14033 Caen Cedex 9, France; 9) Department of Pediatrics, CHU d'Amiens, Hôpital Nord, 80054 Amiens Cedex 1; 10) UMR1324 CSGA Centre des Sciences du Goût et de l'Alimentation, 21065 Dijon, France; 11) Genomics Plateform, IMAGINE Foundation and Paris Descartes University, 75015 Paris, France; 12) Bioinformatics Plateform, Paris Descartes University, 75015 Paris, France; 13) Wessex Clinical Genetics Service, University Hospital Southampton, Southampton, SO 16 5YA, UK; Oxford Brookes University, Oxford OX1 3QX, UK.

Anophthalmia and microphthalmia (A/M) are early eye development anomalies resulting in absent or small ocular globes, respectively. A/Ms occur as syndromic or nonsyndromic forms. They are genetically heterogenous with some mutations in some genes responsible for both anophthalmia and microphthalmia. The purpose of this study was to identify the disease gene involved in A/M in a large inbreed Pakistani Family. A combination of homozygosity mapping, exome sequencing and Sanger sequencing, was used to identify the disease mutation in the Pakistani family and to screen the ALDH1A3 gene for mutations in additional unrelated A/M patients. We identified homozygosity for a missense mutation in the gene encoding the A3 isoform of the aldehyde dehydrogenase 1 (ALDH1A3) in the Pakistani family. The screening of the gene is a cohort of A/M patients excluding known A/M genes allowed identifying two additional homozygote ALDH1A3 mutations including another missense change and a splice-site mutation in two consanguineous families. The review of the clinical files of patients showed that patients with ALDH1A3 mutations had A/M with occasional orbital cystic, neurological and cardiac anomalies. ALDH1A3 is a key enzyme in the formation of a retinoic acid gradient along the dorso-ventral axis during the early eye development. Transitory expression of mutant ALDH1A3 cDNAs showed that both missense mutations reduce the accumulation of the enzyme, potentially leading to altered retinoic acid synthesis. While the role of retinoic acid signaling in eye development is well established, our findings provide genetic evidence of a direct link between retinoic acid synthesis dysfunction and early eye development in human.

2944F

Identification of a novel locus for a recessive congenital myopathy by linkage analysis in an Israeli Bedouin family. *E. Muhammad[1,2], O. Reish[3,4], Y. Ohno[5], T. Scheetz[6,7], A. DeLuca[7], C. Searby[8], M. Regev[3], L. Benyamini[3], Y. Fellig[9], A. Kihara[5], V.C. Sheffield[8], R. Parvari[1,2], Reish, Sheffield, Parvari.* 1) Immunology and Genetics, Ben Gurion University of the Negev, Beer-Sheva , Israel; 2) National Institute of Biotechnology in the Negev, Ben Gurion University of the Negev, Beer-Sheva , Israel; 3) Genetic Institute, Assaf Hrofeh Medical Center, Zerifin 70300, Israel; 4) The Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; 5) Laboratory of Biochemistry, Faculty of Pharmaceutical Sciences, Hokkaido University, Kita 12-jo, Nishi 6-chome, Kita-ku, Sapporo 060-0812, Japan; 6) Department of Ophthalmology, University of Iowa, Iowa City, IA 52242, USA; 7) Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242, USA; 8) Department of Pathology, Hadassah-Hebrew-University-Medical-Center, Jerusalem 91120, Israel; 9) Department of Pediatrics - Division of Medical Genetics, Howard Hughes Medical Institute, University of Iowa, Iowa City, IA 52242, USA.

Congenital myopathy disorders (CMDs) are heterogeneous inherited diseases of muscle characterized by a range of distinctive histologic abnormalities. We have studied a consanguineous family with a non-progressive congenital myopathy. In order to pursue a molecular diagnosis in this family, we performed genotyping on four patients, their parents and a healthy sibling using the Affymetrix GeneChip Human SNP5 array. We determined the genotype calls by using Affymetrix GeneChip Genotyping Analysis Software (GTYPE) and KinSNP software. Based on the consanguinity in the family, we hypothesized homozygosity by descent of a recessive mutation as the likely cause of the disorder. Therefore, we searched for homozygous regions consistent with linkage. Three homozygous blocks (on chromosomes 6, 10 and 14) shared by the three affected individuals, heterozygous in the parents, and not homozygous in the unaffected sib were identified. Exome sequencing revealed a highly suggestive mutation in a gene not previously reported to be associated with myopathy in humans. The variation segregated as expected in the family, it did not appear in dbSNP, evs or the 1000 genome project, nor was it found in 134 Bedouin control individuals.

2945W

Identification of a new locus on chromosome 1 for autosomal dominant retinitis pigmentosa. *K.K. Selmer[1], R. Riise[2], M.D. Vigeland[1], K. Brandal[1], D.E. Undlien[1].* 1) Department of Medical Genetics, Oslo University Hospital and University of Oslo; 2) Department of Ophthalmology, Innland Hospital, Elverum, Norway.

Purpose: The aim of this study was to identify the genetic cause of autosomal dominant retinitis pigmentosa in a large Norwegian family.

Methods: We ascertained 20 members of a Norwegian family with autosomal dominant retinitis pigmentosa in three generations. Maximum possible LOD score was estimated in easyLINKAGE. Peripheral venous blood was drawn and clinical ophthalmological examination performed. We genotyped all 20 individuals using Affymetrix GeneChip® Human Mapping 10K 2.0 array and performed linkage analysis in the software MERLIN. Candidate genes were sequenced using traditional Sanger sequencing.

Results: Assuming full penetrance and no phenocopies, the maximum possible LOD score of this family was estimated to be 3.6. Multipoint linkage analysis revealed a single region achieving this maximum score. The linkage peak spans 5 Mb on the short arm of chromosome 1, and contains more than 100 protein coding genes, of which none are known to cause retinitis pigmentosa. Two candidate genes, the *SLC16A1* and *SLC16A4* were sequenced, but no mutation was identified.

Conclusion: We have identified a new locus for autosomal dominant retinitis pigmentosa on chromosome 1. The linked region is 5 Mb long and contain more than 100 genes. The sequencing of two candidate genes did not lead to the identification of the disease gene.

2946T

Molecular basis of previously uncharacterized mental retardation syndrome. *E. Kaasinen[1,2], E. Rahikkala[3,4], S. Miettinen[3,4], K. Palin[1,2,5], L. Pajunen[3,4], L.A. Aaltonen[3,4].* 1) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Haartman Institute, University of Helsinki, Helsinki, Finland; 3) Department of Clinical Genetics, University of Oulu, Oulu, Finland; 4) Department of Clinical Genetics, Oulu University Hospital, Oulu, Finland; 5) The Finnish Cancer Institute, Finland.

We report here a large Finnish family of nine patients with a profound mental retardation syndrome of yet unknown etiology. The family originates from the homogeneous population isolate from Northeastern Finland. Extensive genealogy work revealed that the parents were all distantly related to a common ancestor from the 17th century suggesting a recessive inheritance of a single gene mutation. The clinical features of the patients were not compatible with any previously characterized mental retardation syndrome. The clinical features included also hypotonia, strabismus, retinal pigment atrophy, difficulties to fix the eyes to an object, epilepsy, obesity, joint hypermobility and coarse facial features that develop gradually during childhood. Standard chromosome and array-CGH analysis showed normal results. Also standard metabolic screening including plasma and urine amino acids, urine organic acids and oligosaccharides were normal. We performed whole-genome SNP genotyping with Illumina's HumanOmni2.5 chip from four patients, and whole-genome sequencing with Complete Genomics from one of the four patients included in the SNP genotyping. Homozygous chromosomal regions common in the four patients were combined with variations observed in the whole-genome sequencing data. The longest shared homozygosity was found in 3p22.1-3p21.1 encompassing 11.5Mb. Sanger sequencing of genetic markers revealed that the region was homozygous in all six patients from which we had blood derived DNA available. The parents were healthy heterozygous carriers of the markers, and the homozygous region was segregating with the disease in the family. Single point parametric linkage with SimWalk2 showed logarithm of the odds (LOD) score of 11 for the locus. Four candidate genes (P4HTM, USP4, PCBP4, TKT) within the region had sequence changes that were potentially protein damaging. The changes were found with 0.3-0.7% allele frequencies in 402 whole-genome sequenced controls from Northeastern Finland, and no homozygous were found in these population-based controls, in additional in-house controls or in the 1000 Genomes data. To conclude, we have identified a new mental retardation syndrome and the underlying disease locus. Further functional studies of the candidate genes are needed to show pathogenicity of the identified sequence changes in the locus.

2947F

Novel frameshift mutation of the ADAR1 gene in a Chinese family with dyschromatosis symmetrica hereditaria. *W. Cao[1], J. Zhang[2], X. Chen[3], Z. Zhang[3], H. Wang[2], L. Guo[2], Y. Liu[2], X. Zhao[2], Q. Xing[2], M. Shao[4].* 1) Clinical Research Center, Zhengzhou People's Hospital, Zhengzhou, China; 2) Children's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai, China; 3) Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 4) Department of Urology, People's Hospital of Henan Province.

Dyschromatosis symmetrica hereditaria (DSH; MIM127400) is a human pigmentary genodermatosis which has been reported predominantly in East Asians. Patients with DSH typically present a mixture of hyperpigmented and hypopigmented macules on the dorsal aspects of the extremities and freckle-like macules on the face. The DSH locus has been mapped to chromosome 1q21.3, and pathogenic mutations were identified in the ADAR1 gene. ADAR1 catalyzes the conversion of adenosine to inosine RNA editing on double-stranded RNA substrates. Human ADAR1 protein includes two translation products due to alternative initiation of transcription: a full-length 150-kDa ADAR1p150 and a shorter 110-kDa ADAR1p110, which is initiated from a downstream AUG at codon 296. In this study, a family with autosomal dominant DSH from Jiangsu province of China was investigated. The pedigree comprised of 13 individuals spanning three generations. The proband was a 9-year-old boy who presented typical hyperpigmented and hypopigmented macules that varied in shape and size on the dorsal aspects of his hands and feet since the age of 5. Direct sequencing of the patient's ADAR1 gene revealed the presence of a novel heterozygous deletion mutation c.271-272delAG (p.R91fsX123) in exon 2. This mutation leads to a frameshift which results in a premature translation peptide consisting of 122 amino acids without translating any functional domains of ADAR1p150 located in exons 2-15. This alteration was not detected in unaffected members or controls. The novel mutation is similar to p.Q102fsX123, p.N205fsX217, p.V211fsX217 and p.H216fsX261 described previously. They all locate and create new stop codons at the upstream of codon 296, thus they cause early stopped ADAR1p150 synthesis in the patients carrying these mutations, but there should be no effect on the synthesis of ADAR1p110 proteins. Moreover, using luciferase reporter assay, for the first time we confirmed that the identified novel mutation has no effect on the synthesis of the p110 protein, although it located in the 5'-UTR of p110. Our data confirmed the dosage of functional p150 is the determinant of DSH.

2948W

Association of genetic marker rs7576183 located downstream of the human Activinβ-B gene with non-syndromic hypodontia. *L.A. Morford[1,2], P.E. DiFranco[1], A.N. Vu[1], K.C. Kirk[2], G. Falcão-Alencar[2], M.S. Gilbey[2], X. Ding[3], D.W. Fardo[3], J.K. Hartsfield, Jr.[1,2].* 1) Dept Oral Health Science, Division of Orthodontics, Univ Kentucky College of Dentistry, Lexington, KY; 2) Center for Oral Health Research, Univ Kentucky College of Dentistry, Lexington, KY; 3) Biostatistics, College of Public Health, Univ Kentucky, Lexington, KY.

Objective: The aim of this case-control study was to determine whether Single Nucleotide Polymorphisms (SNPs) located within or near the Activinβ-A and Activinβ-B genes were associated with non-3rd molar Naturally Missing Teeth (NMT). Materials and Methods: This study was conducted under the approval of the University of Kentucky Internal Review Board for Human Subjects Research. Genomic DNA was isolated from the saliva of 97 orthodontic patients (29 subjects with NMT and 68 controls). The average age ± standard deviation for the subjects with NMT and controls were 16.0 ± 5.5 and 14.5 ± 4.6 years, respectively. Clinical assessment of the dentition, dental histories and radiographic verification were utilized to accurately identify cases of tooth agenesis versus teeth that were missing due to other causes (extractions due to caries, trauma, or orthodontic treatment). Subjects with non-3rd molar NMT were subclassified based on the number of affected teeth as having hypodontia (1 to 6 NMT) or oligodontia (7 or more NMT). Peg-shaped teeth, which represent only partial tooth formation, were also noted. All control subjects exhibited a full complement of adult teeth. Cases of cleft lip and/or cleft palate and other craniofacial syndromes were excluded from the study. The genotypic variation at six SNPs located within or near the Activinβ-A or Activinβ-B genes were assayed using Taqman®-based SNP genotyping. Hardy-Weinberg Equilibrium (HWE) was assessed with the Fisher's test and logistic regression analysis was used for association with significance at $p<0.05$. Results: NMT were identified more frequently in women than men (3.8 to 1). The maxillary lateral incisors were the most common teeth affected by agenesis and/or peg formation, followed by mandibular 2nd premolars and maxillary 2nd premolars. All control genotyping data maintained HWE. The genetic marker rs7576183, located downstream of the ActivinβB gene, was significantly associated with the subjects exhibiting hypodontia ($p=0.033$) under an additive mode of inheritance (MOI). Individuals with hypodontia that were homozygous carriers of the rs7576183 G-allele predominately presented with incisor hypodontia of the maxillary arch. Conclusion: Each rs7576183 A-allele provides a multiplicative reduction of 0.41 (0.17, 0.90) in the odds of hypodontia. Conversely, each copy of the G-allele at rs7576183 confers a 2.44 (1.11, 5.88) increased odds of hypodontia.

2949T

New candidate genes in holoprosencephaly: results from homozygosity mapping in six inbred families. *S. Odent[1,2], M. de Tayrac[2,3], M. Babron[4,5], C. Dubourg[2,3], C. Mouden[3], R. Bouvet[6], S. Gazal[4,5], M. Sahbatou[7], L. Ratié[2], V. Dupé[2], J. Mosser[2,3,6], V. David[2,3].* 1) Genetique Clinique, CHU de Rennes Hosp SUD, Rennes CDX 2, France; 2) UMR 6290 CNRS, IGDR, University Rennes1, Rennes, France; 3) Service de génétique moléculaire et génomique, CHU pontchaillou, Rennes, France; 4) Inserm UMR 946, Genetic variability and human diseases, Paris, France; 5) Institut Universitaire d'Hématologie, Université Paris Diderot, Paris, France; 6) Plateforme Biogenouest, Rennes, France; 7) Fondation Jean Dausset CEPH, Paris, France.

Holoprosencephaly (HPE) is a congenital malformation of the human brain due to an imperfect division of the forebrain during early development. Multiple genetic defects have been identified as involved in this process. It is now currently admitted that HPE is a multihit pathology caused by at least two or more dysfunctional events involving at least 4 major genes (SHH, ZIC2, SIX3 and TGIF) and 10 minor genes belonging to different signaling pathways. However, the mutations and deletions in these genes represent only 30% of HPE cases. Recessive inheritance of HPE can also be suspected in consanguineous families with intrafamilial recurrence. Homozygosity mapping was undertaken in six families with history of consanguineous marriage to search for regions harboring mutations that are identical by descent. Parents and affected children were genotyped on HumanCytoSNP-12 arrays (Illumina). We first determined the population ancestry of each family with OriginMineR [de Tayrac, ASHG 2012] to estimate their specific SNP allele frequencies. Inbreeding coefficients of affected children were estimated from their genomic data by the FEstim method [Leutenegger et al., 2003]. Using the genomic inbreeding coefficient we performed homozygosity mapping without relying on the genealogical information. In parallel we detected the runs of homozygosity across the genome for each individual (PLINK). Three regions of interest were detected by both methods on chromosomes 1, 6 and 10. We applied Endeavor software to prioritize genes in these regions. One of them was previously identified by CGH array on chromosome 6 (DLL1) and other good candidate genes will be presented.

2950F

A frameshift mutation of TBC1D7, a subunit of the TSC1-TSC2 complex upstream of mTORC1, causes a new distinct clinical phenotype with intellectual disabilities. *L. Micale[1], A. Abdullah Alfaiz[2,3], B. Mandriani[1], C. Fusco[1], B. Augello[1], M.T. Pellico[1], J. Chrast[2], L. Zelante[1], A. Reymond[2], G. Merla[1].* 1) IRCCS, Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, FG, Italy; 2) Center for Integrative Genomics, Génopode, Lausanne, Switzerland; 3) Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland.

Mutations in TSC1 or TSC2 cause the tuberous sclerosis complex (TSC), a multisystemic disorder characterized by the development of hamartomas or benign tumors in various organs as well as a high incidence of epilepsy, intellectual disability, and autism. TSC1, TSC2, and TBC1D7 form a complex that inhibits mTORC1 signalling and limits cell growth. Disruption of TBC1D7 has been reported in a family with intellectual disabilities and megalencephaly. Using exome sequencing we identified two sisters with a novel homozygous truncating mutation in TBC1D7: c.18-21delGAGA; p.R7TfsX21. Both siblings are macrocephalic and affected by medium intellectual disabilities. They present myopia, astigmatism, prognathism, osteo-articular defects, behavioural abnormalities, learning difficulties and celiac disease. We assessed the functional consequences of this variant on TBC1D7 mRNA and protein levels in cell lines derived from the affected subjects and their heterozygote unaffected parents. It affects mTORC1 activity, measured by phosphorylation assays of its S6K1 and 4EBP1 substrates and monitoring cellular cell growth and autophagy markers as suggested by the newly described role of TBC1D7 as an upstream regulator of mTORC1. This study reinforces the involvement of TBC1D7 in the regulation of mTORC1 pathway, increases the number and types of TBC1D7 mutations, enlarges the associated phenotypic spectrum, and provide additional proofs about the existence of a new syndrome due to TBC1D7 mutations.

2951W

Exome Sequencing Identifies Novel Genes for Autosomal Recessive Heterotaxy-spectrum Disorders. *A.E. Cast[1], M. Tariq[1], F. Rahim[2,3], M. Jelani[2,4], S.M. Ware[1,5].* 1) Division of Molecular Cardiovascular Biology, Cincinnati children's Hospital Medical center, Cincinnati, OH; 2) Institute of Basic Medical Sciences, Khyber Medical University, Peshawar 25000, Pakistan; 3) Physiology Department Bacha Khan Medical College, Mardan, Pakistan; 4) Princess Al-Jawhara Albrahim Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah 80205, Saudi Arabia; 5) Division of Human Genetics, Cincinnati children's Hospital Medical center, Cincinnati, OH.

Heterotaxy is a rare multiple congenital anomaly syndrome resulting from abnormal left-right (LR) embryonic organ patterning with a prevalence of ~1 in 10,000 newborns. Although heterotaxy is the most highly heritable cardiovascular malformation, the genetic basis is most frequently multifactorial and autosomal recessive inheritance has been described only rarely. In this study, we analyzed whole-exomes of 18 probands with heterotaxy-spectrum disorders. This includes 4 probands belonging to families with obvious autosomal recessive inheritance. Exome analyses identified approximately ~8000 coding variants in each individual which were subsequently filtered using 1000 Genome, dbSNP135 and NIH exome sequencing project (ESP) databases. The remaining variants were further analyzed using regions of homozygosity (ROH) revealed by SNP genotyping and/or inheritance-based exome analysis. This identified suspected disease-causing variants in novel candidate genes in 7 probands as well as in previously known heterotaxy or related phenotype genes in 4 probands. The majority of variants were homozygous and located in ROH (~2Mb-44Mb), supporting the autosomal recessive inheritance in 6 familial cases. Sanger sequencing of probands and available family members confirmed these homozygous mutations and validated their recessive segregation with phenotype. Novel selected candidates (*C21orf59, DNAH6, DNAJB7, MIA3, INTS7, JMJD1C, SORBS2*) were further investigated by expression analyses in mouse, including RT-PCR and whole mount in-situ hybridization, with results suggesting a possible role in LR patterning and early heart development. Knockdown of *c21orf59* via antisense morpholino in *Xenopus laevis* resulted in LR patterning defects and abnormal *pitx2* expression at stage 30, suggesting *c21orf59* is a novel gene involved in laterality. This study proves the importance of exome analysis complemented with ROH segregation revealed by SNP arrays for identification of unknown causes of human laterality defects and these causes will provide insights to understand new molecular mechanisms involved in embryonic LR patterning.

2952T

Age-Dependent Germline Mosaicism of the Most Common Noonan Syndrome Mutation Shows the Signature of Germline Selection. *P. Calabrese[1], S. Yoon[1], S. Choi[1], J. Eboreime[1], B. Gelb[2], N. Arnheim[1].* 1) Mol Comp Biol, Univ Southern California, Los Angeles, CA; 2) Icahn School of Medicine at Mount Sinai, New York, NY.

Noonan syndrome (NS) is among the most common Mendelian genetic diseases (~1/2,000 live births). Most cases (50%-84%) are sporadic, and new mutations are virtually always paternally derived. More than 47 different sites of NS de novo missense mutations are known in the PTPN11 gene that codes for the protein tyrosine phosphatase SHP-2. Surprisingly, many of these mutations are recurrent with nucleotide substitution rates substantially greater than the genome average; the most common mutation, c.922A>G, is at least 2,400 times greater. We examined the spatial distribution of the c.922A>G mutation in testes from 15 unaffected men and found that the mutations were not uniformly distributed across each testis as would be expected for a mutation hot spot but were highly clustered and showed an age-dependent germline mosaicism. Computational modeling that used different stem cell division schemes confirmed that the data were inconsistent with hypermutation, but consistent with germline selection: mutated spermatogonial stem cells gained an advantage that allowed them to increase in frequency. SHP-2 interacts with the transcriptional activator STAT3. Given STAT3's function in mouse spermatogonial stem cells, we suggest that this interaction might explain the mutant's selective advantage by means of repression of stem cell differentiation signals. Repression of STAT3 activity by cyclin D1 might also play a previously unrecognized role in providing a germline-selective advantage to spermatogonia for the recurrent mutations in the receptor tyrosine kinases that cause Apert syndrome and MEN2B. Looking at recurrent mutations driven by germline selection in different gene families can help highlight common causal signaling pathways.

2953F

Clinical spectrum of the families with multiple synostosis syndromes with novel NOG mutations. *J. Kim[1], B. Lee[1,2], G. Kim[1], J. Lee[1], S. Choi[1], J. Lee[1], H. Yoo[1,2].* 1) Medical Genetics Center, Asan Medical Center, Seoul, South Korea; 2) Dept. Pedicatrics, Asan Medical Center, Children's Hospital, Univ. Ulsan College of Medicine, Seoul, South Korea.

Multiple synostosis syndrome (SYNS) is a rare autosomal dominant skeletal disorder, characterized by variable clinical features including facial dysmorphism, progressive fusion of proximal interphalangeal joints, variable degree of fusions of multiple other joints including spine, and conductive hearing loss. Currently, three genes, NOG, GDF5, and FGF9, have been identified as the causative genes of SYNS. According to the haploinsufficiency of the NOG, GDF5, and FGF9 genes, SYNS is classified as SYNS1 (MIM186500), SYNS2 (MIM610017) and SYNS3 (MIM612961), respectively. However, due to its diverse phenotypic and genotypic heterogeneities as well as its extreme rarity, it is difficult to diagnose a patient with SYNS either clinically or genetically. In this report, two unrelated Korean families with SYNS were described. Under the diagnosis of SYNS, genetic testing was done for the NOG, GDF5, and FGF9 genes using genomic DNA from peripheral leukocytes of the patients. Genetic testing revealed two novel NOG mutations: c.452C>A (p.Ser151*) in family 1 and c.261_262insG (p.Pro88Alafs*94) in family 2. The two novel mutations are expected to produce the premature truncated proteins, lacking C-terminal region that is an integral part for its appropriate function. The patients shared the general features of SYNS, but their phenotypic expressivities were different among inter- and intra-family members. In addition, their diversities were irrespective of a patient's age, indicating the importance of surveillance for the full spectrum of SYNS for each affected patient. Our report expands the understanding of this rare condition from the genetic and clinical perspectives.

2954W

***ZFPM2* Mutations in Sporadic and Familial Congenital Diaphragmatic Hernia Patients.** *M. Longoni[1,2], M.K. Russell[1,2], F.A. High[1,2,3], K. Lage[1,2,4], J. Wells[5], C.J. Bult[5], K.G. Ackerman[6], C. Lee[2,7], B.R. Pober[1,2,3], P.K. Donahoe[1,2,4].* 1) Massachusetts General Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Children's Hospital Boston, Boston, MA; 4) Broad Institute, Cambridge, MA; 5) The Jackson Laboratory, Bar Harbor, ME; 6) School of Medicine and Dentistry, University of Rochester, Rochester, NY; 7) Brigham and Women's Hospital, Boston, MA.

Congenital Diaphragmatic Hernia (CDH) is a common birth malformation. CDH is genetically heterogeneous, and evidence points to a multifactorial, possibly polygenic, etiology in many cases. The most prevalent phenotype is isolated left-sided posterolateral hernia, associated with life threatening pulmonary hypoplasia. An intragenic deletion in Zinc finger protein, multitype 2 (*ZFPM2*, also known as *FOG2*), a gene previously implicated in CDH, was identified in a multigenerational family with isolated hernia inherited in an autosomal dominant pattern with incomplete penetrance. Point mutations in *ZFPM2* or in its transcriptional co-repressor C-terminal binding protein (*CTBP2*) were identified in 7 unrelated patients from a cohort of 93 CDH patients by exome sequencing. All tested variants were inherited from an unaffected parent, consistent with the hypothesis that multiple environmental factors or genetic modifiers are necessary for *ZFPM2* haploinsufficiency to display a clinically recognizable diaphragmatic defect. In an effort to identify pathogenic mechanisms of *ZFPM2* mutations resulting in CDH, microarray expression data from the primordial diaphragm of *Zfpm2* null mice at E11.5 were generated from laser captured tissue. The dataset, containing 428 differentially expressed genes (mutant vs. wildtype), was enriched for Mouse Genome Informatics (MGI) Mammalian Phenotype ontology categories such as skeleton, muscle, and respiratory phenotypes. This expression dataset, together with an analogous one from wild type mice collected at three time points during diaphragm development published by Russell and colleagues (2011), was used as a screening tool to identify possible second hits that explain the observed variable penetrance. In most patients with *ZFPM2* deletions or point mutations, we identified additional novel or rare sequence variants, defined as less than 0.1% in control cohorts, mapped to genes deregulated in *Zfpm2* knock-out animals or in genes whose expression levels are highly correlated with *Zfpm2* during ontogenesis. Though further studies are needed to determine that our prioritized second hits are bona fide *Zfpm2* genetic interactors, our results show that developmentally oriented expression studies are a useful tool for the interpretation of exome sequencing data when studying birth defects.

2955T

Whole exome sequencing links TMCO1 with Cerebro-Facio-Thoracic Syndrome. *Y. Bayram[1], E. Karaca[1], D. Pehlivan[1], C.R. Beck[1], C. Gonzaga-Jauregui[1], T. Gambin[1], S.N. Jhangiani[2], H. Aydin[3], W. Wiszniewski[1], A.H. Cebi[4], M.M. Atik[1], D. Muzny[2], R.A. Gibbs[2], J.R. Lupski[1,5,6].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 3) Department of Medical Genetics, Zeynep Kamil Women's and Children's Hospital, Istanbul, Turkey; 4) Department of Medical Genetics, Karadeniz Technical University, Faculty of Medicine, Trabzon, Turkey; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Texas Children's Hospital, Houston, TX, USA.

Next generation massively parallel sequencing, both Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES), is a disruptive technology that has changed the course of clinical/human genetics. As a more efficient and cost effective method, WES is now widely used as a diagnostic tool for identifying the underlying gene, mutation(s), and genetics of challenging genetic syndromes. Here we report a case with a clinical diagnosis of Cerebrofaciothoracic syndrome (CFTS) (MIM#213980) in whom we identified by WES homozygous splice site mutations in the transmembrane and coiled-coil domains 1 (TMCO1) gene. Mutations in TMCO1 gene were shown to be responsible for craniofacial dysmorphism, skeletal anomalies, and mental retardation syndrome (MIM#614132), which is characterized by facial dysmorphism, multiple malformations of the vertebrae and ribs, and mental retardation. Our retrospective review revealed that clinical manifestations of both syndromes are remarkably similar. We propose that, mutations of TMCO1 are not only responsible for craniofacial dysmorphism, skeletal anomalies, and mental retardation syndrome, but also for CFTS. In addition, one should keep in mind that these two syndromes might be different spectrums of the same clinical entity, molecularly diagnosed as TMCO1 defect syndrome.

2956F

Stormorken syndrome: mutation in STIM1 as a cause or a remarkable coincidence? *G. Morin[1], G. Jedraszak[1], A. Rabbind singh[1], B. Roméo[2], E. Bourges-Petit[3], H. Sevestre[4], D. Brémond-Gignac[5], H. Ouadid-Ahidouch[6], M. Mathieu[1], J. Rochette[1].* 1) Genetic department, EA 4666, Amiens University Hospital, Amiens, France; 2) Pediatric pneumology unit, Amiens University Hospital, Amiens, France; 3) Pediatric cardiology unit, Amiens University Hospital, Amiens, France; 4) Pathology service, Amiens University Hospital, Amiens, France; 5) Ophtalmology service, Amiens University Hospital, Amiens, France; 6) University of Picardie Jules Verne, UFR of Sciences, Laboratory of Cellular and Molecular Physiology EA 4667, SFR CAP-SANTE, Amiens, France.

Background Stormorken syndrome is very rare autosomal dominant disease with associated congenital miosis, thrombocytopenia/thrombocytopathia, tubular aggregate myopathy, asplenia, small stature, ichtyosis, and moderate developmental disabilities. When documented the myopathy is characterized by the presence of tubular aggregates. Additionally, most patients present moderate hypocalcemia. Till now, only 4 families have been reported with a total of 6 patients. The moderate hypocalcemia observed in some patients suggested the deregulation of the calcium homeostasis. Furthermore, mutations in the calcium sensor region of STIM1 were reported in the autosomal dominant form of tubular aggregate myopathy. Methods We describe a family study expending over 2 generations with 2 affected individuals displaying Stormorken syndrome. The disease affected father and son confirming its autosomal dominant inheritance. Both proband and father had marked congenital miosis, moderate thrombocytopenia (lower value: son 28G/L, father 97G/L), muscular fatigue, significant elevation of creatine kinase (higher value: son 769UI/L, father 4412UI/L), moderate hypocalcemia, small stature, moderate learning difficulties and ichtyosis. The father was asplenic. The son presented an increased bleeding time with normal values for clotting factors. In order to investigate the genetic cause of the Sormorken syndrome, we investigated STIM1 based on the presence of hypocalcemia and the result of the muscular biopsy. The STIM1 gene was studied by direct sequencing in the affected and unaffected individuals of the family. Results We identified a novel reported heterozygous missense mutation in exon 7 of STIM1 gene (c.910C>T; p.Arg304Trp) in the father and son, and not in the other members of the family. Restriction enzyme analysis confirmed the presence of the C>T substitution in position 910. The identified mutation lies in the coiled coil 1 domain of the protein (Cα2). This amino acid residue is highly conserved in mammals. Data obtained from in silico analysis show a major change in the secondary structure of the Cα2 domain when compared to the wild type.

2957W

SX10 in the development of the olfactory ensheathing cells. *V. Pingault[1,2,3], W. Watanabe[1,2], S. Marcos[4], V. Baral[1,2], A. Chaoui[1,2], M. Goossens[1,2], J.P. Hardelin[5], C. Dodé[4,6], N. Bondurand[1,2].* 1) Genetics, INSERM U955 hopital Henri Mondor, Creteil, France; 2) Université Paris Est, Faculté de Médecine, Créteil, F-94010, France; 3) Hôpital Henri Mondor, Laboratoire de Biochimie et Génétique, Créteil, F-94010, France; 4) INSERM U1016, Institut Cochin, Département de génétique et développement, Université Paris-Descartes, Paris, F-75014, France; 5) INSERM U587, Département de neuroscience, Institut Pasteur, Université Pierre et Marie Curie, Paris, F-75015, France; 6) Laboratoire de biochimie et génétique moléculaire, APHP, Hôpital Cochin, Paris, F-75014, France.

The SOX10 transcription factor plays a role in the maintenance of progenitor cell multipotency, lineage specification, cell differentiation, and is a major actor in the development of the neural crest. It has been implicated in Waardenburg syndrome (WS), a rare disorder characterized by the association of pigmentation abnormalities and deafness, sometimes associated with neurological disorders. We recently found that apart from WS, SOX10 mutations also cause Kallmann syndrome (KS) with deafness. KS is defined by the association of anosmia and hypogonadotropic hypogonadism due to incomplete migration of neuroendocrine GnRH (gonadotropin-releasing hormone)-cells along the olfactory, vomeronasal, and terminal nerves. We provide further study of SOX10 expression in the olfactory ensheathing cells (OECs) during mouse development and of the developmental defects observed in the *So×10^{lacZ}* mutant mouse. This confirms several of the roles previously suspected for these peculiar cells, which are thought to be responsible for the olfactory neurons unique capacity to renew throughout life.

2958T

FGFR1 mutations cause Hartsfield syndrome, the unique association of holoprosencephaly and ectrodactyly. *C. Vilain[1,2], I. Migeotte[3,1], N. Lambert[3,1], C. Perazzolo[3], D. de Silva[4], B. Dimitrov[5], C. Heinrichs[2], S. Janssens[6], B. Kerr[7], G. Mortier[8], G. Van Vliet[9], P. Lepage[2], G. Casimir[2], M. Abramowicz[1,3], N. Simonis[10], G. Smits[1,2].* 1) Dep of Clinical Genetics, ULB Center of Human Genetics, Brussels, Belgium; 2) Department of Paediatrics, Hôpital Universitaire des Enfants Reine Fabiola (HUDERF), Université Libre de Bruxelles (ULB), Brussels, Belgium; 3) Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire (IRIBHM), Université Libre de Bruxelles (ULB), Brussels, Belgium; 4) Department of Physiology, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka; 5) Department of Clinical Genetics, Guy's Hospital, London, United Kingdom; 6) Center for Medical Genetics Ghent, Ghent University Hospital, Ghent, Belgium; 7) Manchester Academic Health Science Centre, University of Manchester, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom; 8) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium; 9) Endocrinology Service and Research Center, Hôpital Sainte-Justine and Department of Pediatrics, Université de Montréal, Montréal, Québec, Canada; 10) Laboratoire de Bioinformatique des Génomes et des Réseaux (BiGRe), Université Libre de Bruxelles (ULB), Brussels, Belgium.

Harstfield syndrome is the rare and unique association of holoprosencephaly (HPE) and ectrodactyly, with or without cleft lip and palate, and variable additional features. All the reported cases occurred sporadically. Although several causal genes of HPE and ectrodactyly have been identified, the genetic cause of Hartsfield syndrome remains unknown. We hypothesized that a single key developmental gene may underlie the co-occurrence of HPE and ectrodactyly. Using whole exome sequencing in four isolated cases including one case-parents trio, and direct Sanger sequencing of three additional cases, we identified a novel FGFR1 mutation in six out of 7 patients with Hartsfield syndrome. Affected residues are highly conserved and are located in the extracellular binding domain of the receptor (two homozygous mutations) or the intracellular tyrosine kinase domain (four heterozygous de novo variants). Strikingly, among the six novel mutations, three are located in close proximity to the ATP's phosphates or the coordinating magnesium, with one position required for kinase activity, and three are adjacent to known mutations involved in Kallmann Syndrome plus other developmental anomalies. Dominant or recessive FGFR1 mutations are responsible for Hartsfield syndrome, consistently with the known roles of FGFR1 in vertebrate ontogeny and conditional Fgfr1-deficient mice. Our study shows that, in humans, lack of accurate FGFR1 activation can disrupt both brain and hand/foot midline development, and that FGFR1 loss-of-function mutations are responsible for a wider spectrum of clinical anomalies than previously thought, ranging in severity from seemingly isolated hypogonadotropic hypogonadism, through Kallmann Syndrome with or without additional features, to Hartsfield Syndrome at its most severe end.

2959F

Exome sequencing in Mendelian cleft lip and palate families: Results of an International Orofacial clefting consortium. *T. Roscioli[1], M. Buckley[2], C. Carels[3], T. Cox[4], E. Haan[5], D. Hanna[6], A.V. Hing[7], K. Khandelwal[3], E. Kirk[8], A. Lidral[9], J. Murray[10], D.A. Nickerson[6], F. Reinier[6], J. Smith[6], E. Thompson[5], H. van Bokhoven[11], H. Zhou[12].* 1) School of Women's and Children's Health, University of New South Wales, Sydney, NSW, Australia; 2) Department of Haematology and Genetics, South Eastern Area Laboratory Services, Sydney, Australia; 3) Department of Orthodontics and Craniofacial Biology, Radboud University, Nijmegen, The Netherlands; 4) Department of Pediatrics, University of Washington, Seattle, USA; 5) South Australian Clinical Genetics Service, Adelaide, Australia; 6) The University of Washington Center for Mendelian Genomics, Seattle, USA; 7) Department of Pediatrics, Division of Craniofacial Medicine, University of Washington and Children's Hospital and Regional Medical Center, Seattle, USA; 8) Department of Medical Genetics, Sydney Children's Hospital, Sydney, Australia; 9) College of Dentistry, University of Iowa, Iowa City, USA; 10) Department of Pediatrics, University of Iowa, Iowa City, USA; 11) Department of Human Genetics, Radboud University, Nijmegen, The Netherlands; 12) Department of Developmental Biology, Radboud University, Nijmegen, The Netherlands.

Cleft lip and palate is one of the most common birth defects, occurring in 1 of 700 live births. The majority of individuals occur in single affected member families. While GWAS studies have identified major loci, there remains an unexplained genetic contribution that we hypothesize is due to rare unidentified mutations. To this end, we have undertaken genomic studies in larger multi-affected families who could be enriched for Mendelian forms of orofacial clefting in order to identify causative high penetrance alleles. A minimum of two exomes have been performed from each family on the Illumina platform in 20 families with patterns of inheritance consistent with dominant, recessive and X-linked traits. The results have identified pathogenic mutations in some known genes, amongst them, ARHGAP29, in approximately 15% of the cohort. The remaining families may harbour mutations in novel genes for which investigations are ongoing, including studies of the collagen family of genes and SOX9. We discuss in addition adjunct methodologies to reduce the number of variants requiring confirmatory Sanger sequencing such as the programs homozygositymapper, dominantmapper and SNP arrays. In conclusion, based on this initial survey, we have identified multiple rare variants which may contribute to isolated cleft lip and palate.

2960W

Identification of a novel mutation Arg118Gly in the TWIST1 gene causing Saethre-Chotzen Syndrome. *M.L.M. CASTRO[1], C.H.P. GRANGEIRO[1], N.R. QUARESEMIN[1], C.M. LEPREVOST[1], J.A. JOSAH-KIAN[1], L.A.F. LAUREANO[2], W.A.R. BARATELLA[1], J. HUBER[1], E.S. RAMOS[1,3], W.A. SILVA JR[3,4], L. MARTELLI[1,3], G.A. MOLFETTA[3,4].* 1) Serviço de Genética Médica - Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 2) Laboratório de Citogenética - Hospital Das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 3) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 4) Centro de Medicina Genomica - HCFMRP - USP.

Saethre-Chotzen syndrome (SCS) or Acrocephalosyndactyly Type III is a craniosynostosis syndrome characterized by craniofacial and limb abnormalities, complete penetrance with variable expressivity and autosomal dominant inheritance. Its prevalence varies between 1:25,000 and 1:50,000 live births. Since the first description of mutations in the TWIST1 gene causing SCS, about 163 different mutations, deletions and insertions have been detected in affected patients. The TWIST1 gene is located on chromosome 7p21.1 and encodes a transcription factor with a DNA-binding domain implicated in cell lineage determination and differentiation. We report on a novel missense mutation in the TWIST1 gene in a patient with SCS phenotype. The patient was referred to the Medical Genetics Division of the University Hospital when he was 7 months old and his dysmorphological examination showed: short stature, microbrachycephaly due to bilateral coronal synostosis, high forehead, low frontal hairline, flat face, midface hypoplasia, prominent eyes, unilateral eyelid ptosis, a broad and depressed nasal bridge, anteverted nostrils, high vaulted and narrow palate, antihelix prominent and prominent crura helicis, short neck, small hands with brachydactyly, clinodactyly of fifth fingers bilaterally. The development was normal, however, at the age of two years old, in the last evaluation during his follow up at the Genetics Division, he demonstrated sensorineural hearing loss and speech delay. Genomic DNA for the genetic analysis was extracted from peripheral venous blood sample by conventional methods. In order to detect any abnormalities in the TWIST1 gene, we carried out the DNA sequencing method using an ABI Big Dye Terminator Kit (Applied Biosystems, USA), following manufacturer's instructions. The sequences were compared to the TWIST1 Accession Number NM_000474 and it was revealed the mutation Arg118Gly in heterozygosis. In silico bioinformatics analysis using the software MutPred indicated a high score of patogenicity. The MutPred data have also revealed that the Arg118Gly mutation causes loss of MoRF binging, loss of stability, loss of phosphorilation at T121, loss of disorder and loss of solvent accessibility. We suggest that this novel mutation may be responsible for the SCS phenotype, but further complementary functional assays should be applied for confirming the patogenicity of the mutation Arg118Gly in the TWIST1 gene.

2961T

Splice site mutation leads to ataxia and retardation in a consanguineous family. *R.M. Burns[1], W. Peng[2], J. Xu[2], J.Z. Li[2], M. Burmeister[1, 2].* 1) Molecular and Behavioral Neuroscience Institute University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics University of Michigan, Ann Arbor, MI.

The cerebellar ataxias are a heterogeneous group of neurological disorders with onset of symptoms occurring between early childhood and late adulthood. While the genetic causes of several ataxias have already been identified, many more, especially rare recessive forms, are still unknown. Heterogeneity, or the large number of genes that can cause recessive ataxias, is the major reason why many genes for recessive forms are still unknown. Individual families with independent mutations are too small to positively identify a chromosomal region by genetic linkage.

Objective: We are studying a consanguineous Turkish family with a non-progressive, congenital ataxia of unknown etiology. Symptoms include hypotonia, developmental delay, mental retardation, nystagmus, truncal and extremity ataxia and cerebellar hypoplasia that do not progress with age. In this study, we aim to elucidate the genetic cause of a rare recessive ataxia.

Methods: We utilized a combination of exome sequencing with homozygosity mapping and expression analysis to identify candidate genes. Sanger sequencing was used to verify damaging variants. Molecular assays, such as RT-PCR and immunoblotting, were used to determine the functional consequences of gene variants.

Results: We have identified a variant in an obligatory splice sequence (the first base of an intron is changed from GT to AT). This mutation is absent in >13,000 American control DNAs and 200 Turkish control chromosomes. RT-PCR demonstrates that this splice mutation causes skipping of an exon that is present in all isoforms and expression of the gene is lower in patients than controls. Immunoblot demonstrates loss of this protein in the affected individuals.

Conclusion: Our results suggest that we have identified a novel gene causing recessive ataxia. Additionally, we are developing a zebrafish animal model, utilizing morpholino-mediated knockdown, to study this disorder. Our research may help identify causes of other ataxias and may lead to novel therapies to treat ataxias.

2962F

A novel gene for Spinocerebellar Ataxia (SCA) linked to chromosome 6 and involved in fatty acid metabolism. *E. Di Gregorio[1,2], B. Borroni[3], E. Giorgio[1], D. Lacerenza[1], C. Mancini[1], A. Calcia[1], I. Mura[4], D. Coviello[4], N. Mitro[5], M. Gaussen[6], N. lo Buono[1], A. Funaro[1], G. Vaula[7], I. Lagroua[6], L. Orsi[7], A. Durr[6,8], C. Costanzi[3], A. Padovani[3], A. Brice[6,8], L. Boccone[9], E. Hoxha[10], F. Tempia[10], D. Caruso[5], G. Stevanin[6,8], A. Brusco[1,2].* 1) Medical Sciences, University of Torino, Torino, to, Italy; 2) S.C.D.U Medical Genetics, Città della Salute e della Scienza, Torino, Italy; 3) University of Brescia, Department of Neurology, Brescia, Italy; 4) Laboratory of Human Genetics, Galliera Hospital, Genova, Italy; 5) University of Milano, Department of Pharmacological Sciences, Italy; 6) Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (INSERM/UPMC Univ. Paris 6, UMR_ S975; CNRS 7225, EPHE), Pitié-Salpetriere Hospital, Paris, France; 7) S.C.D.U. Neurology, Città della Salute e della Scienza, Torino, Italy; 8) APHP, Fédération de génétique, Pitié-Salpêtrière Hospital, Paris, France; 9) Ospedale Regionale Microcitemie, ASL 8, Cagliari, Italy; 10) University of Torino, Neuroscience Institute Cavalieri Ottolenghi (NICO), Torino, Italy.

Spinocerebellar ataxias (SCA) are a highly heterogeneous group of autosomal dominant neurodegenerative disorders phenotypically characterized by gait ataxia, incoordination of eye movements, speech, and hand movements, and usually associated with cerebellar atrophy. More than 30 SCAs have been identified, whose genes may be classified into two main categories: repeat expansion disorders - among which the most common forms SCA 1-3, 6, and 7 - and genes with conventional mutations. Here we report the identification of a novel SCA gene. Genome-wide linkage analysis identified a 92 Mb region on chromosome 6 in an Italian family affected by a pure form of ataxia with disease onset in the 4th decade of life. Next generation sequencing of all coding genes in the smallest interval identified only one possible mutated gene, with a Gly to Val amino acid change. Screening of over 450 SCA independent patients identified this same mutation in two further unrelated Italian families. Haplotyping proved that at least two of the three families shared a common ancestor. Two further missense variants, in three independent families, affected the same exon involved in the Italian families, suggesting this may be a mutational hot-spot. All changes hit conserved amino acids, and were not common polymorphisms. The gene encodes an ubiquitously expressed enzyme involved in fatty acid biosynthesis. In situ hybridization on mouse and human brain demonstrated that Purkinje cells have a peculiar high expression of the enzyme. In agreement with the function of this gene, we showed a reduced level of a subgroup of fatty acids in the serum of three patients with the Gly to Val amino acid change. Furthermore, we found an increased gene expression at the messenger RNA and protein levels in the patients' lymphoblasts vs. controls. We hypothesize that a positive feedback loop, activated by the functional impairment of the enzyme, may lead to an increase of protein expression that in turn may accumulate within cells leading to a toxic gain of function. In conclusion, we suggest that our mutated gene, highly expressed in Purkinje cells, is a good candidate for a new form of pure autosomal dominant cerebellar ataxia and join the group of CNS diseases involving fatty acids metabolism.

2963W

Targeted exome sequencing with copy number variant detection reveals HSD17B4-deficiency in a male with cerebellar ataxia and azoospermia. *D.S. Lieber[1,2,3,4], S.G. Hershman[1,2,3,4], N.G. Slate[1,2], S.E. Calvo[1,2,3,4], K.B. Sims[2,5], J.D. Schmahmann[5], V.K. Mootha[1,2,3,4,6].* 1) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 3) Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA; 4) Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA; 5) Department of Neurology, Massachusetts General Hospital & Harvard Medical School, Boston MA 02114, USA; 6) Department of Medicine, Massachusetts General Hospital, Boston MA 02114, USA.

We describe an adult male with cerebellar ataxia, peripheral neuropathy, hearing loss, and azoospermia. Commercial genetic testing of 18 ataxia and mitochondrial disease genes was negative, but biochemical findings in serum, urine, and muscle biopsy pointed to the possibility of a mitochondrial abnormality. Targeted exome sequencing followed by analysis of single nucleotide variants and small insertions/deletions failed to reveal a genetic basis of disease. We subsequently applied computational algorithms to infer copy number variants (CNVs) from exome data, revealing a heterozygous 12kb deletion of exons 10-13 of *HSD17B4* that was compounded with a rare missense variant (p.A196V). Recessive mutations in *HSD17B4* cause D-bifunctional protein deficiency, a severe, infantile-onset disorder of peroxisomal fatty acid oxidation. Recently, compound heterozygous mutations in *HSD17B4* were reported in two sisters diagnosed with Perrault syndrome (MIM # 233400), who presented with ovarian dysgenesis, hearing loss, and ataxia. Retrospective review of patient records revealed elevated ratios of pristanic:phytanic acid and arachidonic:docosahexaenoic acid, consistent with peroxisomal dysfunction. Our case expands the phenotypic spectrum of Perrault syndrome, representing the first male reported with ataxia, infertility, and HSD17B4 deficiency. Furthermore, our study highlights the importance of exome-based CNV detection in the diagnosis of rare disorders and points to potential crosstalk between mitochondria and peroxisomes in Perrault syndrome.

2964T

Mutations of COQ2 in Familial and Sporadic Multiple System Atrophy. *J. Mitsui[1], T. Matsukawa[1], H. Ishiura[1], Y. Fukuda[1], Y. Ichikawa[1], H. Date[1], B. Ahsan[1], Y. Nakahara[1], Y. Momose[1], Y. Takahashi[1], J. Goto[1], Y. Yamamoto[2], K. Shirahige[3], H. Takahashi[4], O. Onodera[5], M. Nishizawa[5], T. Kondo[6], S. Murayama[7], A. Dürr[8], A. Brice[8], A. Filla[9], T. Klockgether[10], U. Wüllner[10], G. Nicholson[11], S. Gilman[12], S. Tsuji[1], JAMSAC, JGSCAD, JPDSGC, JaCALS, and NAMSA-SG.* 1) Department of Neurology, The University of Tokyo, Tokyo Japan; 2) Tokyo University of Technology, Tokyo, Japan; 3) Institute of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo Japan; 4) Department of Pathology, Brain Research Institute, Niigata University, Niigata, Japan; 5) Department of Neurology, Brain Research Institute, Niigata University, Niigata, Japan; 6) Department of Neurology, Wakayama Medical University, Wakayama, Japan; 7) Department of Neuropathology and the Brain Bank for Aging Research, Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan; 8) Hôpital de la Salpêtrière, Département de Génétique et Cytogénétique, F-75013, Paris, France; 9) Department of Neurological Sciences, University Federico II, Napoli, Italy; 10) Department of Neurology, University of Bonn and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 11) University of Sydney at the ANZAC Research Institute, Concord Hospital, Sydney, Australia; 12) Department of Neurology, University of Michigan, Ann Arbor, MI.

Background Multiple system atrophy (MSA) is an intractable neurodegenerative disease characterized by autonomic failure with various combinations of parkinsonism, cerebellar ataxia, and pyramidal dysfunction. MSA is classified into two subtypes: MSA-C, characterized by predominant cerebellar ataxia, and MSA-P, characterized by predominant parkinsonism. MSA-C has been reported to be more prevalent (8.0 per 100,000) than MSA-P in the Japanese population, whereas MSA-P has been reported to be more prevalent in Europe and North America. Furthermore, multiplex MSA families have been identified. These findings strongly suggest involvement of genetic components underlying MSA. Methods In combination with linkage analysis, we performed whole-genome sequencing of an individual with autopsy-proven MSA of a multiplex family. Mutational analysis was further conducted on the patients in five other families, and, furthermore, on Japanese (363 MSA patients and 520 controls), European (223 MSA patients and 315 controls), and North American (172 MSA patients and 294 controls) series. Functional analysis of COQ2, which encodes para-hydroxybenzoate-polyprenyltransferase, an enzyme essential for coenzyme Q10 (CoQ10) biosynthesis, was conducted using a yeast complementation system and by enzyme activity measurement. CoQ10 levels in lymphoblastoid cells and brains were measured by high-performance liquid chromatography. Results We identified a homozygous mutation (M78V-V343A) and compound heterozygous mutations (R337X/V343A) in COQ2 in two multiplex families. Furthermore, we demonstrated that a common variant (V343A) and multiple rare variants of COQ2, both of which are functionally impaired, are associated with sporadic MSA. The V343A variant was exclusively observed in the Japanese population, while other multiple rare variants are observed irrespective of ethnic background. The ratio of MSA-C to MSA-P was significantly higher in COQ2 variant carriers than in noncarriers. Conclusions Functionally impaired variants of COQ2 associated with an increased risk of MSA in multiplex families and sporadic cases, supporting a role of impaired COQ2 activities in the pathogenesis of MSA.

2965F

Late-onset sacsinopathy diagnosed by exome sequencing and comparative genomic hybridisation. *A. Pyle[1], H. Griffin[1], J. Duff[1], S. Zwolinski[2], T. Smertenko[1], P. Yu-Wai-Man[1], M. Santibanez-Koref[1], R. Horvath[1], P.F. Chinnery[1].* 1) Wellcome Centre for Mitochondrial Research, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom; 2) Northern Genetics Service, Newcastle upon Tyne Hostpitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom.

The molecular diagnosis of adult-onset autosomal recessive cerebellar ataxias remains challenging because of genetic heterogeneity. However, recently developed molecular genetic techniques will potentially revolutionise the diagnostic approach. Here we set out to define the genetic basis of the ataxia in two brothers with no molecular diagnosis. Clinical evaluation was followed by whole exome second generation sequencing and comparative genomic hybridisation to determine the diagnosis. Whole exome sequencing identified a hemizygous novel spastic ataxia of Charlevoix-Saguenay (SACS) stop-codon mutation in both brothers (c.13048G>T, p.E4350X) which was present in the mother, but not the father. Comparative genomic hybridisation revealed a 0.7Mb deletion on chromosome 13q12.12 in both brothers, which included SACs and was heterozygous in the asymptomatic father. The milder phenotype, caused by a whole gene deletion and a stop codon mutation in SACS, indicates a loss-of-function mechanism in late-onset autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), and illustrates the importance of chromosomal rearrangements in the investigation of adult-onset ataxia.

2966W

Autosomal-recessive spastic ataxias: systematic whole-exome sequencing of a large cohort reveals novel phenotypes and gene candidates. *R. Schüle[1,2], M. Gonzalez[2], J. Reichbauer[1], A. Caballero Oteyza[1], S. Wiethoff[1,3], M. Dobler[4], U. Gaiser[4], T. Rattay[1], K. Karle[1], L. Schols[1], S. Zuchner[2], M. Synofzik[1].* 1) Department of Neurodegenerative Disease, Hertie Institute for Clinical Brain Research, Tuebingen, Germany; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida; 3) Brain Research Trust, UCL, Institute of Neurology, London, UK; 4) Department of Paediatric Neurology and Developmental Medicine, Children's Hospital, University of Tuebingen, Tuebingen, Germany.

Autosomal recessive hereditary spastic ataxias are a clinically and genetically poorly defined group of neurodegenerative disorders characterized by the combined occurrence of ataxic and spastic features. Patients are usually classified as either complicated Hereditary Spastic Paraplegia (HSP) or autosomal recessive spinocerebellar cerebellar ataxia (SCAR), often with a large overlap which makes the respective classification arbitrary in many cases . Few genes are known to cause spastic ataxia as the predominant phenotype and they explain only a small fraction of known cases. Among the most common known causes of autosomal recessive spastic ataxias are mutations in the SACS gene, causing autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), late onset Friedreich ataxia, caused by repeat expansions in the FXN gene and HSP Type 7 (SPG7). Using a comprehensive whole-exome sequencing approach, we aimed to systematically unravel the genetic basis of spastic ataxias, investigate atypical phenotypes and identify novel gene candidates. We have collected a cohort of 51 index patients with progressive early onset spastic ataxia and a family history compatible with autosomal recessive disease. Mean age of disease onset was 10.8 years. Additional symptoms present in some of the patients included mental retardation, optic atrophy, peripheral neuropathy, epilepsy and dysgenesis of the corpus callosum on MRI. Whole-exome sequencing was performed in all index cases (enrichment: SureSelect Human All Exon 50Mb kit, Agilent; sequencing: Hiseq 2000, Illumina). After alignment to the hg19 version of the human genome (Wurrows-Wheeler algorithm) and variant calling (Genome Analysis Toolkit GATK software package) data was imported into GEM.app, a browser accessible exome database and collaboration tool. Pathogenic mutations in known genes were identified in about one third of cases These include mutations in known spastic ataxia genes like SACS, SPG7, GBA2, or FA2H, but also genes not yet associated with spastic ataxia like e.g. ADCK3, SYNE1, or TACO1 among other genes Clinico-genetic findings in solved families as well as exciting new candidate genes for autosomal recessive spastic ataxias will be presented. These findings not only provide a systematic account of the genetic basis and phenotypic spectrum of spastic ataxias, but also enlarge the genetic spectrum that needs to be considered in spastic ataxias.

2967T

A dominantly inherited syndrome of adult-onset tremor, ataxia, parkinsonism, seizures and sensory neuropathy associated with c.970-1 G→A *PolG2* splice site mutation. *L. Van Maldergem[1], R.W. Taylor[2], R. Van Coster[3], G. Stevanin[4], A. Rolfs[5], M. Humble[6], E. Blakely[2], L. He[2], J. Piard[1], B. Depaepe[3], P. Hella[7], F.G. Debray[8], J-J. Martin[9], J. Bergmans[10], P. Laloux[11], A. Durr[4], S. Forlani[4], W.C. Copeland[6], E. Mormont[11].* 1) Centre de génétique humaine, Université de Franche-Comté, Besançon,, France; 2) Pathology Wellcome Trust Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, United-Kingdom; 3) Laboratory of Mitochondrial disorders University of Gent, Belgium; 4) CRICM (INSERM / UPMC 975, CNRS 7225) and neurogenetics laboratory, GHU Pitié-Salpêtrière, Paris, France; 5) Albrecht-Kossel-Institute for Neuroregeneration, University of Rostock, Rostock, Germany; 6) Mitochondrial DNA Replication Group National Institute of Environmental Health Sciences, Durham, NC; 7) Dept Neurology, Regional Hospital, Namur, Belgium; 8) Metabolic Unit, Centre of Human Genetics, University Hospital, Liège, Belgium; 9) Born-Bunge Foundation, University of Antwerp, Belgium; 10) Neurophysiology, private practice, Brussels, Belgium; 11) Dept Neurolgy, CHU UCL Mont-Godinne Dinant, Yvoir, Belgium and Institute of Neuroscience (IoNS), Université Catholique de Louvain (UCL), Brussels, Belgium.

Mutations in an increasing number of nuclear genes involved in the maintenance and replication of mitochondrial DNA (mtDNA) are being described, associated with an extensive spectrum of clinical phenotypes and secondary mtDNA defects that can take two forms: mtDNA depletion syndromes which are characterised by a quantitative loss of mtDNA copy number leading to isolated organ or multi-systemic pediatric mitochondrial disease and multiple mtDNA deletions which are associated with late-onset mitochondrial disease, a mosaic pattern of cytochrome c oxidase (COX)-deficiency and in which progressive external ophthalmoplegia (PEO) is the predominant clinical feature. We describe the clinical, pathological and molecular findings in a large Caucasian family segregating a late-onset (sixth to seventh decade) movement disorder comprising head and arms tremor, cerebellar ataxia, parkinsonism and sensory axonal peripheral neuropathy, complicated after a few years by generalized epileptic seizures and cognitive decline, in seven individuals. Spotty white matter T2 hypersignals were observed in cerebellum, cerebellar peduncles, brainstem and periaqueductal grey matter. Strikingly, PEO was mild or even absent. Muscle biopsy of four affected family members revealed the presence (2-25%) of COX deficient fibers, some of which exhibited 'ragged-red' pathology. Fluorescent imaging of mitochondrial membrane potential in patient fibroblasts revealed a mosaic staining pattern, characteristic of a heterogeneous mitochondrial defect. Long-range and real-time PCR analyses confirmed the presence of clonally-expanded mtDNA deletions as the cause of the focal COX deficiency. Candidate gene screening excluded mutations in POLG, PEO1, RRM2B, SLC25A4 and TK2 whereas sequencing of POLG2 revealed all affected individuals to be heterozygous for a c.970-1G>A splicing mutation. cDNA studies on muscle from two affected individuals indicated 10-15% exon5 skipping and more complex exon6 skipping. By contrast, seven unaffected family members were wild type at this POLG2 locus indicating complete segregation of the clinical phenotype with the novel splice-site mutation. This is the first complex neurological phenotype reported so far due to a mutation in the gene encoding the accessory subunit of human polymerase gamma.

2968F

Successful collaborative data mining for novel neuromuscular genes in Genomes Management Application (GEM.app). *S. Zuchner, F. Speziani, A. Abrams, A. Strickland, A.P. Rebelo, R. Ulloa, R. Acosta, M. Gonzalez.* Department for Human Genetics and Hussman Institute for Human Genomics, University of Miami, Miami, FL.

Many neuromuscular diseases are genetically heterogeneous with hundreds of Mendelian genes identified so far. Because of available exome sequencing analysis new genes are being reported on a nearly weekly basis. Still these genes explain only a portion of all patients for these phenotypes, indicating that a sizable number of new genes are still to be identified. In parallel, exome sequencing broadens the understanding of phenotype/genotype relationships and clinico-genetic overlap between diseases. When analyzing large amount of exome data, we identified a need for novel powerful genomic analysis tools that are seamlessly accessible to medical doctors and molecular biologists, and provide secure data sharing if needed. Over the past two years, we have designed the web-based tool GEM.app that allows for analysis and integration of multiple exome/ genome datasets across the world. GEM.app contains >3,000 exomes with extensive variant annotation from over 160 users in 18 different countries. It represents a fully functional analysis suite for exome data, is fast, yet easy to use. As of today GEM.app is used 250 times each day with a total of >35,000 queries performed by its users. Single exomes or families are queried in < 1 second and searches over hundreds of families typically take less than 4 seconds. Thus, it is especially powerful to search for common genes across several hundred patient exomes with the same phenotype. More than a dozen novel genes have been identified or contributed by GEM.app in the past 12 months. Nearly all of these genes represent collaborative data analyses. Recently published examples of hereditary spastic paraplegia genes include DDHD1, DDHD2, GBA2, CYP2U1, RTN2, B4GALNT1; Charcot-Marie-Tooth disease genes MARS, PDK3, AARS, HARS, HINT1, SLC25A; spinal muscular atrophy BICD2; different syndromes ANKRD11, PNPLA6, SARM1, DNAJC5; and many more in the pipeline. In addition, we have begun to use GEM.app for remote and local clinical genetics case conferences with geneticists, neurologists, genetic counselors, nurse practitioners and students to discuss specific findings in exomes of research cases. Taken together, we have created a new genomic platform for studies of neuromuscular diseases and beyond. Access is available to all investigators with exome/genome datasets at no cost. This platform can be used for a multitude of research and clinical applications and will continue to significantly simplify the interpretation of exomes.

2969W

Comprehensive genetic analysis of an Usher I patient cohort. *K.M. Bujakowska[1], M. Consugar[1], E. Place[1], M.F. Farkas[1], D. Taub[1], J. White[1], D. Navarro-Gomez[1], A. Langsdorf[1], M. Sousa[1], C. Weigel DiFranco[2], S. Harper[2], X. Gai[3], E.L. Berson[2], E.A. Pierce[1].* 1) Ocular Genomics, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA; 2) Berman-Gund Laboratory for the Study of Retinal Degenerations, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA; 3) Department of Molecular Pharmacology and Therapeutics, Loyola University Chicago Health Sciences Division, Maywood, Illinois, USA.

Usher syndrome is the most frequently inherited dual impairment of vision and hearing. Usher syndrome type 1 (USH1) is the most severe form characterized by profound congenital deafness, vestibular dysfunction and prepubescent onset of retinal degeneration. Currently, there are six genes associated with USH1: Myosin VIIa (MYO7A), Harmonin (USH1C), Cadherin-23 (CDH23), Protocadherin-15 (PCDH15), Sans (USH1G) and recently identified CIB2. The products of five of these genes have recently been identified to be part of the calyceal processes of photoreceptors. The purpose of the study was to genetically characterize a cohort of 38 USH1 probands. Methods: An USH1 cohort of 38, partially pre-screened for MYO7A mutations, was high-throughput sequenced for targeted exons of genes associated with inherited retinal degenerations. Genetically unsolved samples were subsequently screened for deletions using a comparative genomic hybridization (CGH) array and deep intronic mutations were investigated by sequencing of the targeted whole-gene capture of the known Usher genes. Intronic regions were mapped onto retina transcriptome. When possible, the likely pathogenic variants were confirmed by co-segregation in available family members with Sanger sequencing. The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by Institutional Review Board. Results: Selective exon capture and Illumina sequencing provided excellent coverage of the targeted exons, with >95% of exons having >10X sequence depth. With this sequencing approach, we were able to solve 55% of 38 probands, where mutations in MYO7A and CDH23 were the most frequent (18% each), followed by USH1C (11%). Eleven samples (29%) contained one heterozygous likely pathogenic change in a USH1 gene. The remaining, 16% did not carry mutations in any known USH1 gene. CGH array analyses and Usher whole-gene sequencing is in progress and will be reported. Conclusions: Targeted exon sequencing provides an effective approach for genetic diagnostic testing of patients with Usher syndrome type 1. We report 24 novel variants among 38 probands. This study also shows that there are some USH1 patients presently with an unknown genetic cause of the disease, suggesting possible copy number variation in known genes, deep intronic mutations or additional genetic loci for this condition. This work was supported by grants from the National Eye Institute (EY012910) and the Foundation Fighting Blindness.

2970T

SPG11 and SPG15 are the most frequent genotypes causing spastic paraplegia with thin corpus callosum, white matter changes and mental retardation in Italian patients. *C. Gellera[1], V. Pensato[1], B. Castellotti[1], E. Sarto[1], D. DiBella[1], L. Nanetti[1], D. Pareyson[2], E. Salsano[2], M. Eoli[3], C. Ciano[4], C. Mariotti[1], F. Taroni[1].* 1) Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS - Istituto Neurologico Carlo Besta, Milan, Italy; 2) Unit of Clinic of Central and Peripheral Degenerative Neuropaties, IRCCS - Fondazione Istituto Neurologico Carlo Besta, Milan, Italy; 3) Unit of Molecular Neuro-Oncology,IRCCS - Fondazione Istituto Neurologico Carlo Besta, Milan, Italy; 4) Unit of Clinical Neurophisiology, IRCCS - Fondazione Istituto Neurologico Carlo Besta, Milan, Italy.

Hereditary Spastic Paraplegias (HSPs) are clinically and genetically heterogeneous diseases characterized by lower limb spasticity and weakness. The genetic of HSPs is complex and at least 38 HSP genes have been identified as causative ones. A subgroup of complicate autosomal recessive HSPs (ARHSP) has been distinguished for the presence, of thin corpus callosum (TCC) and mild white matter lesions (WML). The ARHSP-TCC phenotype has been recognized in association with at least seven genes but causative mutations have been most frequently found in the KIAA1840 (SPG11) and ZFYVE26 (SPG15) genes. We selected a cohort of 53 unrelated Italian patients with complicated spastic paraplegia presenting with at least one clinical/neuro-radiological hallmark (mental retardation, TCC and/or WML). DNA samples from the patients were analyzed for mutations in SPG11 (KIAA1840), SPG15 (ZFYVE26), SPG21 (ACP33) and SPG5A (CYP7B1) genes. Patients included in this survey presented juvenile onset of neurological symptoms (14.6±10yrs). In 10 cases family history was compatible with a recessive inherited disorder, while the remaining were sporadic cases. Molecular investigations allowed the genetic diagnosis in 21 index cases: 15 patients, from 14 families, were found to carry pathogenic mutations in the SPG11 gene, and 7 unrelated patients were found to carry mutations in the SPG15 gene. The remaining 32 cases were negative also for mutations in SPG21 and SPG5A genes. SPG11 mutations were found in homozygous form in 4 cases and in compound heterozygous form in the remaining 10 patients. Overall, we found 19 different mutations with a large prevalence (74%) of nonsense mutations. Only, 7 mutations have been previously reported. SPG15 mutations were all newly identified mutations. They have been found in homozygous form in 6 cases (five carried a stop mutation and one carried a missense mutation), while one patient was compound heterozygous for a frameshift and a stop mutation. In all cases we could demonstrated the segregation of the identified mutations within families. In our cohort of Italian patients with complicated HSP we found SPG11 mutations in 26% of the cases, and SPG15 mutations in 13%. The majority of mutations are predicted to cause absence of the protein and 18 mutations were newly reported. Negative molecular results in at least 60% of cases, suggests further genetic heterogeneity among the patients with similar clinical presentation.

2971F

5q15 deletions: clinical and molecular characterization of a new syndrome. *E. Pallesi-Pocachard[1], J. Andrieux[2], A-M. Bisgaard Pedersen[3], R. Steensbjerre Møller[4], E. Buhler[5], A. Carabalona[5], P.H. Kaad[6], E. Parrini[7], D. Héron[8], B. Keren[8], B. Benyahia[8], N. Sobreira[9], V. Malan[10], L. Manouvrier-Hanu[11], T. Wang[9], R. Guerrini[7], A. Represa[5], M. Kirchhoff[3], M. Holder-Espinasse[11], C. Cardoso[5].* 1) Plateforme de Biologie Moléculaire et Cellulaire, INMED INSERM, Marseille, France; 2) Plateforme de génomique, Centre de Biologie Pathologie, CHRU, Lille, France; 3) Department of Clinical Genetics, Rigshospitalet, Denmark; 4) The Wilhelm Johannsen Centre for Functional Genome Research, Department of Molecular and Cellular Medicine, The Panum Institute, University of Copenhagen, Copenhagen, Denmark; 5) INMED, INSERM, Aix-Marseille University, Marseille, France; 6) Department of Paediatrics, Hjoerring Hospital, Hjoerring, Denmark; 7) Pediatric Neurology and Neurogenetics Unit and Laboratories, Children's Hospital A. Meyer-University of Florence, Florence, Italy; 8) Génétique Moléculaire et Chromosomique, Groupe Hospitalier Pitié-Salpêtrière, Paris, France; 9) Institute of Genetic Medicine and Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, USA; 10) Département de Cytogénétique et de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 11) Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, France.

Malformations of Cortical Development (MCD) are important causes of intellectual disability (ID) and account for 20-40% of drug-resistant epilepsy in childhood. A large number of MCD have now been identified and classified using embryology, genetic, and imaging criteria. Contrary to previous assumptions, the majority of these disorders are now thought to have a genetic basis, although environmental causes such as in utero infection or ischemia are still possible. Microdeletions of the 5q14.3-q15 region have been described in patients with abnormal brain development, severe ID and epilepsy. Further delineation of a critical region of overlap in these patients pointed to *MEF2C* as the causative gene. However, we previously reported two patients with a 5q14.3-q15 deletion that did not span *MEF2C* suggesting that other genes may be implicated in these conditions. Here, we investigated 6 additional patients harboring partially overlapping deletions that do not include *MEF2C*. The clinical phenotype includes moderate-to-severe mental retardation, hypotonia, epilepsy and MCD. The minimal critical region shared by all patients spans an interval of approximately 4.2 Mb in the 5q15 region. This interval contains 8 genes: *NR2F1, FAM172A, POU5F2, C5orf36, ANKRD32, MCTP1, FAM81B* and *TTC37*. Among these genes, we considered three of them as the best candidates since they are expressed in human as well as in rodent fetal brain tissues. To identify which gene contribute to MCD phenotype, we developed an experimental approach. Indeed, we used RNAi-based inactivation in rodents, enabling rapid functional evaluation of candidate genes potentially impacting neuronal migration and maturation. Using this approach, we succeeded to confirm that one of the candidate gene mapped to 5q15 is crucial for brain development.

2972W

Evolution of Magnetic Resonance Imaging (MRI) findings in a girl with Megalencephalic Leukoencephalopahty with Subcortical Cysts (MLC). *D.L. Renaud[1], N. Manolakos[2], G.M. Hobson[2].* 1) Div Child/Adolescent Neurology, Mayo Clinic, Rochester, MN; 2) Nemours Biomedical Research, Alfred I. duPont Hospital for Children, Wilmington, DE.

Background: MLC is characterized by macrocephaly and mild gross motor delay with characteristic MRI findings including diffuse white matter abnormalities and subcortical cysts in the bilateral anterior temporal regions. Mutations in the MLC1 and HEPACAM genes have been described. Objective: To describe the evolution of MRI findings in MLC. Case report: A 7 month old girl presented with macrocephaly (with enlarged anterior fontanelle) and mild gross motor delay. Her MRI revealed diffuse white matter abnormalities with some anterior predominance. The white matter of the brainstem and cerebellum appeared normal. MR spectroscopy was normal. Normal evaluation included thyroid function, lysosomal enzymes, peroxisomal panel, organic acids, lactate, GFAP gene analysis and eye exam. At 10 months, the MRI was unchanged with no new myelin formation suggesting hypomyelination. Lumbar puncture for metabolic studies was normal. At 17 months, the MRI revealed unchanged white matter abnormalities but the new finding of subcortical cysts in the anterior temporal lobes. Molecular analysis revealed two heterozygous mutations in the MLC1 gene and one heterozygous mutation in the HEPACAM gene. At 2 years of age, she continues to have normal intelligence with mild gross motor delay. Conclusion: Subcortical cysts may not be present initially in patients with MLC.

2973T

***De novo* mutations in the autophagy gene encoding WDR45 (WIPI4) cause static encephalopathy of childhood with neurodegeneration in adulthood.** *H. Saitsu[1], T. Nishimura[2], K. Muramatsu[3], H. Kodera[1], S. Kumada[4], K. Sugai[5], E. Kasai-Yoshida[4], N. Sawaura[3], H. Nishida[6], A. Hoshino[6], F. Ryujin[7], S. Yoshioka[7], Y. Kondo[1], Y. Tsurusaki[1], M. Nakashima[1], N. Miyake[1], H. Arakawa[3], M. Kato[8], N. Mizushima[2,9], N. Matsumoto[1].* 1) Yokohama City University, Yokohama, Kanagawa, Japan; 2) The University of Tokyo, Tokyo, Japan; 3) Gunma University Graduate School of Medicine, Gunma, Japan; 4) Tokyo Metropolitan Neurological Hospital, Tokyo, Japan; 5) National Center of Neurology and Psychiatry, Tokyo, Japan; 6) National Rehabilitation Center for Children with Disabilities, Tokyo, Japan; 7) Shiga University of Medical Science, Shiga, Japan; 8) Yamagata University Faculty of Medicine, Yamagata, Japan; 9) Tokyo Medical and Dental University, Tokyo, Japan.

Static encephalopathy of childhood with neurodegeneration in adulthood (SENDA) is a recently established subtype of neurodegeneration with brain iron accumulation (NBIA)1-3. By exome sequencing, we found de novo heterozygous mutations in *WDR45* at Xp11.23 in two individuals with SENDA, and three additional *WDR45* mutations in three other subjects by Sanger sequencing. Using lymphoblastoid cell lines (LCLs) derived from the subjects, aberrant splicing was confirmed in two subjects, and protein expression was observed to be severely impaired in all five subjects. *WDR45* encodes WD repeat protein interacting with phosphoinoside 4 (WIPI4), one of the four mammalian homologs of yeast Atg18, which plays an important role in autophagy4,5. Decreased autophagic activity and accumulation of aberrant early autophagic structures were demonstrated in the LCLs of the subjects. These findings provide the direct evidence that an autophagy defect is indeed associated with a neurodegenerative disorder in humans.

2974F

Alteration of ganglioside biosynthesis responsible for complex hereditary spastic paraplegia, SPG26. *G. STEVANIN[1,2,3], A. Boukhris[4], R. Schule[5], C.M. Lourenço[6], J.L. Loureiro[7], M.A. Gonzalez[8], P. Charles[3], I. Rekik[4], J. Gauthier[9], R.F. Acosta-Lebrigio[8], F. Speziani[8], A. Ferbert[10], A. Caballero-Oteyza[5], A. Dionne-Laporte[9], A. Noreau[9], M. Gaussen[1,2], P. Coutinho[7], P. Dion[9], F. Mochel[1,3], J. Pouget[11], C. Mhiri[4], L. Schols[5], F. Darios[1], G.A. Rouleau[9], W. Marques Jr[6], A. Durr[1,3], S. Zuchner[8], A. Brice[1,3].* 1) Institut du Cerveau et de la Moelle épinière / NEB, Paris, France; 2) Ecole Pratique des Hautes Etudes, Neurogenetics team, Paris, France; 3) APHP, Fédération de Génétique, Pitié-Salpêtrière Hospital, Paris, France; 4) Hôpital Universitaire Habib Bourguiba, Sfax, Tunisia; 5) Department of Neurodegenerative Diseases and Hertie-Institute for Clinical Brain Research, University of Tübingen, Germany; 6) Universidade de São Paulo, Ribeirão Preto, Brazil; 7) UnIGENe and Centro de Genética Preditiva e Preventiva, Institute for Molecular and Cellular Biology, Porto, Portugal; 8) University of Miami, FL, USA; 9) McGill University, Montreal, Quebec, Canada; 10) Klinikum Kassel, Germany; 11) CHU La Timone, Marseille, France.

Hereditary spastic paraplegias (HSPs) form a heterogeneous group of neurological disorders. A whole genome linkage mapping effort was made using three HSP families from Spain, Portugal and Tunisia and it allowed us to reduce the SPG26 locus interval from 34 to 9 Mb. Following this, a targeted capture was made to sequence the entire exome of affected individuals from the three above families, as well as from two additional autosomal recessive HSP families of German and Brazilian origins. Five homozygous truncating (n=3) and missense (n=2) mutations were identified in B4GALNT1. Following this finding, we analyzed the entire coding region of this gene in 65 additional cases, and three mutations were identified in two subjects. All mutated cases presented an early onset spastic paraplegia, with frequent intellectual disability, cerebellar ataxia and peripheral neuropathy as well as cortical atrophy and white matter hyperintensities on brain imaging. B4GALNT1 encodes beta-1,4-N-acetyl-galactosaminyl transferase 1 (B4GALNT1), involved in ganglioside biosynthesis. These findings confirm the increasing interest of lipid metabolism in HSPs. Interestingly, while the catabolism of gangliosides is implicated in a variety of neurological diseases, SPG26 is the second human disease involving defects of their biosynthesis.

2975W

Neurofilament light gene mutation causes hereditary motor and sensory neuropathy with pyramidal signs. *H. Takashima, A. Hashiguchi, M. Nomura, H. Arata, A. Yoshimura, Y. Higuchi, Y. Sakiyama, T. Nakamura, Y. Okamoto, E. Matsuura.* Dept Neurology and Geriatrics, Kagoshima Univ, Kagoshima, Japan.

Objective: To identify novel mutations causing hereditary motor and sensory neuropathy type V (HMSN-V), a variant of Charcot-Marie-Tooth disease (CMT), we screened known CMT and related genes in members of a Japanese family with HMSN-V. Methods: The clinical features, peripheral nerve conduction properties, and MRI findings of four patients from the Japanese family were examined to confirm HMSN-V. We then screened 28 CMT disease-causing and related genes using a custom microarray chip. Results: The clinical features included mild weakness of the distal lower limb muscles, foot deformity, mild sensory loss, and late-onset progressive spastic paraparesis. Electrophysiological studies revealed slower sensory and motor nerve conduction and small sensory and muscle compound action potentials in multiple nerves in limbs, indicating widespread neuropathy. Brain MRI revealed an abnormally thin corpus callosum. In all four patients, microarrays detected a novel heterozygous missense mutation c.1166A>G (p.Y389C) in the gene encoding the light chain neurofilament protein (NEFL).This mutation is segregated with disease and patient phenotype in this family. Conclusion: All members of the Japanese family with motor and sensory neuropathy with pyramidal tract involvement harbored the same novel NEFL mutation, indicating that NEFL mutations can result in an HMSN-V phenotype.

2976T

***KIF5A* gene mutations in patients with spastic paraplegia.** *F. Taroni[1], S. Caldarazzo[1], E. Sarto[1], S. Baratta[1], K. Savio[2], G. Galassi[3], A. Ariatti[3], G. Lauria[4], L. Nanetti[1], E. Salsano[5], D. Pareyson[5], C. Mariotti[1], D. Di Bella[1].* 1) Genetics of Neurodegenerative & Metabolic Diseases, IRCCS Ist Neurol Carlo Besta, Milan, Italy; 2) Neurology Unit, Ospedale degli Infermi, Biella, Italy; 3) Neurology Unit, Ospedale di Modena, Modena, Italy; 4) Neuromuscular Disease Unit, IRCCS Ist Neurol Carlo Besta, Milan, Italy; 5) Neurology 8, IRCCS Ist Neurol Carlo Besta, Milan, Italy.

Hereditary spastic paraplegias (HSP) are a clinically and genetically heterogeneous group of neurodegenerative disorders. Pure and complicated forms of the disease have been described. About half of HSP cases result from autosomal dominant (AD) mutations in *spastin* (SPG4), *atlastin-1* (SPG3A), or *REEP1* (SPG31) genes. Mutations in the *KIF5A* gene have been reported to be a relatively frequent cause (approx. 10%) of pure or complicated AD-HSP phenotypes (SPG10) in French and Italian populations. The *KIF5A* gene encodes the neuronal kinesin heavy chain implicated in the anterograde axonal transport. Most mutations identified so far (all missense except one deletion) are located in the motor domain, with the exception of one mutation identified in the neck region and one in the stalk domain of the protein. The motor domain (exons 1-11) of *KIF5A* was analysed by high-resolution melting (HRM) analysis and/or direct sequencing in 350 unrelated HSP index cases, including 185 AD and 165 sporadic cases, negative for SPG4 mutations. We identified 10 different missense mutations, 4 of which are novel, in 11 HSP probands. The novel mutations were not detected in >350 normal alleles. By *in silico* analysis, all mutations are predicted to damage protein structure or function. Six mutations were identified in AD-HSP probands, 3 cases were apparently sporadic while for 2 probands family history was either not available or unclear. Our patients showed a wide range of age-at-onset, with 4 patients exhibiting onset in childhood, 2 patients at >50 yrs, and the remaining patients in the 3rd-4th decade. Interestingly, 9/11 patients had a complex phenotype with axonal neuropathy, parkinsonism, ataxia, and deafness as variably associated symptoms. *Conclusions*: In our study, *KIF5A* mutations account for 3.5% of cases with autosomal dominant inheritance or sporadic occurrence, a frequency lower than that observed in other Italian (Crimella 2012) and French (Goizet 2009) patient series but similar to that (4%) found in a mixed European population (Schüle 2008). Our findings indicate that SPG10 should be considered in the molecular analysis of HSP phenotypes, even in the absence of a clearly positive family history. Clinically, the probands of our SPG10 families present with a rather broad range of phenotypes, with polyneuropathy or Charcot-Marie-Tooth phenotype as the most frequently associated feature. (Partly supported by grant RF-2009-1539841 from Italian Ministry of Health to F.T.).

2977F

A Locus for Autosomal Recessive Congenital Mirror Movement Disorder Maps to 22q13.1. *J.B. Vincent[1], M.A. Rafiq[1], P. John[2], I.A. Balouch[1,2].* 1) Neurogenetics Sect, R30, Ctr Add/Mental Hlth,Clarke Div, Toronto, ON, Canada; 2) Atta-ur-Rehman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad-Pakistan.

Mirror Movement (MRMV) disorder is a rare, mainly autosomal disorder, although it is thought that some sporadic cases may be due to recessive inheritance. MRMV refers to voluntary movements on one side of the body being mirrored by involuntary movements on the opposite side. Using a linkage analysis and candidate gene approach two genes so far has been in implicated to MRMV: DCC on 18q21.2, which encodes a netrin receptor, and on 15q15.1, which is involved in the maintenance of genomic integrity. Here, we describe a large consanguineous Pakistani family with cases 11 cases of MRMV reported across 5 generations, with autosomal recessive inheritance likely. We used Sanger sequencing to exclude DCC and RAD51. We then employed microarray genotyping and autozgosity mapping to identify a shared region of homozygosity-by-descent among the affected individuals. We identified a single autozygous region of ~3.3 Mb on chromosome 22q13.1 (Chr22:36605976-39904648). We used Sanger sequencing to exclude several candidate genes within this region, including DMC1. Whole exome sequencing is currently underway to identify the causative mutation in this family.

2978W

The Whole Genome Sequence of a Jack Russell Terrier with Progressive Spinocerebellar Ataxia and Myokymia Contains a Homozygous Disease-Associated KCNJ10 Missense Mutation. *D. Gilliam[1], J.R. Coates[2], G.S. Johnson[1], L. Hansen[1], T. Mhlanga-Mutangadura[1], J.F. Taylor[3], G.C. Johnson[2], R.D. Schnabel[3], D.P. O'Brien[2].* 1) Department of Veterinary Pathobiology, University of Missouri College of Veterinary Medicine, Columbia, MO; 2) Department of Veterinary Medicine and Surgery, University of Missouri College of Veterinary Medicine, Columbia, MO, USA; 3) Division of Animal Sciences, University of Missouri College of Food Agriculture and Natural Resources, Columbia, MO, USA.

A distinct progressive autosomal recessive spinocerebellar ataxia occurs in three related dog breeds: Jack Russell Terriers, Parson Russell Terriers, and Russell Terriers (JRT/PRT/RTs). Affected dogs develop cerebellar ataxia by 2-to-6 months of age. As the disease progresses, most dogs exhibit myokymia and some have seizures. We used Illumina sequencing technology to generate a whole genome sequence (WGS) with 20-fold coverage from an affected JRT and aligned the reads to the canine reference genome sequence using NextGENe software. The detected sequence variants were prioritized according to predicted functionality and filtered to remove variants also present in any of 22 WGSs generated for dogs with other heritable diseases which served as controls. Among the 107 homozygous sequence variants in 89 genes that survived the filter step, a KCNJ10:c.627C>G transversion predicting a p.I209M amino acid substitution was considered most likely to be causal because cerebellar ataxia and seizures are features of SeSAME syndrome caused by mutations in human KCNJ10. In addition, p.209I and surrounding amino acids are highly conserved components of the functionally important inward rectifier potassium channel domain. We used a TaqMan allelic discrimination assay to determine the genotypes of individual dogs at KCNJ10:c.627. The 13 JRT/PRT/RTs in our collection with spinocerebellar ataxia were all homozygous for the KCNJ10:c.627G allele; whereas, only one of the 879 genotyped JRT/PRT/RTs with undetermined phenotype was a KCNJ10:c.627G homozygote. We were unable to ascertain any clinical information for this KCNJ10:c.627G homozygote. The other genotyped JRT/PRT/RTs were either heterozygotes (n = 100) or homozygous for the reference allele (n = 779). One hundred and two dogs from 62 other breeds were all KCNJ10:c.627C homozygotes. There appear to be species differences between human and canine KCNJ10 deficiency diseases. SeSAME syndrome patients are deaf; whereas, the affected JRT/PRT/RTs have normal hearing. In addition, although myokymia is often exhibited by affected JRT/PRT/RTs, it is not considered a feature of SeSAME syndrome. This report illustrates that the mutation responsible for an unmapped disease can be identified from the WGS of a single affected dog plus controls.

2979T

The *Alu*-rich architecture of *SPAST* predisposes to diverse and functionally distinct CNV alleles. *P.M. Boone[1], I.M. Campbell[1], B.C. Baggett[1], J.C. Scull[2], C.J. Shaw[1], M.A. Withers[1], P. Moretti[1,3,4], W.E. Goodwin[5], J.K. Fink[6], A. Ordóñez-Ugalde[7], B. Quintáns[7], M-J. Sobrido[7], S. Stemmler[8], J.R. Lupski[1,2,9,10].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Medical Genetics Laboratories, Baylor College of Medicine; 3) Dept. of Neurology, Baylor College of Medicine; 4) Michael E. DeBakey VA Medical Center, Houston, TX, USA; 5) Children's Medical Center of Dallas, TX, USA; 6) Department of Neurology, University of Michigan, Ann Arbor, MI, USA; 7) Grupo de Medicina Xenómica, Fundación Pública Galega de Medicina Xenómica, SERGAS, IDIS, CIBERER, Santiago de Compostela, Spain; 8) Department of Human Genetics, Ruhr University, Bochum, Germany; 9) Texas Children's Hospital, Houston, TX, USA; 10) Department of Pediatrics, Baylor College of Medicine.

SPAST is the disease gene for autosomal dominant spastic paraplegia, type 4 (SPG4). While disease-causing deletions and duplications (CNVs) in *SPAST* have been described, their origin and molecular consequences remain obscure. We mapped 44 *SPAST* CNVs with nucleotide resolution in subjects with SPG4. A diverse combination of exons were deleted or duplicated, providing insights into the importance of particular exons for spastin function. 75% of CNV breakpoints localized to *Alu* elements, suggesting that the *Alu*-rich architecture of *SPAST* renders this locus susceptible to a variety of small genome rearrangements and potentially explaining why a large percentage of mutations in SPG4 patients are CNVs. Nine deletions overlapped part of *SPAST* and part of a nearby, directly-oriented gene, creating a chimeric gene in these patients' genomes. An additional 10 deletions of the 3' end of *SPAST* suggest that splicing from *SPAST* to *SLC30A6*, the most proximal downstream gene, may occur or that *SLC30A6* expression may be altered; PCR and sequencing of cDNA from a patient with a *SPAST* exon 17 deletion revealed multiple *SPAST*:*SLC30A6* fusion transcripts. As *SLC30A6* has been implicated in Alzheimer disease, these data may explain a report of spastic paraplegia and dementia co-segregating in a family with deletion of *SPAST* exon 17. In summary, we provide evidence that the *Alu* architecture of *SPAST* predisposes to diverse, disease-causing CNV alleles with distinct transcriptional - and possibly phenotypic - consequences. †The authors acknowledge Athena Diagnostics, a subsidiary of Quest Diagnostics, for contributing a cohort of anonymized DNA samples.

2980F

Tank-binding kinase 1 (*TBK1*), autophagy, and glaucoma. *J.H. Fingert[1,2], K. Anfinson[1,2], F. Solivan-Timpe[1,2], B.R. Roos[1,2], A. Khanna[1,2], A.L. Robin[3,4], R.F. Mullins[1,2], B.A. Tucker[1,2].* 1) Ophthalmology and Visual Sciences, Carver College of Medicine, University of Iowa, Iowa City, IA; 2) Institute for Vision Research, Iowa City, IA; 3) Department of Ophthalmology and International Health, Johns Hopkins University, Baltimore, MD; 4) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

Purpose: Retinal ganglion cell (RGC) loss is a key feature of glaucoma. We have previously shown that duplication of the TANK binding kinase 1 (*TBK1*) gene is a cause of normal tension glaucoma (NTG). *TBK1* and other NTG genes (*OPTN* and *TLR4*) have important roles in autophagy. Consequently, we derived RGCs from human skin cells to investigate the effects of *TBK1* gene duplication on autophagy in the most relevant tissue for glaucoma pathophysiology. **Methods**: Adult induced pluripotent stem cells (iPSCs) were generated from human skin cells (fibroblasts or keratinocytes) via infection with a single lentiviral vector, which was designed to drive expression of the Yamanaka factors Oct4, Sox2, Klf4, and c-MYC. Pluripotency was assessed by testing for the induced expression of the Yamanaka factors with RT-PCR. Neuronal cells were produced by differentiating iPSCs via culture in a series differentiation media containing Noggin, Dkk-1, IGF1, bFGF, DAPT, and aFGF. Differentiation into iPSC-derived retinal ganglion-like cells was assessed by testing for production of the RGC markers BRN3B, NF200, and THY1 with immunocytochemistry. Additional immunocytochemistry and Western blot experiments were used to search for altered production or altered localization of key factors in autophagy, including LC3-II (a marker of activated autophagy). **Results**: Generation of iPSCs from skin cells of patients with *TBK1*-related glaucoma and from controls was confirmed by RT-PCR amplification of all four Yamanaka factors. iPSC-derived neurons from both glaucoma patients with a *TBK1* gene duplication and from control subjects were positive for the RGC markers BRN3B, NF200, and THY1. Finally, Western blot analysis showed significantly increased levels of LC3-II in cells from glaucoma patients with a *TBK1* gene duplication. **Discussion**. We have successfully generated cultured neurons that expressed RGC markers and harbor a glaucoma-causing *TBK1* gene duplication. Initial studies show that a key marker of autophagy is greatly increased in cells with a *TBK1* gene duplication. These data provide strong evidence that activation of autophagy may be an important event in the pathophysiology of glaucoma. Further studies of the autophagy pathway using these iPSC-derived RGCs may provide additional insights into the causes of glaucoma at the molecular level and may also suggest new therapeutic interventions.

2981W

Lamin B1 over-expression is associated with nuclear defects and alterations in nuclear structure. *H. Rolyan[1], K. Dahl[2], N. Dahl[3], A. Melberg[4], C.F.M. de Souza[5], F.P. Vairo[5], A. Brusco[6], Q.S. Padiath[1].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 2) Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, USA; 3) Dept. of Immunology, Genetics and Pathology, Section of Clinical Genetics, The Rudbeck laboratory, Uppsala University Children's Hospital, Uppsala, Sweden; 4) Department of Neuroscience, Neurology, Uppsala University, Uppsala Sweden; 5) Hospital de Clínicas de Porto Alegre & Universidade Federal do Rio Grande do Sul; Brasil; 6) University of Torino, Department of Medical Sciences, Torino, Italy.

Autosomal Dominant Leukodystrophy (ADLD) is an adult onset demyelinating disease caused by duplications of the lamin B1 (LMNB1) gene. LMNB1 is a component of the nuclear lamina and together with providing structural integrity to the nucleus, it also plays an important role in gene regulation and chromatin modification. The mechanism by which LMNB1 duplications cause ADLD is unknown and it is also not known whether LMNB1 duplications lead to altered nuclear structure. To address this question, we have analyzed the nuclear structure in fibroblasts from ADLD patients. We have shown for the first time that patient fibroblasts show a greater number of nuclear abnormalities, alterations in nuclear shape and size and have altered chromatin landscapes. These results were confirmed in mouse fibroblast lines over expressing LMNB1. Our results suggest that alterations in nuclear structure are a functional consequence of LMNB1 over expression that may play an important role in disease causation.

2982T

Mechanisms underlying non-recurrent microdeletions causing neurofibromatosis type-1 (NF1). *J. Vogt[1], K. Bengesser[1], K. Claes[2], K. Wimmer[3], L. Messiaen[4], V.-F. Mautner[5], R. van Minkelen[6], E. Legius[7], H. Brems[7], T. Rosenbaum[8], M. Upadhyaya[9], D.N. Cooper[9], H. Kehrer-Sawatzki[1].* 1) Institute of Human Genetics, Ulm University, Ulm, Baden-Wuerttemberg, Germany; 2) Center for Medical Genetics, Ghent University Hospital, 9000 Ghent, Belgium; 3) Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria; 4) Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, Birmingham, USA; 5) Department of Neurology, University Hospital Hamburg Eppendorf, Hamburg, Germany; 6) Department of Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 7) Department of Human Genetics, KU Leuven, Leuven, Belgium; 8) Clinical Centre Duisburg, Children's Hospital, Germany; 9) Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK.

NF1 microdeletions encompassing the NF1 gene region at 17q11.2 are present in ~5% of patients with NF1. Although the mechanisms underlying recurrent NF1 microdeletions have been investigated in great detail, those underlying non-recurrent (atypical) NF1 microdeletions are not well delineated. NF1 microdeletions with non-recurrent breakpoints are heterogeneous in terms of their size, breakpoint position and number of deleted genes. In this study, we investigated 20 atypical NF1 deletions using high-resolution custom aCGH and performed breakpoint-spanning PCRs with primers located in non-deleted regions closely flanking the deletion boundaries. Sequence analysis of breakpoint-spanning PCR products indicated that 15 deletions exhibit simple breakpoint junctions without further complexities and hence were probably mediated by NHEJ/MMEJ. We therefore conclude that NHEJ/MMEJ is the most common mechanism underlying atypical NF1 deletions. However, two of the 20 deletions investigated exhibited complex breakpoints characterized by SVA-F element insertion at the deletion junctions, as determined by inverse PCR, semi-specific PCR and Genome Walker analysis. Large genomic deletions associated with SVA-element insertions at the breakpoints have not been previously reported. We postulate that SVA insertions, occurring via retrotransposition, that trigger large genomic rearrangements may represent an as yet uncharacterized mechanism responsible for causing CNVs with non-recurrent breakpoints. We noted that 14 (70%) of the 20 atypical NF1 microdeletions displayed proximal breakpoints located within a 38-kb region of the SUZ12P. The enrichment of atypical NF1 deletion breakpoints in SUZ12P is remarkable since the breakpoints of recurrent type-2 NF1 deletions are also located within SUZ12P. However, only two atypical NF1 deletion breakpoints co-localized with the breakpoints of previously analysed type-2 NF1 deletions. Hence, a clustering of breakpoints within a single region of SUZ12P was not observed. Nevertheless, the accumulation of breakpoints associated with recurrent and non-recurrent NF1 deletions within SUZ12P is indicative of its genomic instability. The analysis of FISH, microsatellite markers and insertion/deletion polymorphisms revealed somatic mosaicism in 12 of the 20 (60%) atypical NF1 deletion patients investigated. Thus, not only recurrent type-2 NF1 deletions, but also a considerable proportion of atypical NF1 deletions, are of postzygotic origin.

2983F

West syndrome caused by homozygous mutation in an evolutionary conserved mitochondrial elongation factor. *A.A. Alfaiz[1,2,3], N. Boutry-Kryza[4,5,6], D. Ville[7], N. Guex[2], J. Bellescize[8], C. Rivier[9], A. Labalme[4], V. Portes[5,7], P. Edery[4,5,6], M. Till[4], I. Xenarios[1,2], J. Herrmann[10], D. Sanlaville[4,5,6], G. Lesca[4,5,6], A. Reymond[1].* 1) Center for Integrative Genomics , University of Lausanne, Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics Quartier Sorge - Batiment Genopode 1015 Lausanne Switzerland; 3) King Abdullah International Medical Research Center, National Guard Health Affairs King Abdulaziz Medical City- Riyadh P.O. BOX 22490 Riyadh 11426; 4) Department of Constitutional Cytogenetics, Lyon Hospices Civils, Lyon, France; 5) Claude Bernard Lyon I University, Lyon, France; 6) CRNL, CNRS UMR 5292, INSERM U1028, Lyon, France; 7) Center of reference for Tuberous Sclerosis and rare epileptic disorders, Hospices Civils de Lyon, Lyon, France; 8) Department of Pediatric Epileptology, Lyon Hospices Civils, Lyon, France; 9) Department of Pediatrics, Hopital Nord-Ouest, Villefranche-sur-saône, France; 10) AG Zellbiologie TU Kaiserslautern - FB Biologie Erwin Schrödinger-Straβe 13 (Gebäude 13) 67663 Kaiserslautern.

Infantile spasms (IS), pathognomonic hypsarrhythmia and developmental regression are the triad of features that identify West's Syndrome (WS), a rare epileptic disease affecting about 1:3,500 live births. It occurs either in isolation or as part of a more complex disease syndrome. This heterogeneity often impedes the identification of the etiology of WS within the clinical practice. We identified using exome sequencing a homozygous mutation in the GUF1 gene in three siblings from a consanguineous family affected with isolated West syndrome and severe neurological impairment. The encoded GUF1 protein is a universally conserved GTPase in mitochondria and chloroplasts. Its prokaryote homolog, named EF4 or LepA, encodes a non-essential translational elongation factor able to back-translocate the tRNA on ribosomes. The EF4 gene is essential when suboptimal conditions trigger the formation of aberrant elongating ribosomes probably because the encoded mitochondrial translational elongation factor has the capacity of remobilizing stuck ribosomes under these suboptimal conditions. It was proposed that the strong selective advantage conferred by the presence of this elongation factor both in eu- and prokaryote could explain its ubiquitous distribution, as well as why it is one of the most conserved proteins. The variant we identified and which segregates with WS modifies an Alanine residue conserved in ALL eukaryotic organisms from budding yeast to human. It is positioned within the beta-strand moiety of the GTPase that specifically recognizes the acceptor stem and D-loop of tRNA molecules. Nuclear defects affecting mitochondrial protein synthesis might lead to severe epileptic disorders in infants such as IS/WS. They are, however, easily overlooked possibly because of the difficulty to correctly gauge the activity of the respiratory chain complexes, especially when no lactate and/or pyruvate increase is identified in blood and/or cerebrospinal fluid. We show that exome sequencing is a suitable strategy for identifying causative genes in familial IS/WS since many of the causative nuclear genes are probably yet unknown. We establish a new link between alteration of mitochondrial protein synthesis and IS/WS.

2984W

Gene discovery and high-throughput resequencing of candidate genes in epileptic encephalopathy. *H. Mefford[1], E. Sherr[2], I. Scheffer[3], A. Poduri[4], D. Dlugos[5], Epi4K Investigators.* 1) Pediatrics, University of Washington, Seattle, WA; 2) Neurology, University of California, San Francisco, CA; 3) Neurology, University of Melbourne, Australia; 4) Neurology, Children's Hospital, Boston, MA; 5) Neurology, Children's Hospital of Philadelphia, PA.

Epileptic encephalopathies (EE) are severe childhood epilepsy syndromes characterized by multiple refractory seizure types, developmental arrest or regression and a poor prognosis. In most cases, the cause is unknown, but *de novo* genetic mutations have been identified in some patients. We recently performed exome sequencing in 264 probands and their unaffected parents for two classic EE syndromes: infantile spasms (IS; n=149) and Lennox Gastaut syndrome (LGS; n=115). We confirmed 329 *de novo* mutations in 309 genes and found an excess of *de novo* mutations in the ~4000 genes that are most intolerant to functional genetic variation in the human population (Epi4k & EPGP Investigators, 2013, *Nature*). Notably, we identified four patients with *de novo* mutations in *GABRB3* and two patients with the identical *de novo* mutation in *ALG13*, confirming these two genes as novel EE genes. While many of the other genes in which there were *de novo* mutations are reasonable candidate genes for EE, identification of additional patients with mutations is necessary to confirm the role of each gene in the genetic etiology of EE. Here, we perform targeted capture and high-throughput resequencing of 31 genes in which a *de novo* mutation was identified in one or more patients with IS or LGS in our prior study. We selected genes in which two unrelated probands had a *de novo* mutation (n=4, including *GABRB3* and *ALG13*), genes that are among those intolerant to genetic variation and in which the *de novo* mutation was severe (nonsense, frameshift, splice; n=8), and genes in which mutations have previously been reported in other epilepsy syndromes, brain malformation syndromes, intellectual disability, autism or related phenotypes (n=19). We are screening ~600 patients with a range of EE syndromes, as we anticipate that many of the genes will cause EE disorders other than IS or LGS. We will report the frequency of *de novo* mutations of each candidate gene in our cohort and investigate genotype-phenotype correlations for each gene in which multiple patients harbor mutations.

2985T

Mutations of *TBC1D24* can cause nonsyndromic deafness DFNB86 or unrelated epilepsy disorders. *A.U. Rehman[1], K. Lee[2], R.J. Morell[1], T. Ito[3], M.C. Drummond[1], S. Riazuddin[4], R.L.P. Santos-Cortez[2], S.N. Khan[5], A.J. Griffith[3], S. Riazuddin[5], S.M. Leal[2], T.B. Friedman[1].* 1) Laboratory of Molecular Genetics, NIDCD, National Institutes of Health, Rockville, MD; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Otolaryngology Branch, NIDCD, National Institutes of Health, Rockville, MD; 4) Division of Pediatric Otolaryngology, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH; 5) Allama Iqbal Medical College/Jinnah Hospital Complex, Lahore, Pakistan.

Many nonsyndromic deafness genes remain to be identified. We previously reported that *DFNB86*, a locus for autosomal recessive nonsyndromic deafness, is located on chromosome 16p (maximum two-point LOD score 8.54; Ali and Rehman et al., 2012). We subsequently identified three additional large families segregating nonsyndromic deafness that showed statistically significant linkage to the *DFNB86* interval. To identify the causative gene for DFNB86 deafness, we sequenced the exome of one affected individual from three of the four DFNB86 families. In each exome dataset we identified two to nine homozygous DNA variants in the *DFNB86* interval that were present at an allele frequency of <1% in the public SNP databases and were predicted to be pathogenic by at least one *in silico* prediction program. Assuming there is a single gene responsible for all of the deafness linked to the *DFNB86* locus, we searched for mutations of a single gene across the three exome datasets. We found that affected individuals from two of the three families were homozygous for the same missense mutation, p.Asp70Tyr, of *TBC1D24*. Affected individuals of the third family were homozygous for another missense mutation, p.Arg293Pro, of *TBC1D24*. The fourth DFNB86 family was also segregating the p.Asp70Tyr allele of *TBC1D24*. These missense mutations co-segregated with deafness in the respective families and we did not observe them in 358 control chromosomes from an ethnically matched cohort. Both missense mutations alter highly conserved residues of the wild type TBC1D24 protein and are predicted to be pathogenic by a variety of *in silico* prediction programs including MutationTaster and PolyPhen-2. The splice isoforms, expression patterns and possible functions of *TBC1D24* in the auditory system will be discussed. Enigmatically, mutations of *TBC1D24* were previously reported to cause focal epilepsy and intellectual disability syndrome or familial infantile myoclonic epilepsy, with no reports of hearing loss in any of these patients. Clinical re-evaluations of the affected subjects of our DFNB86 families indicated that deafness appears to be nonsyndromic.

2986F

HIPPOCAMPAL DEFECTS IN THE FMR1 KNOCKOUT MOUSE. *F. Kooy[1], I. Heulens[1], V. Sabanov[2], T. Ahmed[2], A. Popp[3], R. Willemsen[3], R. D'Hooge[2], D. Baltschun[2], L. Rooms[1].* 1) Dept Medical Genetics, Univ Antwerp, Antwerp, Belgium; 2) Laboratory of Biological Psychology, KU Leuven, Belgium; 3) Department of Clinical Genetics, Erasmus MC Rotterdam, The Netherlands.

Whole-cell patch-clamp recordings on hippocampal slices from CA1 pyramidal neurons revealed a significant decrease (p=0.0127) of IPSCs in three-week-old Fmr1 knockout mice. Several drugs that modulate the GABA(A) receptor are available for clinical use. Ganaxolone, a positive allosteric modulator of GABA(A) receptors, prevents seizures in Fmr1 knockout mice. To further explore the therapeutic effect of ganaxolone, we evaluated the effect of the drug in the marble burying paradigm. Acute ganaxolone treatment is able to correct the abnormal marble burying behavior of Fmr1 knockout mice without having an effect in wild-type littermates (p<0.001). Previous studies reported enhanced mGluR5-dependent long-term depression (LTD) in Fmr1 knockout mice. We examined NMDA-dependent LTD by induction with low-frequency stimulation. Our data reveal a significant decrease in this form of LTD in Fmr1 knockout mice (p<0.05). Previously, we found a reduced expression of Larg in Fmr1 knockout mice. LARG is a GEF which activates RhoA, involved in actin cytoskeleton remodeling and spine morphology. We show that several genes of the LARG-RhoA pathway are underexpressed in the fragile X syndrome. DiOlistic labeling shows that spine density was significantly increased in the CA1 region of the hippocampus in Larg+/- mice (p<0.01), while spine length remained unchanged. Underexpression of LARG thus results in dendritic spine abnormalities with striking similarities to those observed in the fragile X syndrome. Our findings thus confirmed the importance of the GABAergic pathway as a powerful therapeutic target as well as unveiled novel pathways involved in key aspects of the fragile X phenotype.

2987W

Analysis of Genomic DNA sequence based on new molecular diagnostic strategies for Neurofibromatosis Type I. *S. Choi[1], J. Kim[1], J. Lee[1], J. Lee[1], G. Kim[1], B. Lee[1,2], H. Yoo[1,2].* 1) Medical Genetics Center, Asan Medical center, Seoul, South Korea; 2) Dept. Pediatrics Asan Medical Center, Children's Hospital, University of Ulsan College of Medicine, Seoul, South Korea.

Neurofibromatosis I (NF1, OMIM 162200) is characterized by multiple café au lait spots, axillary and inguinal freckling, multiple cutaneous neurofibromas, and iris Lisch nodules. NF1 is diagnosed based on the constellation of clinical findings. The genetic testing in NF1 is also important for the diagnostic confirmation. More importantly, it is essential for the appropriate genetic counseling. However, the molecular genetic testing is problematic due to the large size of the causative gene, NF1. In addition, the gene has homologous domains such as Cystein-Serine rich domain (CSD), GTPase related domain (GRD), and adenylated kinase pseudogene (AK3). These domains are dispersed in multiple chromosomal loci, leading to false positive results in routine medical sequencing tests in NF1. To avoid interferences by these homologous sequences, labor-intensive RNA-based sequencing methods have been used. To facilitate the genetic diagnostic process with high accuracy, we introduced the new strategy of genomic DNA-based molecular test for NF1. A total of eighteen patients were enrolled. All patients were diagnosed with NF1 on the basis of typical clinical features. Genomic DNAs were isolated from peripheral leukocytes. Homologous sequence domains on NF1 were amplified with the range of 6~13kb size using five NF1-specified primer sets. Nested PCRs were done in the five long-PCR products by exon-by-exon primers of the NF1 gene. The nested PCR products were sequenced using the same nested PCR primers. With this method, pseudo-variants by CSD, GRD, and AK3 regions located on exon9-to-31 of NF1 were excluded in all patients. Subsequently, we identified 18 pathogenic mutations including 11 novel variants in all NF-1 patients studied. Analytical validities of this diagnostic strategy will be evaluated in a large patient cohort with NF1.

2988T

Comprehensive Mutation Analysis for Hereditary Sensory and Autonomic Neuropathy Using a Next-Generation Sequencing System. *J. YUAN, E. Matsuura, Y. Higuchi, A. Hashiguchi, T. Nakamura, S. Nozuma, Y. Sakiyama, A. Yoshimura, H. Takashima.* Kagoshima University 8-35-1 Sakuragaoka, Kagoshima City, Kagoshima, Japan.

Objective To establish an effective diagnostic procedure and identify new molecular mechanism in patients with hereditary sensory and autonomic neuropathy (HSAN). Methods Sixteen Japanese patients with predominant sensory and/or autonomic nervous dysfunctions were enrolled. Nerve conduction studies revealed multiple sensory nerve involvement with or without mild motor nerve abnormalities. Sural nerve biopsy was performed in eight patients, and the specimens were examined by light and electron microscopy. Using a next-generation sequencing system (Illumina, MiSeq), all coding regions and flanking intron sequences of 11 known HSAN disease-causing genes and another five genes (SCN9A, CCT5, PRNP, FLVCR1, and RNF170) that may also generate sensory and autonomic symptoms were screened. Subsequently, the suspected pathogenic mutations and low-coverage domains (depth of less than 10) were confirmed by Sanger sequencing. Cases without pathogenic mutations in these genes proceeded to exome sequencing. Results The mean age of the included patients (11 males and 5 females) was 50.1 years (range, 20-80 years), and the median age at disease onset was 35 years (range, 1-71 years). Six patients (37.5%) displayed central nervous system involvement, whereas hearing loss was detected in five patients (31.3%). Pathological studies of the sural nerve revealed the loss of large and/or small myelinated nerve fibers at variable severity. Using the MiSeq system, 97.6% of target regions were covered in more than 10 reads. A frameshift mutation was discovered in SCN9A in two unrelated patients. Meanwhile, in another two patients, novel missense mutations were detected in DNMT1. Conclusions The next-generation sequencing system proved to be effective and trustworthy for genotyping the HSAN gene panel. However, we could not identify any pathogenic mutation in these 16 genes in 75% patients with the HSAN phenotype. Using exome sequencing, we can expect that more HSAN disease-causing genes will be identified.

2989F

Whole Exome Sequencing Identifies Rare Variants in Turkish Patients with Brain Malformation. *E. KARACA[1], D. PEHLIVAN[1], C. GONZAGA-JAUREGUI[1], T. GAMBIN[1], W. WISZNIEWSKI[1], A.H. CEBI[2], M.M. ATIK[1], V. GELOWANI[1], M. WITHERS[1], D. MUZNY[3], S.N. JHANGIANI[3], R.A. GIBBS[3], J.R. LUPSKI[1,4,5].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) 2.Department of Medical Genetics, Karadeniz Technical University, Faculty of Medicine, Trabzon, Turkey; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 5) Texas Children's Hospital, Houston, TX, USA.

Disorders disrupting normal brain development represent a clinically and genetically heterogeneous group. Genomic technologies are being used widely, especially in recent years, and enabling the discovery of many novel genes that are involved in brain development. However, many more are likely waiting to be brought to light. Here we report whole exome sequencing (WES) data from 25 Turkish probands from 22 families, 18 of which are consanguineous, with diagnoses of neurodevelopmental delay with brain malformation. Exome sequencing revealed deleterious mutations in candidate genes that have not been described before as mutated in brain malformation syndromes in 20 probands (80%), while the other 5 probands (20%) had mutations in known genes. Functional studies of the identified variants and genes will be performed in order to elucidate their functional impact in normal brain development. Increasing use of genomic sequencing approaches, along with other genome-wide interrogation technologies are helping to identify causative rare variants not only in disorders of brain development but also in other disorders showing Mendelian inheritance.

2990W

Investigating the genetic basis of amyotrophic lateral sclerosis using next-gen techniques. *J.A. Fifita[1], K.L. Williams[1], G.A. Nicholson[1,2], I.P. Blair[1].* 1) Macquarie University, Sydney, New South Wales, Australia; 2) Molecular Medicine Laboratory, Concord Hospital, New South Wales, Australia.

Amyotrophic lateral sclerosis (ALS) is a fatal form of motor neuron disease characterized by the progressive degeneration of both the upper and lower motor neurons. ALS may occur as familial or sporadic disease. To date, genetic mutations are the only known cause of ALS. ALS is heterogeneous with mutations in over 15 genes now described in both familial and sporadic ALS, including *SOD1*, *TARDBP*, *FUS*, *C9ORF72*, *UBQLN2*, *OPTN*, *PFN1*, *hnRNPA1* and *SS18L1*. Nevertheless, the genes are yet to be identified in 30-40% of familial ALS. We aim to identify novel genes involved in ALS using next-generation sequencing techniques. We performed exome capture (Illumina TruSeq Exome Enrichment Kit) and massively parallel sequencing (Illumina HiSeq2000) on 14 DNA samples from five ALS families that are negative for all known ALS genes. The families used in this study have DNA samples available from either two affected individuals and a 'married-in' control, or more than two affected individuals. Bioinformatic analysis of exome data identified several candidate variants among four families. Additional segregation analysis in one of these families reduced the number of potential causative variants to two. Analysis of large control cohorts is underway to validate these variants. Functional studies will then be undertaken to examine the pathological consequences of these variants in an attempt to further understand the disease mechanisms underlying both sporadic and familial ALS. Using a combination of exome sequencing and Sanger sequencing we have also determined the prevalence of known ALS genes in Australian familial and sporadic ALS cohorts. Mutations in known ALS genes account for 57.2% of familial ALS and 5% of sporadic ALS. The familial mutations comprise of *SOD1* (13.9%, n=26), *TARDBP* (2.1%, n=4), *FUS* (2.7%, n=5) *C9ORF72* (38%, n=71), *OPTN* (0.5%, n=1), *UBQLN2* (1.1%, n=2), and *SS18L1* (0.5%, n=1). Currently, the cause of 40% of ALS families, and 95% of sporadic ALS is still unknown. Additional exome sequencing projects are underway in Australian familial index patients to identify novel ALS genes.

2991T

Molecular genetics of primary microcephaly in Indian population: mutations in WDR62, ASPM and STIL genes. *A. Kumar[1], V. Bhat[1], G. Mohan[1], S.C. Girimaji[2].* 1) Molecular Reproduction, Development & Genetics, Indian Institute of Science, Bangalore, Karnataka, India; 2) Child and Adolescent Psychiatry, National Institute of Mental Health and Neuro Sciences, Bangalore, Karnataka, India.

Development of the human brain, specially the cerebral cortex, is an intricate and complex process that involves generation of neural progenitors in periventricular zones, cell proliferation through symmetric and asymmetric cell divisions, and finally the migration of post-mitotic neurons to their final destinations. An understanding of the molecular mechanisms guiding these intricate and complex processes is facilitated by discovery of causative genes for a Mendelian form of the cortical (brain) malformation disorder, primary microcephaly (MCPH), which is characterized by a smaller than normal brain and mental retardation. Interestingly, the brain volume (400 cc) of an MCPH patient is similar to early hominids, such as Australopithecus who lived 2.0-4.4 myr ago. Occasionally, MCPH patients also have other brain malformations, such as pachygyria, polymicrogyria, simplified gyral pattern etc. We have been working on the genetics of MCPH for some time, which led to discovery of the seventh MCPH gene STIL in 2009. So far, a total of nine genes and 10 loci have been identified for MCPH: MCPH1 on ch8p22-ter:MCPH1, MCPH2 on ch19q13.1-q13.2:WDR62, MCPH3 on ch9q34:CDK5RAP2, MCPH4A on ch15q21.3-q13.3:CASC5, MCPH4B on ch15q21.3-q13.3:CEP152, MCPH5 on ch1q25-q32:ASPM, MCPH6 on ch13q12.2:CENPJ, MCPH7 on ch1p32.3-p33:STIL, MCPH8 on ch4q12:CEP135, MCPH9 on ch3q22.2:CEP63 and an orphan locus MCPH10 on ch10q11.23-21.3. We have continued to ascertain MCPH families and collected so far a total of 60 families. Using homozygosity mapping and Sanger sequencing, we have identified mutations in WDR62 (3 families), ASPM (8 families) and STIL (6 families) genes only in our families, with a mutation detection rate of 28.33% (17/60 families) in known nine MCPH genes. Thus, the ASPM gene is found to be the most commonly mutated gene followed by STIL in Indian population. There are still many families in our family dataset which do not harbor mutations in any of the nine known genes, suggesting the discovery of additional novel genes for MCPH in future. We are currently using homozygosity mapping and whole-exome sequencing to identify the novel gene(s). A comprehensive analysis of our molecular genetic analysis of Indian MCPH families will be presented and discussed.

2992F

Genetic background of inner ear malformation. *R. Birkenhager, E. Prera, S. Arndt, W. Maier, R. Laszig, A. Aschendorff.* Otorhinolaryngology H&N Surg, Univ Freiburg, Freiburg, Germany.

Introduction: Hearing impairment is the most common sensory disorder in humans, affecting approximately one to three in 1000 newborns, with 50 % due to genetic causes. The majority of these cases (70 %) are non-syndromic, about 2 % of these are X-linked. So far five different X-linked loci have been mapped, the causative gene POU3F4 (MIM 300039) has been identified for the gene locus DFNX2. This hearing loss is often progressive, with temporal bone abnormalities and stapes fixation. Temporal bone anomalies include dilation of the fundus of the internal acoustic canal (IAC), that can increase the risk of CSF gusher during cochlear implant surgery. During inner ear development, the homologous mouse Pou3f4 gene is exclusively detected in the mesenchymal tissue adjacent to the otic epithelium. Knockout Pou3f4 mice have abnormalities in mesenchyme derived structures including the scala tympani, spiral limbus and spiral ligament fibrocytes of the cochlea as well as variances in the temporal bone. POU3F4 belongs to a sub-family of transcription factors, which are characterized by two conserved DNA binding domains, a POU and a HOX domain, both helix-turn-helix (HTH) structural DNA binding motifs. Several molecular genetic studies have identified mutations in the POU3F4 gene, including partial or complete deletions of the gene, as well as deletions, inversions, and duplications of the DFNX2 genomic region not encompassing the POU3F4 coding sequence. Here we present the clinical characteristics of four patients from independent families. Mutations analysis in the POU3F4 gene was brought by direct sequencing. Methods: Hearing testing BERA/Electrocochleography was only done in the Patient other family members were not tested because they have no hearing impairment or clinical abnormalities. Radiological a high-resolution CT scan was made. Mutational analysis of the affected individual and first degree relatives was performed using direct sequencing of the coding exon and intron transitions of the POU3F4, including deletion analysis. Results: Sequence analyses revealed a novel deletion of 2600 nucleotides [TAG (Stop) +21nt - 2620nt] and two novel missense mutations c.902C>T, (p.P301L) and c.973 T>A, (p.W325R) and in a fourth patient an already known missense mutation c.907C>T, (p.P303S). Conclusions: Until now three novel strongly conserved mutations in the HOX-domain of the POU3F4 transcription factor were identified leading to an effect in the protein function.

2993W

Using whole exome sequencing to diagnose primary microcephaly caused by mutations in ASPM and WDR62. *A.H. Cebi[1], D. Pehlivan[2], E. Karaca[2], M.M. Atik[2], T. Tos[3], W. Wiszniewski[2], Y. Bayram[2], D. Muzny[4], R.A. Gibbs[4], J.R. Lupski[2,5,6].* 1) Medical Genetics, Karadeniz Technical University, Faculty of Medicine, Trabzon, Turkey; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Department of Medical Genetics, Sami Ulus Children's Hospital, Ankara, Turkey; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Texas Children's Hospital, Houston, TX, USA.

Whole-exome sequencing (WES) is a new groundbreaking technology that has changed the course of human genetics. As a more efficient and cost effective method, WES is now widely used as a diagnostic tool for identifying genetic disorders. Microcephaly is defined by an occipito-frontal head circumference less than 2 standard deviations (SD) under age and gender-based averages. Mutations in the ASPM and WDR62 genes have been found to cause primary microcephaly. The ASPM gene probably plays a role in the coordination of mitotic processes as well as mitotic spindle regulation and may also have a function in regulating neurogenesis, whereas the WDR62 gene codes for a spindle pole protein that serves in cerebral cortical development. Here, we present two separately consanguineous families, each with two affected individuals with primary microcephaly. We identified homozygous frameshift deletion in ASPM gene in the affected individuals from one family and WDR62 gene in the other family, whereas parents were at heterozygous state for this mutation.

2994T

ZBTB20-dependent transcription regulation imbalance in autism and intellectual disability. *R. Koul[1,2], L. Dukes-Rimsky[1], Y. Luo[1,8], K.A. Jones[3], E. Lemyre[5], S.M. Sowell[1], D.P. Srivastava[3], S. Ladd[1], B.R. DuPont[1], C.M. Wilson[6], C. Skinner[1], F. Gurrieri[7], R.E. Stevenson[1], E. Boyd[6], J.L. Michaud[5], L. Wang[2], P. Penzes[3,4], A.K. Srivastava[1,2].* 1) J.C. Self Research Institute of Human Genetics, Greenwood Genetic Center, Greenwood, SC; 2) Department of Genetics and Biochemistry, Clemson University, Clemson, SC; 3) Department of Physiology, Northwestern University Feinberg School of Medicine, Chicago, IL; 4) Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL; 5) CHU Sainte-Justine, Centre de recherche, Montréal (QC), Canada; 6) Fullerton Genetics Center, Asheville, NC; 7) Institute of Medical Genetics, Catholic University, Rome, Italy; 8) Present address: Department of Pediatrics, Emory University School of Medicine, Atlanta, GA.

Dendritic spine morphology and dendritic arborization are key determinants of neuronal connectivity, and have been found to be altered in many neurodevelopmental disorders such as autism spectrum disorders (ASDs) and intellectual disability (ID). Several studies suggest that the changes in neuronal gene expression controlled by selective expression of transcription factors (TFs) affect the formation of dendritic spines and synapses. Recently, we identified the disruption of two transcription factor genes, *ZBTB20*, a member of the BTB/POZ (poxvirus and zinc finger) family of transcription factors, and *SOX5*, a member of the SOX (SRY-related HMG-box) family of transcription factors, by a chromosome translocation t(3;12) in a patient with developmental delay and autistic features. In addition, we identified two *SOX5* mutations, one missense and one nonsense, unique to non-syndromic ID, two nonsynonymous *ZBTB20* mutations unique to ASD and or ASD/ID, and a *de novo* intragenic deletion of *ZBTB20* in a patient with borderline intelligence, attention deficit and impulsivity. Recently, several genomic deletions involving either *ZBTB20* or *SOX5* have also been reported in patients with ASD and or ID. Studies in mice brain have shown that Zbtb20 functions as a transcriptional repressor that binds and represses several TF genes, including *SOX5,* which are involved in neurogenesis, dendritogenesis, and neuronal circuit formation. We found that in human brain, *ZBTB20* expression was largely negatively correlated with these co-expressed TF genes and genes enriched with Gene Ontology terms including synapse and synaptic transmission functions. In cultured rat cortical pyramidal neurons, overexpression of *ZBTB20* affected spine morphology and dendritic arborization. Furthermore, overexpression of *ZBTB20* in HEK293H cells resulted in a significantly reduced expression of TF genes such as *TBR1* and *MEF2C*, previously associated with ASD and ID. The ASD/ID-associated *ZBTB20* mutants had no repressor activities and caused no change in the expression of *TBR1* and *MEF2C*. Taken together, our study suggests the association of SOX5 and its transcription regulator ZBTB20 with ASD and ID and provides potential molecular and cellular mechanisms for ZBTB20 physiological actions, and suggests a potential contribution of a ZBTB20-dependent transcription regulation mechanism in neurocognitive and behavioral disorders.

2995F

Rare Deletion of the *DMD* Brain-Specific Promoter Results in ASD without Muscle Findings. *H. Mason-Suares[1], F. Probst[2], M. Hedge[1], B. Coffee[1].* 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

The *DMD* gene is expressed from eight separate promoters, three of which code for the full length protein with changes in only the first exon. One of these promoters, known as Dp427c or Dp427b, is expressed predominantly in neurons of the cortex and the CA regions of the hippocampus. This brain-specific isoform is identical to the muscle-specific isoform of *DMD* except that the first 11 amino acids of the muscle-specific isoform are replaced with the three amino acids MED. While the function of the brain-specific isoform is unknown, it has been hypothesized that it may be responsible for the intellectual disability defect seen in about one-third of Duchenne muscular dystrophy patients. Supporting this, the brain-specific isoform is associated with the postsynaptic membrane and thought to play a role in synapse structure or function. Here we present the first case of a deletion that only removes the brain-specific promoter, leaving the rest of the *DMD* gene intact including the muscle-specific promoter, in a pair of male twins. The twins presented with developmental delay and an autism spectrum disorder including speech delay and repetitive behaviors. One of the twins also had intellectual disability and intractable epilepsy, which was resolved via left temporal lobectomy. Neither of the twins have had any muscle findings nor a positive Gower's sign consistent with an intact muscle-specific *DMD* transcript. This case suggests that the brain-specific *DMD* transcript is necessary for normal brain function, but does not result in muscular dystrophy.

2996W

Mitochondrial DNA deletions heteroplasmy analysis using the Next Generation Sequencing. *Y. Okamoto, J. Yuan, A. Yoshimura, S. Nozuma, Y. Higuchi, Y. Sakiyama, A. Hashiguchi, I. Higuchi, H. Takashima.* Department of Neurology and Geriatrics, Kagoshima University, Kagoshima, Japan.

Mitochondrial disorders are genetically heterogeneous group of diseases, involving two genomes, the 16.6 kb mitochondrial genome and approximately 1500 genes encoded in the nuclear genome. Mitochondrial myopathy with episodic hyper-creatine kinase-emia (MIMECK) is characterized by episodic or persistent muscle weakness and elevated serum CK levels triggered by infections, drugs, or stressful situations. We previously reported a similar phenotype with same mitochondrial DNA (mtDNA) alterations. However, the pathophysiological mechanism of MIMECK is still unknown. The development of Next Generation Sequencing (NGS) has revolutionized the diagnostic approach. New approaches make genetic analyses much faster, more sensitive and more efficient. The MiSeq personal sequencing system was performed on 30 samples and revealed mtDNA deletions in muscles samples of genetically proven cases, clinically and pathologically suspected cases and MIMECK. We found low rate mtDNA deletions in MIMECK and also found various rate mtDNA deletions heteroplasmy in other mitochondrilal disease including MELAS with 3243A>G mutation. Further study we need to perform analysis for nuclear gene or pathological study to detect pathophysiology. Sanger sequencing does not have the sensitivity to detect heteroplasmic mutations below about 20%. It also does not detect large deletions. Southern blot analysis is not sensitive to detect low levels of heteroplasmic deletions. Low rate mtDNA deletions analysis by NGS could be more sensitive tool before exome-sequence for nuclear mitochondrial related gene or underdiagnosed mitochondrial disease.

2997T

Severe Congenital *RYR1* Associated Myopathy: AR and AD *RYR1* Mutations that Expand the Genetic, Clinical and Pathologic Spectrum. *L. Medne[1], D.X. Bharucha-Goebel[1], M. Santi[2], K. Zukosky[3], J. Dastgir[3], T. Winder[4], P.B. Shieh[5], G. Tennekoon[1], R.S. Finkel[6], J.J. Dowling[7], N. Monnier[8], C.G. Bönnemann[3].* 1) Div of Neurology, CHOP, Philadelphia, PA. 19104; 2) Dept of Pathology & Laboratory Medicine, CHOP, Philadelphia, PA 19104; 3) Neurogenetics Branch, NINDA, NIH, Bethesda, MD; 4) Prevention Genetics, Marshfield, WI; 5) Dept of Neurology, University of California, Los Angeles, CA; 6) Division of Neurology, Nemours Children's Hospital, Orlando, FL; 7) Department of Pediatrics, University of Michigan Medical Center, Ann Arbor, MI; 8) Biochimie et Genetique Moleculaire, Institut de Biologie et Pathologie, CHU, Grenoble, France.

RYR1 gene mutations are known to cause central core or minicore myopathy, typically with a static or slowly progressive course, and pharmacogenetic predisposition to malignant hyperthermia (MH). We report a series of 13 patients with early neonatal onset *RYR1* associated myopathy due to both autosomal recessive and dominant acting *RYR1* mutations representing the severe end of the clinical spectrum, thus expanding the genetic, clinical and histological spectrum of *RYR1* associated myopathy. Clinical features associated with the severe neonatal presentation of *RYR1* associated myopathy included decreased fetal movement, hypotonia, poor feeding, respiratory involvement, arthrogryposis, ophthalmoplegia in four patients, and femur fractures or hip dislocation at birth. One patient had a cleft palate, and another a congenital rigid spine phenotype. Muscle biopsy was performed in all patients at ages ranging from 3 weeks of life to 4 years old. Marked type I fiber predominance approaching type I uniformity was common. All dominant cases had cores while recessive cases had extremely diverse pathology including central cores, multi minicores, more subtle unevenness or normal staining on NADH. Centralized nuclei, extensive fatty and connective tissue infiltration, extremely small fibers in the absence of active degeneration and regeneration were also seen as part of this spectrum. Dominant *RYR1* mutations (*de novo* or due to mosaic parent) were present in 4 patients and 9 patients had recessive *RYR1* mutations. Two of the carrier parents showed muscle weakness and one had ophthalmoplegia, thus showing that those mutations can be dominant acting. This series confirms and expands the clinical and histological variability associated with severe congenital *RYR1* associated myopathy. Both dominant and recessive mutations of the *RYR1* gene can result in a severe neonatal onset phenotype, but more clinical and histological heterogeneity has been seen in those with recessive RYR1 gene mutations. Absence of cores on muscle biopsy should not preclude the consideration of *RYR1* associated myopathy. Furthermore, identification of predominantly recessive *RYR1* mutations with severe neonatal phenotype alters recurrence risk counseling for parents and their own medical management suggesting the need for life-long MH precautions in heterozygous carrier parents and siblings given the AD nature of isolated MH predisposition mutations.

2998F

Subtle mutations in the SMN1 gene in Chinese patients with SMA: p.Arg288Met mutation causing SMN1 transcript exclusion of exon7. *Y. QU[1], J. DU[1], E. LI[2], J. BAI[1], Y. JIN[1], H. WANG[1], F. SONG[1].* 1) Capital Institute of Pediatrics, Beijing, China; 2) Children's Hospital Affiliated Capital Institute of Pediatrics, Beijing, China.

Background: Proximal spinal muscular atrophy (SMA) is a common neuromuscular disorder resulting in death during childhood. Around 81 ~ 95% of SMA cases are a result of homozygous deletions of survival motor neuron gene 1 (SMN1) gene or gene conversions from SMN1 to SMN2. Less than 5% of cases showed rare subtle mutations in SMN1. Our aim was to identify subtle mutations in Chinese SMA patients carrying a single SMN1 copy. Methods: We examined 14 patients from 13 unrelated families. Multiplex ligation-dependent probe amplification analysis was carried out to determine the copy numbers of SMN1 and SMN2. Reverse transcription polymerase chain reaction (RT-PCR) and clone sequencing were used to detect subtle mutations in SMN1. SMN transcript levels were determined using quantitative RT-PCR. Results: Six subtle mutations (p.Ser8LysfsX23, p.Glu134Lys, p.Leu228X, p.Ser230Leu, p.Tyr277Cys, and p.Arg288Met) were identified in 12 patients. The p.Tyr277Cys mutation has not been reported previously. The p.Ser8LysfsX23, p.Leu228X, and p.Tyr277Cys mutations have only been reported in Chinese SMA patients and the first two mutations seem to be the common ones. Levels of full length SMN1 (fl-SMN1) transcripts were very low in patients carrying p.Ser8LysfsX23, p.Leu228X or p.Arg288-Met compared with healthy carriers. In patients carrying p.Glu134Lys or p.Ser230Leu, levels of fl-SMN1 transcripts were reduced but not significant. The SMN1 transcript almost skipped exon 7 entirely in patients with the p.Arg288Met mutation. Conclusions: Our study reveals a distinct spectrum of subtle mutations in SMN1 of Chinese SMA patients from that of other ethnicities. The p.Arg288Met missense mutation possibly influences the correct splicing of exon 7 in SMN1. Mutation analysis of the SMN1 gene in Chinese patients may contribute to the identification of potential ethnic differences and enrich the SMN1 subtle mutation database.

2999W

Exome sequencing identifies PINCH2 mutations associated with early-onset autosomal recessive LGMD with severe cardiomyopathy and triangular tongue. *J. Warman Chardon[1], A. Smith[2], J. Woulfe[3], K. Rakhra[4], C. Dennie[4], J. Schwartzentruber[5], C. Beaulieu[2], FORGE. Canada Consortium[6], J. Majewski[7], D.E. Bulman[2], K.M. Boycott[1,2], D. Dyment[1,2].* 1) Genetics, Chidren's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 3) Pathology, The Ottawa Hospital, Ottawa, Ontario, Canada; 4) Medical Imaging, The Ottawa Hospital, Ottawa, Ontario, Canada; 5) McGill University and Genome Quebec Innovation Centre, Montréal, Québec; 6) Canada; 7) Department of Human Genetics, McGill University, Montreal, Quebec.

Limb Girdle Muscular Dystrophy (LGMD) is a heterogeneous group of inherited disorders leading to progressive muscle degeneration often associated with cardiac complications. We investigated affected siblings with a childhood onset LGMD with macroglossia and calf enlargement. The siblings developed decreased ejection fraction with global left ventricular dysfunction in their 3rd decade, severe quadriparesis and relative sparing of the face, with a broad based triangular tongue. After negative genetic evaluation for known LGMD genes, we performed whole exome sequencing of the affected siblings and identified shared compound heterozygous missense mutations in exon 5 (c.C356T; p.Pro119Leu and c.C342G; p.Asn114Lys) and exon 11 (c.T1034C; p.Leu345Pro) of the PINCH2 gene, which segregated appropriately in the family. Both novel variants in exon 5 are in cis. PINCH2 is an important member of the IPP (ILK, Parvin, PINCH) heterotrimeric complex and is essential for signaling through integrin adhesion receptors that regulate cell migration, spreading and adhesion and is found in skeletal and cardiac muscle cells. The IPP complex stabilizes expression of the component proteins by reducing proteasomal degradation. PINCH1 and PINCH2 demonstrate overlapping expression, competitive binding to ILK and functional redundancy. There is evidence in murine models that PINCH1 may compensate for loss of PINCH2 in some tissues, suggesting that a potential therapeutic approach may involve upregulation of PINCH1 to stabilize the IPP complex in patients with PINCH2 mutations. Although well demonstrated in animal models, this is a novel discovery of IPP complex dysfunction implicated in LGMD.

3000T

C19ORF12 Mutations in Neurodegeneration with Brain Iron Accumulation (NBIA). *G. Annesi[1], P. Tarantino[1], M. Gagliardi[1], G. Lesca[2,3], E. Broussolle[4,5], A. Gambardella[1], A. Quattrone[6].* 1) Inst. of Neurological Science, National Research Council, Mangone (CS), CS, Italy; 2) Department of Medical Genetics, Hospice Civiles de Lyon and ClaudeBernard Lyon, University Lyon France, Lyon, France,; 3) CRNL, CNRS UMR 5292, INSERM U1028, Lyon, France, Lyon, France; 4) Université Lyon I; Centre de Neurosciences Cognitives, CNRS UMR5229, Lyon, France,; 5) Service de Neurologie C, Hôpital Neurologique Pierre Wertheimer, Lyon, France, Lyon, France,; 6) Neuroimaging Research Unit, National Research Council, Catanzaro, Italy, Catanzaro, Italy.

Neurodegeneration with Brain Iron Accumulation (NBIA) comprise heterogeneous group of progressive neurodegenerative disorders that present with a progressive extrapyramidal syndrome and excessive iron deposition in the brain, particularly affecting the basal ganglia, mainly the globus pallidus. Genetic defects in PANK2 gene are the most common cause of NBIA, followed by mutations in PLA2G6, few reported NBIA families are known to carry mutations in FA2H, a gene previously associated with familial leukodystrophy and spastic paraparesis. Mutations within C19orf12 have recently been identified in patients with NBIA. This gene C19orf12, codes for a mitochondrial membrane protein and the acronym MPAN (mitochondrial membrane protein-associated neurodegeneration). In this study we report the clinical description of five patients from five families with NBIA and the subsequent molecular genetic investigation. Five patients diagnosed with NBIA by movement disorders specialists. As required by clinical regulations, all provided written informed consent for genetic testing.The initial diagnosis was based on manifestation of progressive movement disorder and demonstration of a hypointense signal in the globus pallidus on T2-weighted MRI scans. The exons of PANK2, PLA2G6, FA2H and C19orf12 were amplified by PCR and sequenced. Results.Sequencing of PANK2, PLA2G6 and FA2H was normal; whereas the molecular analysis of C19orf12 gene revealed a novel homozygous and heterozygous C19orf12 mutations in our patients with NBIA. Further studies are needed to explore the function of C19orf12 in NBIA, and extended genetic analysis of larger patients cohorts will provide more information about the frequency of this disease.

3001F

Hypoxanthine guanine phosphoribosyltransferase (HPRT) deficiencies: *HPRT1* mutations in new Japanese families and PRPP concentration. *Y. Yamada[1], N. Nomura[1], A. Yamano[1], K. Yamada[1], R. Kimura[1], D. Fukushi[1], H. Hasegawa[2], M. Nakamura[2], K. Ichida[2], N. Wakamatsu[1].* 1) Dept Gen, Inst Dev Res, Aichi Human Service Ctr, Kasugai Aichi, Japan; 2) Dept Pathophysiol, Sch Parma, Tokyo Univ Pharm and Life Sci, Tokyo, Japan.

Mutation of hypoxanthine guanine phosphoribosyltransferase (HPRT), gives rise to Lesch-Nyhan syndrome, which is characterized by hyperuricemia, severe motor disability, and self-injurious behavior, or HPRT-related gout (Kelley-Seegmiller syndrome). The marked heterogeneity of HPRT deficiency is well known, as reported more than 500 mutations at the HPRT gene (*HPRT1*) locus. We had identified a number of *HPRT1*mutations in Asian patients manifesting different clinical phenotypes. In this study, we clarified the mutations of four new Japanese families with the HPRT deficiencies and observed the increase of erythrocyte PRPP concentrations in the deficiencies. A new mutation of G to TT (c.456delinsTT) resulting in a frame shift (p.152fs154X) in exon 3 has been identified from a Lesch-Nyhan family. In the other Lesch-Nyhan family, a point mutation on intron 7 (c.532+5G>T) causing splicing error (exon 7 excluded) was detected. In two partial deficiency cases with hyperuricemia, two missense mutations of p.D20V (c.59A>T) and p.H60R (c.179A>G) were found, respectively. The mutations of c.532+5G>T and p.D20V were detected firstly in Japanese patients but had been reported in European families. The p.H60R was reported previously as an asymptomatic missense substitution. The PRPP concentrations in erythrocyte from three Lesch-Nyhan patients (18.5; 33.4; 40.8 nmol/mL) were markedly increased compared with those from normal control (0.8 ± 0.3 nmol/mL, n=8). The PRPP concentrations from patients with partial deficiency (1.3 - 7.7 nmol/mL) were also increased, but the degrees were smaller than those of the complete deficiency. The increases of erythrocyte PRPP concentration were correlated approximately with the phenotypic severity. Furthermore, we summarize the spectrum of Asian HPRT mutations.

3002W

PLEKHG5 deficiency leads to an intermediate form of autosomal recessive Charcot-Marie-Tooth disease. *H. Azzedine[1], P. Zavadakova[1], V. Planté-Bordeneuve[2], M. Vaz Pato[3], N. Pinto[3], L. Bartesaghi[1], J. Zenker[1], O. Poirot[1], N. Bernard-Marissal[1], E. Arnaud[1], R. Cartoni[1], A. Title[1], G. Venturini[1], J.J. Médard[1], E. Makowski[4], L. Schöls[5], K.G. Claeys[6,7], C. Stendel[7], A. Roos[6], J. Weis[6], O. Dubourg[8], J.L. Loureiro[9], G. Stevanin[10, 11], G. Said[12], A. Amato[13], J. Baraban[4], E. LeGuern[10], J. Senderek[14], C. Rivolta[1], R. Chrast[1].* 1) Medical genetics, DGM Faculty of biology and medicine, Unil, Lausanne, Switzerland; 2) Service de Neurologie, CHU Henri Mondor, 94000 Créteil, France; 3) CICS- Health Science Research Centre, Universidade da Beira Interior, 6200-506 Covilhã, Portugal; 4) Solomon H. Snyder Department of Neuroscience, John Hopkins University, Baltimore, MD 21205, USA; 5) Department of Neurodegenerative Disease, Hertie-Institute for Clinical Brain Research and Center for Neurology, 72076 Tübingen, Germany; 6) Institute of Neuropathology, University Hospital RWTH Aachen, 52074 Aachen, Germany; 7) Department of Neurology, University Hospital RWTH Aachen, 52074 Aachen, Germany; 8) Centre de reference des maladies neuromusculaires, Assistance Publique-Hôpitaux de Paris, CHU Pitié-Salpêtrière, 75013 Paris, France; 9) UnIGENe and Centro de Genética Preditiva e Preventiva, Institute for Molecular and Cellular Biology, 4050 Porto and Serviço de Neurologia, Centro Hospitalar entre Douro e Vouga, 4520-211 Santa Maria da Feira, Portugal; 10) Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (Inserm/UPMC 975 ; CNRS 7225) and Departement of Genetics, Assistance Publique-Hôpitaux de Paris, CHU Pitié-Salpêtrière, 75013 Paris, France; 11) Neurogenetics Group, Ecole Pratique des Hautes Etudes, CHU Pitié-Salpêtrière, 75013 Paris, France; 12) Department of Neurology, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris and University Pierre et Marie Curie, 75013 Paris, France; 13) Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; 14) Department of Neurology, Friedrich-Baur-Institute, Ludwig-Maximilian-University, 81377 Munich, Germany.

Charcot-Marie-Tooth disease (CMT) comprises a clinically and genetically heterogeneous group of peripheral neuropathies characterized by progressive distal muscle weakness and atrophy, foot deformities, and distal sensory loss. Following the analysis of two consanguineous families affected by a childhood to late-onset recessive form of intermediate CMT, we identified overlapping regions of homozygosity on chromosome 1p36 with a combined maximum LOD score of 5.4. Molecular investigation of the genes from this region allowed identification of two homozygous mutations in PLEKHG5 that produce premature stop codons and are predicted to result in functional null alleles. Analysis of Plekhg5 in the mouse revealed that this gene is expressed in neurons and glial cells of the peripheral nervous system, and that knockout mice display reduced nerve conduction velocities that are comparable to those of affected individuals from both families. Interestingly, a homozygous PLEKHG5 missense mutation was previously reported in a recessive form of severe childhood onset lower motor neuron disease (LMND) leading to loss of the ability to walk and need for respiratory assistance. Together, these observations indicate that different mutations in PLEKHG5 lead to clinically diverse outcomes (intermediate CMT or LMND) affecting the function of neurons and glial cells.

3003T

Characterization of a homologue of the Batten disease protein CLN3 in the model eukaryote Dictyostelium. *R.J. Huber, S.L. Cotman, M.A. Myre.* Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA.

Batten disease refers to a group of mostly childhood neurodegenerative disorders also known as the neuronal ceroid lipofuscinoses (NCL). NCL disorders characteristically display dysregulated lysosomal function and an excessive accumulation of lipofuscin in neurons and other cell types. Juvenile NCL is the most common subtype of Batten disease and results from recessive, loss-of-function mutations in the CLN3 gene. Clinical symptoms of the disease include the progressive loss of motor function and psychological ability, blindness, seizures, and a reduced lifespan. Despite substantial research efforts using a variety of systems, the precise function of CLN3 remains unknown. The social amoebozoan Dictyostelium discoideum is used as a model system for studying a variety of cell and developmental processes and has been successfully used in biomedical research for studying a number of human diseases, including neurodegenerative disorders (e.g., Alzheimer's, Parkinson's, and Huntington's). To better understand the function of CLN3 in humans, we knocked out the CLN3 homologue in Dictyostelium and assayed the mutant cells for obvious and subtle phenotypic abnormalities during Dictyostelium growth and development. CLN3- cells proliferate at an enhanced rate and are able to complete the developmental cycle. However, phenotypes are observed during early, mid-, and late Dictyostelium development suggesting an involvement of CLN3 in cell movement, aggregation, and differentiation. Our current work is focused on expressing human CLN3 in Dictyostelium to see whether its expression can rescue the phenotypes we observe in CLN3- cells. Together, our work has generated a new cellular model for studying CLN3 function, which will provide new insight into the function of CLN3 in normal and diseased human cells. Our work also further validates the use of Dictyostelium for studying the function of proteins linked to neurodegeneration.

3004F

Whole exome sequencing identifies a new gene for Charcot-Marie-Tooth Type 2 in a Polish family. *M.L. Kennerson[1,2,3], G. Pérez-Siles[1], A. Kochański[4], A. Kidambi[1], A.P. Drew[1], J. Kosińska[5], M. Gonzalez[7], D. Kabzińska[4], R. Ploski[5], M. Menezes[6], I. Hausmanowa-Petrusewicz[4], S. Zuchner[7], G.A. Nicholson[1,2,3].* 1) Northcott Neuroscience, ANZAC Res Inst, Concord, Australia; 2) Molecular Medicine Laboratory, Concord Hospital, Sydney, Australia; 3) Sydney Medical School, University of Sydney, Sydney, Australia; 4) Neuromuscular Unit, Mossakowski Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland; 5) Department of Medical Genetics, Warsaw Medical University, Warsaw Poland; 6) Kids Research Institute, The Childrens Hospital, Westmead, Australia; 7) Hussman Institute for Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA.

Charcot Marie Tooth (CMT) disease is the most common inherited peripheral neuropathy affecting 1 in 2500 people. The disorder affects both motor and sensory neurons of the peripheral nervous system resulting in the 'dying back' (axonal degeneration) of long nerves. CMT is both clinically and genetically heterogeneous and is traditionally divided into demyelinating (CMT1) and axonal (CMT2) forms based on electrophysiological and neuropathalogical criteria. For CMT type 2 up to 80% of the causative genes remain unidentified. CMT research in the last 20 years has made significant discoveries that are elucidating the complexities of motor and sensory nerve biology and the pathogenic process of axonal degeneration. The advancement of sequencing technologies is providing affordable tools to identify gene mutations in families too small for traditional positional cloning approaches. The strategy of whole exome sequencing (WES) is therefore a very feasible approach for gene discovery in smaller CMT families. We have previously reported a Polish family with CMT type 2 in which all known CMT2 genes have been excluded (1). The family is a three generation kindred with affected individuals in two generations. We have performed WES on four members of the family (3 affected and one normal). Bioinformatic analysis and variant filtering identified a non-synonymous exonic variant in a member of the tubulin gene family. The variant fully segregates with the affected phenotype and has been excluded in 1400 normal control chromosomes. The variant occurs at a highly conserved amino acid residue of the protein and is predicted to be pathogenic. By querying 164 CMT2 index exomes in the GEMapp database (University of Miami, Miller School of Medicine URL: https://genomics.med.miami.edu/gem.app) we have identified an additional family with a different variant in the same gene. Transient transfection experiments with wild type and mutant expression constructs have shown alteration of the cytoskeleton. Genetic validation and functional studies supporting pathogenicity of the variant will be presented. This study demonstrates the power of WES for gene discovery in small families, further genetic heterogeneity of CMT type 2 and highlights a pathogenic mechanism altering cellular scaffolding. (1) Kochański A., Kennerson M. et al. 2005 Neurology 64:533-535.

3005W

Distributions of Degenerative Myelopathy Associated SOD1 Alleles among Privately Owned Dogs. *A. Kolicheski[1], R. Zeng[1], G.S. Johnson[1], M.L. Katz[2], T. Mhlanga-Mutangadura[1], L. Hansen[1], E. Ivansson[3], K. Lindblad-Toh[3,4], J. Guo[1], D.P. O'Brien[5], G.C. Johnson[5], J.R. Coates[5].* 1) Department of Veterinary Pathobiology, University of Missouri College of Veterinary Medicine, Columbia, MO; 2) Mason Eye Institute, University of Missouri School of Medicine, Columbia, MO, USA; 3) Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; 4) Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA; 5) Department of Veterinary Medicine and Surgery, University of Missouri College of Veterinary Medicine, Columbia, MO, USA.

Canine degenerative myelopathy (DM) is a common neurodegenerative disease in dogs 8 years-old or older. Similar to human amyotrophic lateral sclerosis patients (ALS) patients, dogs with DM show both upper and lower motor neuron signs at end stage. DM-affected dogs may prove useful as a naturally-occurring disease model to investigate ALS pathogenic mechanisms and to evaluate the efficacy of therapeutic interventions. Two SOD1 mutations, SOD1:c.118G>A and SOD1:c.52A>T, are associated with DM. We here report the distribution of alleles at SOD1:c.118 and/or SOD1:c.52 in 33,747 privately owned dogs including cross-bred dogs and representatives from 222 breeds. Members of 124 different dog breeds harbored the SOD1:c.118A allele; whereas, the SOD1:c.52T allele was found exclusively in Bernese Mountain Dogs. Among the 914 genotyped Bernese Mountain Dogs, the SOD1:c.52T allele frequency was only 3.5%, much lower than the 37% frequency of the SOD1:c.118A allele in this breed. Twenty-four Bernese Mountain Dogs were compound heterozygotes, inheriting the SOD1:c.52T allele from one parent and the SOD1:c.118A allele from the other. We ascertained the clinical status of six 8 years-old or older compound heterozygotes Bernese Mountain Dogs. Four of them developed clinical signs consistent with DM. DM was histopathologically confirmed in eight SOD1:c.118A/G heterozygotes from 5 breeds. These eight dogs had no other SOD1 sequence variants, suggesting that heterozygosity at SOD1:c.118A/G increases the risk of developing DM. To estimate the relative risk of developing DM for SOD1:c.118A/G heterozygotes versus SOD1:c.118A homozygotes or SOD1:c.118G homozygotes, we reviewed clinical histories of 104 dogs that were under 8 years old and without signs of DM when their samples were submitted for genotyping but are currently over 10 years old. Sixteen of 26 SOD1:c.118A homozygotes (62%) exhibited DM-like signs; whereas, none of the 40 SOD1:c.118A/G heterozygous dogs and none of the 38 SOD1:c.118G homozygous dogs developed these clinical signs. Because the SOD1:c.118A allele is so common among privately owned dogs, it should be possible to recruit a cohort of homozygotes to test the efficacy of therapeutic interventions.

3006T

Homozygous null mutations of *FIG4* in Yunis-Varón syndrome. *G.M. Lenk[1], P.M. Campeau[2], J.T. Lu[3,4], Y. Bae[2], L. Burrage[2], P. Turnpenny[5], J.R. Corona-Rivera[6,7], L. Morandi[8], M. Mora[8], H. Reutter[9], A.T. Vulto-van Silfhout[10], L. Olivier-Faivre[11,12], E. Haan[13], R.A. Gibbs[3], M.H. Meisler[1], B.H. Lee[2,14].* 1) Dept Hum Genet, Univ Mich, Ann Arbor; 2) Dept Molec Hum Genet, Baylor Coll Med, Houston; 3) Hum Gen Seq Cntr, Baylor; 4) Dept Str Comput Biol Molec Biophys, Baylor; 5) Clin Genet Dept, Royal Devon & Exeter Hosp, Exeter, UK; 6) Div Ped New Civil Hosp, Guadalajara; 7) Instit Hum Genet, Centro Univ Ciencias Salud, Univ Guadalajara; 8) Neuromusc Dis Neuroimm, Carlo Besta Inst Neurol - IRCCS, Milan; 9) Dept Neonat, Child Hosp, Univ Bonn; 10) Dept Hum Genet, Radboud Univ, Nijmegen Med Cntr; 11) Cent Génét, Hôspt d'Enfants, Dijon; 12) Equipe GAD EA4271, Univ Bourgogne, Dijon; 13) S. Austral Clin Genet Serv, Women & Child Hosp, N Adelaide; 14) HHMI, Houston.

The *FIG4* gene encodes a phospholipid phosphatase that removes the 5-phosphate from the signaling lipid $PI(3,5)P_2$, a regulator of late endosome/lysosome vesicle trafficking. Deficiency of $PI(3,5)P_2$ in *Fig4* null mice results in enlarged cytoplasmic vacuoles, autophagy defects and neurodegeneration (Ferguson et al, HMG 2012). Hypomorphic mutations of human *FIG4* result in the peripheral neuropathy Charcot-Marie-Tooth type 4J (Nicholson et al, Brain 2012). Yunis-Varón syndrome (OMIM 216340) is an autosomal-recessive disorder with neurological and bone involvement including cleidocranial dysplasia and digital anomalies, as well as enlarged vacuoles in neurons, muscle, and cartilage. Using whole-exome and Sanger sequencing, we identified frameshift and missense mutations of *FIG4* in affected individuals from three unrelated families with Yunis-Varon Syndrome (Campeau et al, AJHG 2013). The *FIG4* genotypes of affected individuals included homozygosity for protein truncation alleles in one family, with heterozygosity for missense and null mutations or homozygosity for missense mutations in the other two families. In a functional assay, the missense substitutions from these families failed to correct the vacuolar defect of *Fig4* null mouse fibroblasts, demonstrating loss of function. We identified previously unrecognized skeletal defects in *Fig4* null mice, including small skeletons, reduced trabecular bone volume and cortical thickness. Comparison of patients with Yunis-Varon Syndrome and CMT4J indicates that complete absence of functional FIG4 results in central and peripheral nervous system dysfunction as well as skeletal anomalies, while retention of as little as 0.2% of normal FIG4 activity protects the CNS and causes peripheral nerve impairment only. Neurological findings in null individuals of both species include degeneration with vacuolization of the cerebral cortex and cerebellar and thalamic nuclei, as well as hypoplasia of the corpus callosum. Our findings demonstrate that null mutations of *FIG4* are responsible for a subset of Yunis-Varon Syndrome, the most severe known human disorder caused by defective phosphoinositide metabolism, and demonstrate a previously unrecognized role for $PI(3,5)P_2$ signaling in skeletal development. Studies in progress include genomic analysis of three Yunis-Varon families lacking *FIG4* mutations, and further investigation of defective bone development in homozygous null mice.

3007F

SLC25A12 homozygous missense mutation reduces neuronal AGC1 activity to cause global developmental delay, intractable epilepsy, and reduced N-acetylaspartate in consanguineous.
D. Li[1,9], M.J. Falk[2,3,9,10], X. Gai[4], E. McCormick[2,3], E. Place[2,3], F.M. Lasorsa[5], F.G. Otieno[1], C. Hou[1], C.E. Kim[1], N. Abdel-Magid[1], L. Vazquez[1], F.D. Mentch[1], R. Chiavacci[1], G. Giannuzzi[6], E.D. Marsh[7], Y. Guo[1], L. Tian[1], F. Palmieri[5,6,10], H. Hakonarson[1,2,8,10]. 1) The Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA; 2) Division of Human Genetics Department of Pediatrics, The Children's Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 3) Division of Child Development and Metabolic Disease, Department of Pediatrics, The Children's Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 4) Department of Molecular Pharmacology and Therapeutics, Loyola University Stritch School of Medicine, Mayfield, IL 60153, USA; 5) CNR Institute of Biomembranes and Bioenergetics, 70125 Bari, Italy; 6) Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, 70125 Bari, Italy; 7) Division of Pediatric Neurology, Departments of Pediatrics and Neurology, The Children's Hospital of Philadelphia and The Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, USA; 8) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 9) Equal contribution; 10) Corresponding authors.

Whole exome sequencing analysis identified a novel homozygous c.1058G>A (p.Arg353Glu) missense mutation in *SLC25A12* that segregated with disease in a consanguineous Indian kindred in which a brother and sister were similarly affected with severely delayed global development, congenital hypotonia, mixed epilepsy, delayed myelination, and notable reduction of the N-acetylaspartate peak on brain spectroscopy. Extensive clinical diagnostic genetic testing was unrevealing, although both affected siblings shared three large homozygous chromosomal regions that were identified by genome-wide microarray analysis. Research-based whole exome sequencing (WES) analysis identified two rare homozygous variants that were predicted to be disease-causing, with only the *SLC25A12* mutation found by Sanger validation to segregate with disease in this family. *SLC25A12* encodes the neuronal aspartate-glutamate carrier 1 (AGC1) protein, which is an essential component of the malate-aspartate shuttle that transfers the reducing equivalents of NADH and H+ from cytosol to mitochondria in neurons. Functional studies demonstrated that the activity of recombinant mutant (Arg353Glu) AGC1 protein from the proband's fibroblasts was reduced to fifteen percent of the wild-type rate when measuring either aspartate or glutamate transport. This is the second reported family in which AGC1 has been shown to be causal of a severe, infantile-onset neurologic disease. An ACG1 mutation was identified in the initial case by candidate gene sequencing with fibroblast validation of absent AGC1 activity following identification of severely reduced brain levels of N-acetylaspartate and hypomyelination. Indeed, AGC1 activity enables neuronal export of aspartate, which is the glial substrate necessary for proper neuronal myelination. In summary, these data confirm that autosomal recessive *SLC25A12* disease impairs neuronal AGC1 activity and highlight the importance of considering *SLC25A12* in the differential diagnosis of infantile-onset seizures, hypotonia, and global developmental delay with abnormal myelination and reduced N-acetylaspartate levels in brain.

3008W

Defect of TARG1/C6orf130 gene causing an autosomal form of familial neurodegeneration with seizure.
E. Ozkan[1], B. Chioza[1], M. Patton[1], A.G. Ladurner[2], G. Timinszky[2], I. Ahel[3], R. Sharifi[1]. 1) BMS Genetics, St.George's University of London, London, United Kingdom; 2) Faculty of Medicine, Butenandt Institute of Physiological Chemistry, Ludwig Maximilians University of Munich, Munich, Germany; 3) Cancer Research UK, Paterson Institute for Cancer Research, University of Manchester, Manchester, UK.

The genetic causes of neurodegenerative disorders, a heterogeneous group of chronic progressive diseases, remain largely unknown. Here, we report the discovery of TARG1/c6orf130 gene (NM_145063.2) causing severe neurodegeneration with an autosomal recessive trait. The clinical presentations are; progressive global neural dysfunction and severe neurological development delay, generalized tonic- clonic (grandmal) epilepsy, tetraplegia, absence of tendon reflexes, communication disability, and lack of swallowing reflex. We genotyped the family members on genome-wide SNP microarrays and used the data to determine a single 20 cM homozygosity-by-descent (HBD) locus in region 6p21 (LOD score of 7.4); with performing next generation sequencing we identified the missense change c.227C>T that causes an early stop codon (NP_659500.1:p.R76X) at the conserved C-terminus within TARG1/C6orf130. This gene encodes C6orf130 protein belongs to the macrodomains family which are evolutionarily conserved structural protein involving in adenosine diphosphate (ADP)-ribosylation. Adenosine diphosphate (ADP)-ribosylation is an evolutionarily conserved reversible post-translational protein modification that regulates a wide range of cellular processes, including DNA repair, transcription, telomere dynamics, cell differentiation and proliferation, the inflammatory and immune responses and apoptosis. We identified C6orf130 as a PARP (poly(ADP-ribose) (PAR) polymerase)-interacting protein that removes mono(ADP-ribosyl)ation on glutamate amino acid residues in PARP-modified proteins. Our biochemical analysis of C6orf130 suggests a mechanism of catalytic reversal involving a transient C6orf130 lysyl-(ADP-ribose) intermediate. In human neuron cells, we show that C6orf130 localise to the nucleus. Furthermore, depletion of C6orf130 protein in cells leads to proliferation and DNA repair defects. Collectively, our data suggest that C6orf130 enzymatic activity has a role in the turnover and recycling of protein ADP-ribosylation, and we have implicated the importance of this protein in supporting normal neuron cell function in humans. This is for the first time that we show the importance of C6orf130 protein (newly named Terminal ADP-Ribose protein Glycohydrolase; TARG1) in neuron cells by reversing protein ADP-ribosylation and protein modulation.

3009T

Application of high-throughput sequencing to pediatric patients with unresolved diagnoses. *R.J. Taft[1], C. Simons[1], A. Vanderver[2], R.J. Leventer[3,4], M.S. van der Knaap[5], N.I. Wolf[5], R. Schiffmann[6], S. Damiani[7], P. Pearl[1], M. Bloom[8], S.M. Grimmond[9], D. Miller[9], D.R. Thorburn[10], J. Christodoulou[10], M. Gabbet[11], J. McGaughran[12].* 1) Institute for Molecular Bioscience, University of Queensland, St Lucia, Australia; 2) Department of Neurology, Children's National Medical Center, Washington, DC, USA; 3) Department of Neurology, Royal Children's Hospital, Melbourne, Victoria, Australia; 4) Murdoch Children's Research Institute, Melbourne, Victoria, Australia; 5) Department of Child Neurology, VU University Medical Center, Amsterdam, The Netherlands; 6) Institute of Metabolic Disease, Baylor Research Institute, Dallas, TX, USA; 7) Mission Massimo Foundation, Victoria, Australia; 8) Department of Pediatrics, Children's National Medical Center, Washington, DC, USA; 9) Genetic Metabolic Disorders Research Unit, Children's Hospital at Westmead, Sydney, Australia; 10) Murdoch Childrens Research Institute; 11) Genetic Health Queensland at the Royal Brisbane and Women's Hospital, Herston, Queensland, Australia; 12) Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia.

Rare diseases, the vast majority of which are genetic in origin, have a particularly detrimental effect on children. More than 50% of rare disease patients are paediatric, and at least 30% of these children will not live to see their first birthday. Additionally, many of these patients will be incorrectly diagnosed (at least 40% according to EURODIS), or will remain without a final diagnosis. For example, it is well established that half of the patients with leukoencephalopathies, rare central nervous system white-matter disorders, remain a diagnostic mystery despite the fact that they are known to have a genetic basis. Here we present a series of studies showing the successful application of high-throughput genome and exome sequencing to paediatric patients with unresolved diagnoses. These include i) identification of the mutations in DARS responsible for an young boy's leukoencephalopathy and the characterisation of a novel disease (HBSL), ii) identification of a de novo TUBB4A mutation responsible for a leukoencephalopathy of unknown genetic aetiology, H-ABC, iii) identification of a novel de novo mutation in KCNT1, a potassium transporter, in a child with non-specific white matter abnormalities and severe epilepsy, which led to treatment with specific channel therapies, and iv) in less than a month, identification of the mutation responsible for an unexplainable case of Leigh disease, which facilitated patient enrolment in an ongoing clinical trial. In collaboration with the Global Leukodystrophy Initiative (GLIA) and others, we are now working with large cohorts of children with rare diseases with the aim of dramatically reducing the number of unresolved cases.

3010F

Rai1 dosage in early development is essential to PTLS-like phenotype development in Dt-Rai1 animals. *L. Cao[1], C. Abad[1], J. Molina[2], J.I. Young[1,3], K. Walz[1,3,4].* 1) Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, FL, USA; 2) Centro de Estudios Científicos, CECS, Valdivia, Chile; 3) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, FL, USA; 4) Department of Medicine, Miller School of Medicine, University of Miami, FL, USA.

In humans, duplication of Rai1 is associated with the Potocki-Lupski syndrome, a neurodevelopmental disorder that causes failure to thrive, hypotonia, intellectual disabilities and autism. Previously, we had generated a mouse line carrying a transgene encoding Rai1 under the control of a tTa responsive promoter. By crossing this mouse line with camkII-tTa transgenic mice which express tTa downstream of CamkII promoter, we obtained double-transgenic mice (Dt-Rai1) which express transgenic Rai1 specifically in mouse forebrain neurons. Surprisingly, Dt-Rai1 mice showed most of the PTLS-like phenotypes including underweight, lower abdominal fat content, hyperactivity, and impaired learning and memory ability. Doxycycline administration can inactivate tTA mediated transcription and thus turn off the transgene expression. A critical question for possible therapeutic approaches to PTLS and related disorders concerns phenotypic prevention and reversibility. Here we applied doxycycline treatment to Dt-Rai1 animals at different time points: 3-5 months, to test phenotypic reversibility, 1-3 months to test phenotypic prevention. Our study showed that turning off the transgene expression after the onset of the phenotypes (3-5 month) cannot rescue the phenotypes and turn off the transgene expression before the onset of the phenotypes (1-3 month) cannot prevent the development of the phenotypes in mice either. Interestingly, when the mice were treated with doxycycline from embryonic development to 3 month, some of the phenotypes are rescued as activity in open field and context memory in fear conditioning test. At 3 month of age, Body weight showed an increase of ~3 g compared with mice treated with doxycycline from 1-3 month. Our results showed that Rai1 dosage in early developmental stage is essential to the development of PTLS-like phenotype in mice.

3011W

Role of REEP1 (SPG31) in mitochondrial structure and energetic function. *C. Goizet[1,2], J. Lavie[1], N. Bellance[1], S. Melser[1], G. Solé[3], D. Hannequin[4], S. Lyonnet[5], S. Forlani[6,7,8], A. Brice[6,7,8,9], G. Stevanin[6,7,8,9], A. Durr[6,7,8,9], R. Rossignol[1], D. Lacombe[1,2], G. Bénard[1].* 1) Univ. Bordeaux Segalen, Laboratoire Maladies Rares : Génétique et Métabolisme (MRGM), EA 4576, Bordeaux, France; 2) CHU Bordeaux, Hopital Pellegrin, Medical Genetic, Bordeaux, France; 3) CHU Bordeaux, Pôle des Neurosciences Cliniques, Pessac, France; 4) INSERM, U614, Service de Neurologie, CHU Rouen, France; 5) Université Paris Descartes and APHP, Département de Génétique, Hôpital Necker-Enfants Malades, Paris, France; 6) INSERM, U675, Paris, France,; 7) Université Pierre et Marie Curie - Paris 6, UMR-S975, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, GHU Pitié-Salpêtrière, Paris, France; 8) CNRS, UMR7225, Paris, France; 9) APHP, GHU Pitié-Salpêtrière, Département de Génétique et Cytogénétique, Fédération de Génétique, France.

Mutations in REEP1 cause SPG31, an autosomal dominant form of hereditary spastic paraplegia (HSP). In a previous study, we showed that fibroblasts and muscle biopsies from one SPG31 patient with a heterozygous truncative mutation of REEP1 (c.106delG ; p.V36SfsX4) displayed defective mitochondrial energy production as well as altered structure of mitochondrial network. We now confirme an alteration of mitochondrial function and architecture using fibroblasts from additional patients with missense mutations (c.166G>A, p.D56N and c.124T>C, p.W42R, respectively). Interestingly, we also find that SPG31 patients reveal higher level of mitochondrial DNA (around 3-fold increase) compared to controls. To analyze whether REEP1 is involved in mitochondrial physiology, we have expressed different isoforms of REEP1 in HeLa cells and we demonstrate different localization profiles for REEP1. We observed that one pool of REEP1 strongly locates to mitochondria inducing the organelle fragmentation. Our results validate the role of REEP1 in mitochondrial functions.

3012T

Comprehensive genetic analysis of autosomal dominant spinocerebellar ataxia using a next-generation sequencing system. *Y. Higuchi, A. Yoshimura, J. Yuan, Y. Sakiyama, R. Saigo, R. Hirano, A. Hashiguchi, Y. Okamoto, R. Okubo, H. Takashima.* Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Science, Kagoshima City, Kagoshima, Japan.

Objective: Autosomal dominant spinocerebellar ataxias (AD-SCAs) are a group of clinically and genetically heterogeneous neurodegenerative disorders. Genetic studies revealed at least 20 disease-causing genes in patients with AD-SCA. Concomitant with the increase in the number of genes, labor, and reagent costs for molecular genetic testing has increased significantly. We established a diagnostic procedure for AD-SCA using a next-generation sequencing (NGS) system. Methods: First, 208 patients with AD-SCA were screened for repeat expansions in the genes encoding SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA12, SCA31, and DRPLA. Second, 31 patients who were negative for the first screening were enrolled. Using the NGS system (Illumina MiSeq®), we screened for mutations in nine known AD-SCA disease-causing genes, including SPTBN2, TTBK2, KCNC3, PRKCG, ITPR1, PDYN, FGF14, AFG3L2, and TGM6 (genes encoding SCA5, SCA11, SCA13, SCA14, SCA15, SCA23, SCA27, SCA28, and SCA35, respectively). Results: In the first screening, 171/208 (82.2%) patients received a molecular diagnosis; SCA31 was the most frequent subtype followed by SCA6, SCA3, and DRPLA. Secondly, mutation screening using the NGS system revealed four novel missense variations in SPTBN2, TTBK2, and PRKCG. Conclusion: We established a comprehensive diagnostic procedure for AD-SCA using an NGS system. NGS facilitates target resequencing for rapid, accurate, and low-cost applications. No pathogenic mutations were conclusively detected in approximately 18% patients, suggesting that more causative genes should be related with the AD-SCA phenotype. It is expectable that exome sequencing and whole-genome sequencing will contribute to the identification of the novel causative genes.

3013F

The phenotype associated with *ASPM* mutation expands to severe Seckel syndrome. *D.J. Morris-Rosendahl[1], G. Yigit[2], G. Carpenter[3], R. Colnaghi[3], K. Mueller[1], G. Borck[4], M. Trimborn[5], F. Beleggia[2], M. Leipoldt[1], G. Nuernberg[6], K. Seufert[1], B. Wollnik[2], M. O'Driscoll[3].* 1) Institute of human Genetics, Albert-Ludwigs University of Freiburg Medical Center, Freiburg, Germany; 2) Institute of Human Genetics, University of Cologne, Germany; 3) Genome Damage and Stability Centre, University of Sussex, Brighton, UK; 4) Institute of Human Genetics, University of Ulm, Ulm, Germany; 5) Institute of Medical and Human Genetics, Charité University, Berlin, Germany; 6) Cologne Genomics Center, Cologne, Germany.

ASPM is one of the 10 genes known to be associated with primary autosomal recessive microcephaly (MCPH), characterized by reduced head circumference at birth and varying degrees of intellectual disability (ID). Mutations in ASPM are responsible for 25-50% of MCPH, depending on the ethnic background of the patients. The phenotypes associated with the MCPH genes are continually being re-defined and mutations in the *CENPJ* (MCPH6) and *CEP152* (MCPH9) genes are now known to also cause Seckel syndrome. Clinical findings so far associated with *ASPM* mutation include late-onset seizures, pyramidal tract involvement, simplified gyral pattern of the cortex, ventricle enlargement, partial corpus callosum agenesis, mild cerebellar hypoplasia, focal cortical dysplasia and unilateral polymicrogyria. Non-neurologic abnormalities include short stature, idiopathic premature puberty and renal dysplasia. We now describe a family with two daughters (P1 and P2) with a severe Seckel syndrome phenotype including severe microcephaly (P1: OFC at birth 25,5 cm, 38th week; P2: OFC at birth 24,5 cm, 37th week), severe brain malformation and short stature (P1 at 8y8m: weight 10 kg, length 1m; P2 at 5y8m: weight 7 kg, length 86 cm). Both girls have severe intellectual disability, P1 learned to walk with support at 3,5 years and can say a few words, P2 can neither walk nor talk. Extensive cytogenetic and molecular genetic work-up included karyotyping, analysis of chromosome condensation, array CGH (Agilent 244K), genome-wide linkage analysis on an Affymetrix SNP array and sequencing of the following genes: *MCPH1, CEP152, CENPJ, RCC2*, with the only possibly pathological finding being a novel heterozygous missense change, c.2378C>A, in the *MCPH1* gene, in P1, P2 and their father. Subsequent Whole Exome Sequencing (WES) revealed compound heterozygosity for novel truncating *ASPM* mutations: c.T1674A>T (p.Y558X) and c.9324delT (p.L3108fsX) in both affected daughters and heterozygosity for a splice mutation, c.688-2A>C in *CEP164* in P1. Analysis of the ATM- and ATR-dependent G2-M checkpoint responses in EBV-transformed lymphocyte cell lines from P1 and her father, together with the absence of premature chromatin condensation, made MCPH1 an unlikely contributing candidate for the disorder. Functional cellular analysis confirmed ASPM pathogenicity in lymphocyte cell lines from the patients.

3014W

Mutation in the *SYNJ1* gene associated with autosomal recessive, early-onset parkinsonism. *S. Olgiati[1], M. Fang[2], M. Picillo[3], M. Quadri[1], G.J. Breedveld[1], J. Graafland[1], B. Wu[2], F. Xu[2], R. Erro[3], M. Amboni[4,5], S. Pappata[6], M. Quarantelli[6], H.F. Chien[7], E.R. Barbosa[7], B.A. Oostra[1], P. Barone[5], J. Wang[2,8,9,10], V. Bonifati[1].* 1) Dept. of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 2) BGI-Shenzhen, Shenzhen, China; 3) Dept. of Neurological Sciences, University of Naples "Federico II", Naples, Italy; 4) IDC Hermitage-Capodimonte Institute, Naples, Italy; 5) Dept. of Medicine and Surgery, CEMAND, University of Salerno, Italy; 6) Biostructure and Bioimaging Institute, National Research Council, Naples, Italy; 7) Dept. of Neurology, University of Sao Paulo, Brazil; 8) Dept. of Biology, University of Copenhagen, Copenhagen, Denmark; 9) King Abdulaziz University, Jeddah, Saudi Arabia; 10) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark.

Autosomal recessive, early-onset parkinsonism is clinically and genetically heterogeneous. Here, we report the identification, by homozygosity mapping and exome sequencing, of a *SYNJ1* homozygous mutation (p.Arg258Gln) segregating with disease in an Italian consanguineous family with parkinsonism, dystonia, and cognitive deterioration. Response to levodopa was poor, and limited by side effects. Neuroimaging revealed brain atrophy, nigrostriatal dopaminergic defects, and cerebral hypometabolism. *SYNJ1* encodes synaptojanin 1, a phosphoinositide phosphatase protein with essential roles in the post-endocytic recycling of synaptic vesicles. The mutation is absent in human variation databases, is damaging according to all prediction programs, and replaces an amino acid that is extremely conserved in the synaptojanin 1 homologues and in SAC1-like domains of other proteins. Sequencing the *SYNJ1* ORF in unrelated patients revealed another heterozygous mutation (p.Ser1422Arg), predicted as damaging, in a patient who also carries a heterozygous *PINK1* truncating mutation. The *SYNJ1* gene is a compelling candidate for parkinsonism; mutations in the functionally linked protein auxilin cause a similar early-onset phenotype, and other findings implicate endosomal disturbances in the pathogenesis. Our data delineate a novel form of human Mendelian parkinsonism, and provide further evidence for abnormal synaptic vesicle recycling as a central theme in the pathogenesis.

3015T

Copy Number Variants and deletion classes as modifier elements of phenotype in Angelman deleted patients. *S. Russo[1], V. Giorgini[1], E. Mainini[1], G. Randazzo[2], P. Bonanni[2], M. Viri[3], A. Vignoli[4], M. Elia[5], M.T. Bonati[1], L. Larizza[1,6].* 1) Molecular Biology, Institute Auxologico Italiano, Milano, Italy; 2) Epilepsy and Clinical Neurophysiology Unit, Institute Eugeneo Medea, Conegliano Veneto, Italy; 3) Epileptology Center, Hospital Fatebenefratelli, Milano; 4) Epileptology Center, Hospital San Paolo, Milano; 5) IRCCS L'Oasi, Troina (EN), Italy; 6) Medical Genetics, San Paolo School of Medicine, University of Milan, Italy.

Angelman syndrome is a very rare (1:15000) neurodevelopment disease defined by the co-occurrence of major clinical signs, reported in all patients as psychomotor delay, speech absence, peculiar behavior and ataxia. Frequently other features are present: epilepsia occurring in 80% of AS cases, with a very heterogeneous presentation varying for seizure control and age of onset, and associated to a typical EGG profile, microcephaly and sleep disturbance. Molecular bases of disease consist in genetic and epigenetic defects within 15q11-12 region, which affect the expression of UBE3A gene (MIM 601623) coding for an ubiquitinE3 ligase involved in protein degradation through the ubiquitin-proteasome. Clinical presentation appears heterogeneous, both in the severity of the major features and in the kind and number of less frequent signs. This heterogeneity may only partially be accounted by different genetic defects and by mosaicism occurrence. The most common defect (60-70%) is a deletion originating de novo on the maternally inherited chromosome, with different extension in size. Other mechanisms are 15 paternal uniparental disomy, imprinting defects and point mutations of UBE3A gene. Aiming at disclosing the role of modifier factors, we selected a cohort of 50 AS deleted cases. They have been characterized for the deletion size and for harboring rare CNVs by microsatellite segregation and SNP array (Human660W-Quad e Omni 1 Quad). In agreement with the literature most cases grouped in the two main deletion classes, BP1-BP3 and BP2-BP3, while a fraction higher than that reported (17% vs 10%) carried atypical deletions mediated by novel breakpoints. However cases within a specific deletion class displayed a fairly heterogeneous clinical presentation. SNPs array results highlighted several variants yet unreported in databases: although no rearrangement was shared among our AS cohort , a number of interrupted, deleted and duplicated genes were disclosed. Most of the affected genes were reported as susceptibility factors for epilepsy, schizophrenia and autism and code for channels implicated in Na/Ca exchange, receptors and protein involved in cellular trafficking or oxidative stress. Validation of the most interesting modifier genes and the resulting genotype-phenotype correlations is ongoing.

3016F

A Charcot-Marie-Tooth type 2 family with two neuropathy gene variants identified through next generation sequencing using a targeted panel. *F.H. Sansbury[1], T. Antoniadi[2], C. Buxton[2], R. Whittington[2], N.J. Gutowski[3], P.D. Turnpenny[1].* 1) Peninsula Clinical Genetics Service, Royal Devon & Exeter NHS Foundation Trust, Royal Devon & Exeter Hospital (Heavitree), Gladstone Road, Exeter, Devon EX1 2ED, United Kingdom; 2) Bristol Genetics Laboratory, North Bristol NHS Trust, Southmead Hospital, Bristol BS10 5NB, United Kingdom; 3) Department of Neurology, Royal Devon & Exeter NHS Foundation Trust, Royal Devon & Exeter Hospital (Wonford), Barrack Road, Exeter, Devon EX2 5DW, United Kingdom.

We present a 3-generation family with Charcot-Marie-Tooth type 2. Sequencing two affected family members against a targeted panel of neuropathy genes identified a novel *MFN2* c.321G>A p.Asn107Lys variant present in both, plus a previously reported *SCN9A* c.2215G>A p.Ile739Val variant present only in 1 patient. Sanger sequencing of the variants showed the *MFN2* variant was present in 6 of 7 affected family members. The *SCN9A* variant was present in 3 affected family members (including the individual who tested negative for the *MFN2* variant), but also in 1 unaffected family member. Functional studies of the *SCN9A* variant suggest an effect on autonomic function (Han et al, 2012), but its presence in normal controls led the authors to suggest it is a risk factor for small fibre neuropathy in combination with other factors. Interestingly, neuropathic pain is a symptom in the affected family member who does not have the *MFN2* variant. We discuss the possibility of more than one genetic basis for CMT2 in this family and how a neuropathy gene panel approach may identify multiple possible variants.

3017W

A Japanese girl with severe form of vanishing white matter disease resembling Cree leukoencephalopathy. *K. Takano[1], T. Wada[1], H. Osaka[1], Y. Tsurusaki[2], H. Saitsu[2], N. Matsumoto[2].* 1) Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Human Genetics, Yokohama City University, Yokohama, Kanagawa, Japan.

Vanishing white matter disease (VWM)/ childhood ataxia with central hypomyelination (CACH) is an autosomal recessive brain disorder caused by mutations in each of the 5 genes (*EIF2B1-5*), encoding the 5 subunits of eukaryotic translation initiation factor 2B (eIF2B). Brain magnetic resonance imaging (MRI) shows progressive rarefaction and cystic degeneration of the white matter and its replacement by cerebrospinal fluid. The classical phenotype is characterized by early childhood onset and chronic progressive neurological deterioration with cerebellar ataxia, spasticity, optic atrophy and epilepsy. Recently, it has become apparent that the onset of disease varies from antenatal period to adulthood. There is some genotype-phenotype correlation in VWM patients. Cree leukoencephalopathy is a severe variant of VWM and caused by homozygous mutation (p.R195H) of the *EIF2B5* gene. Our patient is a daughter of healthy, unrelated Japanese parents. She developed lethargy, vomiting and seizure shortly after an oral poliovirus vaccination at the age of 4 month. She presented rapid neurological deterioration, loss of head control, axial hypotonia and pyramidal tract sign within a month of onset. Brain MRI showed abnormal white matter intensity. Exome sequencing identified two mutations at the heterozygous state in the *EIF2B5* gene, a known mutation identified in Cree leukoencephalopathy c.584G>A (p.R195H) and a novel mutation c.1223T>C (p.I408T) in the patient. Ile408 is highly conserved and located within I-patch domain which is important in the interactions between eIF2B subunits. Our patient has a severe phenotype with early onset and progressive disease course resembling Cree leukoencephalopathy. Our patient's genotype may correlate with a severe phenotype.

3018T

Proteolipid protein 1 and gap junction α12 gene mutations in 72 Chinese patients with Pelizaeus-Merzbacher disease/ Pelizaeus-Merzbacher like disease and prenatal diagnosis of 15 fetuses in twelve Chinese families with PMD probands. *J. Wang[1], D. Li[1], Y. Wu[1], J. Xiao[2], Q. Gu[1], H. Zhao[1,3], J. Shang[1,4], Y. Yang[1], X. Bao[1], H. Xiong[1], Y. Zhang[1], T. Ji[1], M. Li[1], Y. Meng[5], F. Fang[6], Z. Niu[3], J. Qin[1], H. Shi.[5], F. Zhang[7], X. Wu[1], Y. Jiang[1].* 1) Peking University First Hospital, Beijing, China; 2) Department of Image, Peking University First Hospital, Beijing 100034, China; 3) Department of Neurology, First Hospital of Shanxi Medical University, Taiyuan 030001, China; 4) Department of Neurology, Shanxi Dayi Hospital Affiliated to Shanxi Medical University, Taiyuan 030001, China; 5) Department of Pediatrics, Peking Union Medical College Hospital, Beijing 100730, China; 6) Department of Neurology, Beijing Children's Hospital Affiliated to Capital Medical University, Beijing 100045, China; 7) State Key Laboratory, School of Life Science, Fudan University, Shanghai 200433, China.

Purpose The object of this study was to identify proteolipid protein 1(PLP1) and gap junction α12 mutations in 72 Chinese patients (P1-72) with Pelizaeus-Merzbacher disease (PMD) /PMLD and prenatal diagnosis of fifteen fetuses in twelve Chinese families with PMD probands. Methods Genomic DNA was extracted from peripheral bloods samples. At 19 or 20 weeks gestation, amniotic fluid/chorionic villus sampling (AFS/CVS) was performed. Gene dosage was determined by Multiplex Ligation-dependent Probe Amplification (MLPA). All 7 exons and exon-intron boundaries of PLP1 gene were amplified and analyzed by direct DNA sequencing. Results Of these 72 patients, there were 18 transitional, 45 classical, and 9 congenital PMD according to the clinical and radiological presentation. PLP1 duplications were identified in patients 1-52 with PMD, account for 72.2% (52/72). Their mothers were PLP1 duplications carriers except P52'mother was wildtype. 15 hemizygous missense mutations including eight novel mutations and one reported splicing mutation (IVS5-1G>A) were found in 17 Patients (P53-69) with PMD(23.6%), 12 out of 17 for their mother were showed to be a heterozygote of those mutation, and the remains 5 patients demonstrated de novo. For three patients without PLP1 mutation, we then tested GJC2 mutations with c.925_938del(p.A309Pfs342X), c.201C>G(p.C67T), c.689delG (p.G230Afs), c.735C>A(p.C245X), and c.1199C>A (p.A400E). For the results of prenatal diagnosis(male 9 and female 6), 9 fetuses were PLP1 wildtype, 1 was with PLP1 duplication carrier, and 5 found PLP1 duplication and 1 with c.623G>T(G208V). Conclusions We identified 52 genomic duplications and fifteen missense/splicing mutation of PLP1 gene in 69 Chinese patients with PMD and five missense/framshift mutations in three patients with PMLD. Prenatal diagnoses for fifteen fetuses in twelve PMD proband families were performed, which is useful and helpful for those families. This would be the first report about PLP1 mutations and prenatal diagnosis in PMD patients in mainland of China. Correspondence to: Dr. JIANG Yuwu, Department of Pediatrics, Peking University First Hospital, Beijing 100034, China (Email: jiangyw@263.net) This research was supported by the grants from '973' Project of the Science and Technology Ministry of China (2012CB944602), National Key Research Project '12-5' (2012BAI09B04), National Natural Science Foundation of China (81271257), Beijing Natural Science Foundation (13G20083).

3019F

Comprehensive Mutation Analysis of 421 Cases with Charcot-Marie-Tooth Disease using Microarray and Exome Sequencing. *A. Yoshimura[1], A. Hashiguchi[1], Y. Higuchi[1], J. Yuan[1], Y. Okamoto[1], T. Nakamura[1], J. Mitsui[2], H. Ishiura[2], Y. Takahashi[2], J. Yoshimura[2], K. Doi[2], S. Morishita[2], S. Tsuji[2], H. Takashima[1].* 1) Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima City, Japan; 2) University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.

[Objective] Charcot-Marie-Tooth (CMT) disease comprises a group of clinically and genetically heterogeneous inherited peripheral neuropathies. In our study, to identify causative genes and new molecular mechanisms of CMT, DNA microarray and exome sequencing were successively applied. [Methods] From April 2005 to December 2011, after excluding the PMP22 duplication mutation using fluorescence in situ hybridization, 421 cases with suspected inherited peripheral neuropathies were enrolled from Japan. Mutation screening was initially performed on a custom DNA microarray (Affymetrix) containing all 28 known CMT disease-related genes. The negative cases subsequently proceeded to exome sequencing with Illumina HiSeq2000. Using Sanger's method, we confirmed the suspected variants, and a segregation study was performed if the variants were novel. [Result] In 421 cases of suspected CMT, causative mutations were discovered in 56 cases (13.3%) using DNA microarray. Exome sequencing then revealed another 28 reported mutations in the known CMT disease-related genes. Correspondingly, the positive rate of mutation detection was increased in demyelinating CMT (from 15.0% to 23.9%), and axonal CMT (from 12.2% to 22.4%) as well. The total positive rate of mutation detection was improved from 13% to 20%. [Conclusion] We compared the ability of DNA microarray analysis and exome sequencing to detect known pathogenic mutations in CMT, and exome sequencing proved to be more effective. Next-generation sequencing is a trustworthy method in high-thoughput genetic analysis. However, no pathogenic mutation was found in approximately 80% of cases with suspected CMT. In future, we can expect that a large number of novel disease-related genes will be discovered.

3020W

ABNORMAL COPY NUMBER VARIANTS ARE FREQUENT IN PATIENTS WITH MALFORMATIONS OF CORTICAL DEVELOPMENT ASSOCIATED WITH EPILEPSY. *F.R. Torres[1], D.A. Souza[1], M.M. Guerreiro[2], M.A. Montenegro[2], A.C. dos Santos[3], V.C. Terra[3], A.C. Sakamoto[3], F. Cendes[2], I. Lopes-Cendes[2].* 1) Department of Medical Genetics, UNICAMP, Campinas, Sao Paulo, Brazil; 2) Department of Neurology, UNICAMP, Campinas, Sao Paulo, Brazil; 3) Department of Neurosciences, USP, Ribeirao Preto, Sao Paulo, Brazil.

Patients with malformations of cortical development (MCD) often suffer from seizures which are frequently refractory to treatment with antiepileptic drugs. Advances in molecular genetics have led to a better understanding of the mechanisms underlying several types of MCDs. However, mutation events remain unidentified in the majority of patients. Recent studies have implicated large, rare copy number variants(CNVs) in a range of neurodevelopmental disorders. Therefore, the aim of this study was to investigate whether CNVs could be involved in different types of MCDs. We used a high resolution SNP-array, CytoScan® HD, to investigate CNVs in a cohort of 33 patients with MCDs, including lissencephaly spectrum, periventricular nodular heterotopia and schizencephaly. To assess the clinical significance of our insertions/deletions findings, we searched all CNVs found in the Database of Genomic Variants (DGV) and The International Standards for Cytogenomic Arrays Consortium (ISCA). Genes located within rare CNVs were subsequently subjected to specific Gene Ontology Terms (Go) search. We detected at least one rare CNV in 13 patients (40%), with a total of 19 rare CNVs identified. The average size of CNVs found was 223kb, ranging from 113kb to 456kb. Each CNV contained approximately two genes. Potentially pathogenic CNVs, according to DGV and ISCA databases, contained genes involved in cell division (*NCAPG2*, *HAUS7*), vesicle mediated transport (*TSNARE1*), actin cytoskeleton organization (*DAAM1*, *ACTR6*) and axon guidance or neuronal migration (*PLXNA1*, *KIRREL3*, *ARX*, *DCX*). Several genes located in regions spanned by rare and potentially pathogenic CNVs described in this work are related to molecular pathways previously shown to be involved in MCDs. In addition, deleterious mutations in genes controlling mitosis, vesicle mediated transport, cytoskeleton organization and neuronal migration, have been previously reported in patients with MCDs. In conclusion, we have shown that SNP-array study is a powerful tool for identifying deleterious variants in patients with MCDs as well as to indicate new candidate genes potentially involved in normal and abnormal cortical development.

3021T

Sensorineural hearing loss in OPA1-linked disorders. *D. Bonneau[1], S. Leruez[2], E. Colin[1], D. Milea[2], S. Defoort-Dhellemmes[3], M. Crochet[3], C. Verny[10], M. Ferré[1], J. Lamblin[4], V. Drouin[5], C. Vincent-Delorme[6], G. Lenaers[7], C. Blanchet[8], G. Juul[9], M. Larsen[9], V. Procaccio[1], P. Reynier[1], P. Amati-Bonneau[1].* 1) Département de Biochimie et Génétique, CHU Angers, CNRS UMR6214, INSERM UMR1083 France; 2) Département d'Ophtalmologie CHU Angers, France; 3) Département d'Ophtalmologie CHU Lille, France; 4) Département d'Oto-Rhino-Laryngologie, CHU Lille, France; 5) Département de Biochimie et Génétique, CHU Rouen, France; 6) Service de Génétique Médicale, Centre Hospitalier Arras, France; 7) INSERM U1051, Montpellier, France; 8) Département d'Oto-Rhino-Laryngologie, CHU Montpellier; 9) National Eye Clinic, Kennedy Center and University of Copenhagen, Denmark; 10) Département de Neurologie, CHU Angers, France.

Heterozygous mutations of OPA1 (Optic Atrophy 1), that encodes for a dynamin-related GTPase involved in mitochondrial DNA maintenance, are linked to autosomal dominant optic atrophy (DOA). Optic neuropathy usually occurs insidiously in the first decade of life, about 20% of patients later developing extra-ocular symptoms (DOA+) in particular deafness. We retrospectively reviewed the files of 1380 DOA patients referred to our laboratory from 2003 to 2011. OPA1 mutation was identified in 327 patients (24%) of whom 21 patients (6.4%) had hearing impairment. In 10 patients deafness was detected under age 20 (~48%), in 3 patients over age 20 (~14%) and in 8 patients the age of onset was unknown (38%). Molecular screening of OPA1 identified three mutations, p.Arg445His, p.Gly401Asp and p.Leu243 previously reported associated to optic atrophy and hearing loss; two mutations, p.Val291_Phe328del and p.Ile463_Phe464dup reported in isolated optic atrophy and two novel mutations, p.Arg437Glu and p.Ala357Leufs*4, implicated for the first time in DOA and hearing loss. Optic atrophy and deafness are the only signs in 13/21 patients (62%) additional neurological signs are present in 8/21 patients (38%). In 54% of patients hearing loss started prior to visual abnormalities. A particularly careful should be performed in patients carrying an OPA1 missense mutation because there is a 2-4 fold increased risk of developing 'DOA plus' with missense mutations affecting the GTPase domain or the dynamin domain. In conclusion, deafness linked to OPA1 is probably underestimated since audiological investigations are performed in only a minority of OPA1 mutation carriers.

3022F

Targeted re-sequencing of the human X chromosome exome expands the phenotypic spectrum of ATP7A mutations. *S. MOUTTON[1], A. DELAHODDE[2], M. LANGOUET[1], C. BOLE-FEYSOT[3], K. SIQUIER-PERNET[1], P. NITSCHKE[4], C. VASNIER[2], S. HADJ-RABIA[1], L. YI[5], A. DRECOURT[1], K. PATEL[5], N. BODDAERT[6], A. MUNNICH[1], J.P. BONNEFONT[1], S.G. KALER[5], J. AMIEL[1], L. COLLEAUX[1].* 1) INSERM U781, Université Paris Descartes, Sorbonne Paris Cité, Institut IMAGINE, PARIS, France; 2) Université Paris-Sud, CNRS, UMR 8621, Institut de Génétique et Microbiologie, Orsay, F-91405, France; 3) Plateforme de Génomique, Institut IMAGINE, 75015 Paris, France; 4) Plateforme de Bioinformatique, Université Paris Descartes, Sorbonne Paris Cité, Institut IMAGINE, 75015 Paris, France; 5) Unit on Human Copper Metabolism, Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, USA; 6) Department of Paediatric Radiology, Hôpital Necker Enfants Malades, APHP, Paris, France.

We investigated a large pedigree consistent with X-linked inheritance by a combination of linkage analysis and X-exome sequencing using RainDance Technologies®. All 6 affected males exhibited epileptic encephalopathy. Epilepsy onset ranged from 2 to 22 months and consisted of myoclonic seizures, febrile seizures and Lennox-Gastaut syndrome. A single variant perfectly segregating with the disease was identified. This variant (c.3670G>A; p.Gly1224Ser) is located in the gene encoding the P-type copper-transporter ATPase (*ATP7A*). Yeast complementation assays using the *S. cerevisiae* copper transport mutant ccc2Δ demonstrated its pathogenicity since the p.Gly1224Ser mutant *ATP7A* cDNA was unable to restore normal growth on nutritionally restricted media. Mutations in *ATP7A* yield three distinct X-linked conditions. Menkes disease (MD) presents in the first monthes of life with failure to thrive, developmental delay, and seizures. The diagnosis of MD is usually confirmed by decreased serum copper and ceruloplasmin levels and intracellular copper retention. Occipital horn syndrome features more subtle developmental delay, dysautonomia, and connective tissue anomalies beginning in early childhood. Finally, ATP7A-related distal motor neuropathy is characterized by a late-onset neurological phenotype that resembles Charcot-Marie-Tooth disease, type 2.

Because none of these diagnoses had been suggested in our family we undertook detailed re-evaluation of the patients. Brain MRI showed normal cerebral blood vessels morphology but metal overload in basal ganglia. Despite normal hair texture, optic microscopy revealed rare pili torti. Biochemical analysis showed not only normal ceruloplasmin and mildly increased serum copper levels but, surprisingly, low intracellular copper level in patient's fibroblasts compared to controls (15.5 ng/mg of protein). Finally, immunohistochemical analyses performed in patient's fibroblasts showed normal localization to the trans-Golgi network (TGN) of the mutant protein and normal trafficking from the TGN to the plasma membrane in response to increased copper concentration. Our report further expands the phenotypic spectrum associated with ATP7A mutations. Additional experiments are now underway to elucidate the pathophysiological mechanism underlying this condition which is likely different from known MD/OHS-causing mutations.

3023W

Phenotypic and transcriptomic characterization of the RPE affected by mutations that cause RNA splicing factor retinitis pigmentosa. *M.H. Farkas[1], E.F. Nandrot[2], E.D. Au[1], D. Lew[2], K. Bujakowska[1], M.E. Sousa[1], D.G. Taub[1], S.S. Bhattacharya[2], E.A. Pierce[1].* 1) Ocular Genomics Institute, Berman-Gund Laboratory, Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA; 2) Institut de la Vision, Paris, France.

Mutations in the Pre-mRNA Processing Factors 3, 8, and 31 (PRPF3, 8, and 31) cause non-syndromic retinitis pigmentosa (RP) in humans, an inherited retinal dystrophy (IRD). It is currently unclear what mechanisms, or which tissues, are affected when mutations are present in these ubiquitously expressed proteins. Mice with the human mutations in these genes show limited vision loss, accompanied by late-onset morphological changes in the retinal pigment epithelium (RPE), specifically a loss of basal infoldings, vacuole formation, and basal deposits. Since the RPE is critically important for the overall health and maintenance of the retina, we set out to determine if the observed morphological changes are preceded by abnormal function. Specifically, we investigate, in detail, the phagocytic mechanism of the RPE in Prpf3, Prpf8, and Prpf31-mutant mice. Phagocytosis deficiencies of up to 50% are observed in primary RPE cultures, along with a nearly 60% decrease in RPE adhesion. The RPE diurnal rhythm is also affected in all 3 models. In addition, RNA-Seq data show alterations in gene expression in RPE and retina, as well as in brain and muscle. The greatest number of transcripts found to be differentially expressed (DE) are associated with the RPE of the Prpf3-mutant mice, with over 2000 identified. In comparison, all other tissues/models had a few hundred DE transcripts. Categorization of DE transcripts in the RPE of all 3 models revealed an over abundance of transcription factors, IRD genes and extracellular matrix components. Interestingly, there is no overlap of DE transcripts among all 3 models. All 3 mutant models also display an increase in unique novel splicing events, and we find differential expression of novel genes between the mutant and control mice. The RPE appears to be the primary site of pathogenesis for the RNA splicing factor forms of RP. While degenerative changes appear in 2-year old mice, consistent with the late disease onset observed in humans, our characterization suggests the RPE is dysfunctional soon after birth. RNA-Seq analyses, to date, suggest that aberrant splicing occurs globally, yet only affects retinal tissue. The divergent altered transcripts among the 3 models suggest different mechanisms of pathogenesis, even though all 3 Prpf proteins are found in the same core component of the spliceosome. These results provide the greatest insight into the pathogenesis and mechanism of RNA splicing factor RP to date.

3024T

THE RECRIUTMENT OF RNA POLYMERASE II IS IMPAIRED IN CORNELIA DE LANGE SYNDROME. *A. Musio[1], S. Bilodeau[2], C. Amato[1], V. Quarantotti[1], F. Cucco[1,3], I. Krantz[4], L. Mannini[1].* 1) Istituto di Ricerca Genetica e Biomedica, CNR, Pisa, Italy; 2) Centre de recherche en cancérologie de l'Université Laval and Centre de recherche du CHU de Québec (Hôtel-Dieu de Québec), Département de biologie moléculaire, biochimie médicale et pathologie, Faculté de Médecine, Université Laval, Québec, Canada; 3) Dipartimento di Biologia, Università di Pisa; 4) Division of Human Genetics, The Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania.

Cohesin ensures correct chromosome segregation by holding sister chromatids together from the S phase until their separation in anaphase. Mutations in cohesin and regulatory cohesin genes, NIPBL, SMC1A, HDAC8 and SMC3 are causative of Cornelia de Lange syndrome (CdLS), a rare human development disorder. Since CdLS cell lines show no clear defects in sister chromatid cohesion, the molecular basis underlying CdLS remains elusive. Beyond its structural function in maintaining sister chromatid cohesion, cohesin plays a key role in transcription regulation. It has been suggested that cohesin regulates higher order chromatin organization by the formation of chromatin loops. Recently it has been shown that gene expression profiling of NIPBL- and SMC1A-mutated cell lines exhibit a down-regulation of many genes providing a mechanism for CdLS. However, the molecular mechanisms by which cohesin mutations affect gene expression are not well understood. We hypothesized that the gene transcriptional machinery was somehow affected by mutations in SMC1A found in cell lines derived from CdLS patients. To investigate this we used chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-seq) to identify genomic regions co-occupied by cohesin, NIPBL and RNA pol II in normal human lymphoblastoid cells. Regions occupied by cohesin, NIPBL and RNA pol II were assigned to genes and compared to available gene expression data from CdLS cell lines. We selected 6 genes for our mechanistic studies, including cell-type specific and housekeeping genes that were differentially expressed between normal and SMC1A mutant cells. We found that the recruitment of RNA pol II to the regulatory regions is strongly decreased in CdLS cell lines irrespective of SMC1A mutations. These findings highlight the pivotal role of cohesin in transcription regulation, and the deficiency of RNA pol II may explain the typical gene dysregulation occurring in CdLS cell lines. This work was supported by a grant from Tuscany Region to A.M.

3025F

Characterization of Novel mutations found in Classical and Infertile CF males of Indian population: A molecular approach to establish genotype and phenotype correlation. *R. Prasad[1], H. Sharma[1], F. Becq[2].* 1) Department of Biochemistry, PGIMER, Chandigarh, India; 2) Institute of Physiology and Cell Biology, University of Poitiers, France.

Cystic Fibrosis an autosomal recessive disorder is usually considered as rare disease for Indian population hence much less is known about this disease in context with Indian sub continent. Recently we have established a spectrum of mutations in CFTR gene from classical Cystic Fibrosis as well as from infertile male patients with CBAVD/CUAVD in Indian populations. Among them S549N, L69H in classical CF and G126S, Y852F in CBAVD males were among the rare missense mutations (Sharma et al ., 2009). In this study we have attempted to conduct in vitro gene expression analysis to establish genotype and phenotype correlation and to characterize these four rare missense mutations according to the mechanism that disrupt CFTR protein function. All four mutations from Indian population were characterized by expressing pEGFP-CFTR constructs in BHK-21 cells via 3 step technique viz; CFTR cellular localization was determined by confocal microscopy, where as Western blot analysis and automated iodide efflux assays was used to determine CFTR maturation processes and its channel activity respectively. In Western blot analysis only b-band is obtained for L69H substitution as for F508del whereas in the case of other mutants both b and c bands were found, indicating L69H mutation impair CFTR maturation process, the finding was again confirmed by confocal imaging. Iodide Efflux assay reveled significant decrease in channel activity for L69H and S549N mutants CFTR expressing cells in comparison to WT, although this decrease in channel activity was rescued when cells were incubated at 27 °C. When the effect of CFTR correctors was checked on different mutants, it was found that VX809 significantly ameliorate the defect caused by L69H mutation. Y852F and G126S have no impact on CFTR maturation and function. In conclusion L69H mutation is a class II CF mutation causing impaired maturation leading to protein degradation and Chloride ions impermeability like F508del mutation. This defect is rescued by the corrector VX 809. Whereas S549N mutation can be categorized into a class II/III mutation causing impaired maturation and reduced channel activity.

3026W

Distinct prevalence of homozygous p.V37I variant of GJB2 in Chinese Hans with severe-to-profound, mild-to-moderate or normal hearing phenotype. *Y. Chai[1,2], Y. Tao[1,2], H. Wu[1,2].* 1) Ear Institute, Shanghai Jiaotong University School of medicine, shanghai, China; 2) Department of Otolaryngology-Head and Neck Surgery, Xinhua Hospital, Shanghai Jiaotong University School of medicine, Shanghai, China.

The p.V37I variant of *GJB2* is highly prevalent in East Asians, with a carrier frequency of 6.2% in Chinese Hans. It has been shown that homozygous p.V37I may lead to mild-to-moderate hearing impairment and is a genetic risk for postnatal childhood hearing impairment. On the other side, the complete spectrum of hearing phenotypes associated with the p.V37I variant remains elusive. In this study, we recruited 858 and 88 Chinese Han probands with severe-to-profound and mild-to-moderate childhood hearing impairment, respectively, as well as 1000 Chinese Han adults with normal hearing. Direct sequencing of all subjects for *GJB2* showed that a significant percentage (12.5%, 11/88) of probands with mild-to-moderate hearing impairment was homozygous for p.V37I, in contrast to 0.1% (1/1000) in adults with normal hearing ($P<0.001$). Homozygous p.V37I was also moderately over-represented in probands with severe-to-profound hearing impairment (1.63%;, 14/858, $P<0.01$). Targeted next-generation sequencing of 79 known deafness genes in those 11 probands, however, identified one proband with homozygous p.Y1995X mutation in *CDH23* in addition to the homozygous p.V37I variant in *GJB2*. Our results suggested that the highly prevalent p.V37I variant is associated with a broad spectrum of hearing phenotypes. Though present most frequently in patients with milder degree of hearing impairment, it may also exhibit much severer or completely normal hearing phenotypes. Cautions should be taken when interpret the pathogenic causes for patients carrying the homozygous p.V37I variant though, as some may have separate pathogenic causes in addition to a non-penetrant homozygous p.V37I.

3027T

Mosaic Missense Mutations in the RNase IIIb Domain of DICER1 Cause JANUS, a Novel Overgrowth Syndrome. *S. Klein[1,8], H. Lee[1,6,8], S. Ghahremani[2,8], P. Kempert[3,8], M. Ischander[4,8], M.A. Teitell[6,7,8], S.F. Nelson[1,6,8], J.A. Martinez-Agosto[1,5,7,8].* 1) Department of Human Genetics; 2) Department of Radiology; 3) Division of Hematology-Oncology, Department of Pediatrics; 4) Division of Pulmonary Medicine, Department of Pediatrics; 5) Division of Medical Genetics, Department of Pediatrics,; 6) Department of Pathology and Laboratory Medicine; 7) Jonsson Cancer Center; 8) David Geffen School of Medicine at UCLA, Los Angeles, CA.

Specific genetic conditions that predispose to overgrowth and tumor development have been linked to alterations in tumor suppressor genes. Here we report a distinct syndrome of macrocephaly, developmental delay, bilateral Wilms tumor and bilateral neonatal lung cysts. We name this syndrome JANUS an acronym for the core phenotypic findings, which include a Juxtaposition of Abnormal Head Size, Neonatal Pulmonary Cysts, Urologic Tumors, and Somatic Overgrowth. Although there is phenotypic overlap with previously described overgrowth syndromes, this phenotypic association is novel and its genetic etiology unknown. We performed whole exome sequencing on an affected proband and parental blood samples and identified a de novo missense mutation in the RNase IIIb domain of DICER1. We confirmed an additional missense mutation in the same domain of another unrelated patient by Sanger sequencing. These missense mutations in the RNase IIIb domain of DICER1 affect the metal binding located within this domain. Similar mutations have been previously identified in specific tumor types, including Wilms tumor. Pyrosequencing was used to determine the relative abundance of mutant alleles in various tissue types. The relative mutation abundance is highest in Wilms tumor and kidney samples when compared to blood, confirming that the mutation is mosaic. We have performed bioinformatics analysis of microRNAs affected by these specific mutations and have identified target genes that are over-represented in MTOR, MAPK, and TGF- β signaling pathways. We propose that mutations affecting the metal binding sites of the DICER1 RNase IIIb domain alter the balance of 3p and 5p microRNAs leading to deregulation of these three growth signaling pathways, causing JANUS syndrome. We conclude that mutations in the catalytic RNase domain of DICER1 cause deregulation of specific microRNAs (miRNAs) leading to alterations in central growth pathways, namely MTOR, MAPK, and TGF-β. Similar to mutations in the RAS/MAPK pathway, in which germline mutations result in syndromic presentation or 'rasopathies' while somatic mutations result in neoplasm, somatic mutations in the essential RNase IIIb domain of DICER1 are associated with a number of tumors while mosaic mutations in the same domain cause a novel human overgrowth syndrome.

3028F

Gain and loss of function mutations in *PDE4D* result in two different developmental disorders with mirrored phenotypes. *A. Lindstrand[1], G. Grigelioniene[1], D. Nilsson[1], M. Pettersson[1], BM. Anderlid[1,2], P. Gustavsson[1], H. Valta[5], S. Geiberger[3], K. Lagerstedt[1], F. Taylan[1], J. Wincent[1], T. Laurell[4], M. Pekkinen[6], M. Nordenskjöld[1], O. Mäkitie[1,5,6], A. Nordgren[1].* 1) Department of Molecular Medicine and Surgery and Center of Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Women's and Children's Health Karolinska Institutet, Stockholm, Sweden; 3) Department of Pediatric Radiology, Karolinska University Hospital Solna, Stockholm, Sweden; 4) Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden; 5) Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; 6) Folkhälsan Institute of Genetics, Helsinki, Finland.

Point mutations in *PDE4D* have recently been linked to acrodysostosis, an autosomal dominant disorder characterized by facial dysostosis, severe brachydactyly, nasal hypoplasia, short stature, intellectual disability and hormone resistance. We found *PDE4D* mutations in five unrelated patients with acrodysostosis. In addition, two patients with large heterozygous deletions encompassing the entire *PDE4D* locus were identified. These patients do not have acrodysostosis but suffer from a novel intellectual disability syndrome with severe anorexia and characteristic facial features including a long nose and a small chin. Careful clinical comparisons between these patients raised the possibility that different types of *PDE4D* mutations result in opposite phenotype in humans. To further explore the function and pathogenicity of the identified alleles we used zebrafish embryos. We found that zebrafish *pde4d* was expressed in the brain of the developing embryo. Embryos injected with two different morpholinos showed a consistent phenotype (short body length, large head, small jaw), as did embryos injected with mRNA harboring acrodysostosis point mutations (curved body, small head, large jaw). Interestingly, the jaw of embryos injected with mRNA was severely enlarged while in embryos where *pde4d* was suppressed the jaw size decreased, hence capturing the mirror phenotype seen in humans. In addition, results from co-injections of mutated mRNA and wild-type mRNA showed a more severe phenotype compared to embryos injected with the mutated mRNA alone when the mutations were located in the upstream conserved regions of the PDE4D protein, while this effect could not be observed when the mutations were located further down, in the catalytic domain of the protein. In summary, haploinsufficiency of the *PDE4D* gene results in a previously undescribed syndrome with several characteristic traits that represent the extreme opposite of what we see in patients with acrodysostosis. By comparing overexpression of human mutated transcripts to *pde4d* knockdown in zebrafish embryos we could successfully replicate this mirror phenotype. Our data also suggest that the genetic mechanism of the dominant mutations causing acrodysostosis might be different depending on where in the protein they are situated.

3029W

Loss of function mutations in Carbonic Anhydrase XII result in hyponatremic dehydration and elevated sweat chloride concentration. *B. Vecchio-Pagán[1], M. Lee[1], N. Sharma[1], A. Waheed[2], D. Belchis[3], J. Hertecant[4], W. Sly[2], G.R. Cutting[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medicine, Baltimore, MD; 2) Department of Biochemistry, St. Louis University School of Medicine, St. Louis, MD; 3) Pathology, Johns Hopkins School of Medicine, Baltimore, MD; 4) Tawan Hospital, UAE National University, Al Ain United Arab Emirates.

Functional analysis can provide key evidence that a gene linked to a rare disease by exome sequencing is causative; however, the appropriate physiological context needs to be considered when testing the functional effect of mutations. We provide an illustrative example that demonstrates a new role for for CA XII mediated bicarbonate metabolism in determining the ionic composition of sweat. Exome sequencing revealed a novel homozygous mutation (H121Q) in carbonic anhydrase XII (CA XII) in 2 siblings exhibiting hyponatremic dehydration, salt craving, and elevated sweat chloride (Cl^-) levels in a consanguineous Omani pedigree with three normal siblings. CA XII maintains proper physiological pH through the reversible dehydration of HCO_3^- to CO_2 and H_2O at its extracellular catalytic site. To assess the functional effect of H121Q, we measured the rate of CO_2 hydration. In this assay, H121Q displayed reduced activity (15.4±3.6 %WT activity, n=11). Prior publications have described a different point mutation in CA XII, E143K, in an Israeli Bedouin tribe which segregated with an autosomal recessive phenotype similar to that of the Omani family (OMIM #143860). However, the reported functional assays revealed that the E143K mutation only reduced CA XII function to ~70% WT activity, which we confirmed (75.5±4.9 %WT, n=11). To reconcile the apparent dramatic difference in function between these two mutations, we sought to determine whether either led to altered cellular localization of CA XII. Confocal microscopy using a Protein Tech anti-CA XII antibody was performed upon polarized MDCK cells expressing WT and both CAXII mutants. Wild-type, H121Q and E143K CA XII showed similar localization, primarily at the basolateral cell membrane, suggesting that the catalytic sites of these mutants retain an extracellular orientation. As the E143K mutant had previously been reported to have altered anion sensitivity, we assayed both mutants under Cl^- concentrations which more accurately reflect the extracellular environment of CA XII's catalytic site (100mM Cl^-). In this context, the activity of the E143K mutant was dramatically reduced (3.3±1.3 %WT; n=8) to a level similar to that of H121Q (0.9±0.6 %WT; n=8). Thus, by measuring CA XII in the proper physiological context, we show that the disease causing mutations in both autosomal recessive pedigrees produce severe loss-of-function in CA XII activity.

3030T

A novel splicing silencer generated by dystrophin exon 45 deletion could explain exon 44 skipping that modifies dystrophinopathy. *M. Matsuo[2], EM. Dwianingsih[1], RG. Malueka[1], A. Nishida[1], T. Lee[2], M. Yagi[2], K. Iijima[2], Y. Takeshima[2].* 1) Dept Medical Rehabilitation, Fac Rehabilitation, Kobegakuin Univ, Kobe, Japan; 2) Dept Pediatrics, Graduate School of Medicine,Kobe University, Kobe, Japan.

Background. Duchenne muscular dystrophy (DMD), a progressive muscle-wasting disease, is mostly caused by exon deletion mutations in the dystrophin gene. The reading frame rule explains that out-of-frame deletions lead to muscle dystrophin deficiency in DMD. In outliers to this rule, deletion junction sequences have never previously been explored as splicing modulators. Methods. In a Japanese case, we analyzed the dystrophin gene and mRNA by PCR and real-time reverse transcription (RT)-PCR amplification, respectively, and examined dystrophin expression immunohistochemically. We examined the splicing regulatory activity of the cloned junction sequence in an in vitro splicing system using a chimeric doublesex gene pre-mRNA. Results. We identified a single exon 45 deletion in the patient's genome, indicating out-of-frame mutation. However, his muscle showed weak dystrophin signals. RT-PCR amplification of dystrophin exons 42 to 47 revealed a major normally spliced product with exon 45 deletion and an additional product with deletion of both exons 44 and 45, an in-frame transcript, accounting for approximately 6% of the total mRNA. We considered the latter to underlie the observed dystrophin expression. Remarkably, the junction sequence cloned by PCR walking abolished the splicing enhancer activity of the upstream intron in a chimeric doublesex pre-RNA. Furthermore, antisense oligonucleotides directed against the junction site counteracted this effect. This indicated that the junction sequence was a splicing silencer. Conclusions. This is the first evidence that a novel deletion junction sequence is a splicing silencer. It was strongly suggested that creation of splicing regulator is a modifier of dystrophinopathy.

3031F

Skin-specific Kallikrein-5 transgenic mice recapitulate the main features of Netherton syndrome and provide a viable model for therapeutic approaches. *L. Furio[1,2], S. de Veer[1,2], C. Deraison[1,2], A. Briot[1,2], C. Bonnart[1,2], A. Robin[1,2], A. Hovnanian[1,2,3].* 1) University Paris Descartes Sorbonne Cite, Paris, France; 2) INSERM U781 and Imagine Institute of Genetic Diseases, Paris, France; 3) Department of Genetics, Necker hospital, Paris, France.

Netherton syndrome (NS) is a severe genetic skin disease caused by mutations in *SPINK5* encoding the lymphoepithelial Kazal-type-related serine protease inhibitor (LEKTI). Of the proteases which show overactivity in NS (kallikrein-related peptidase (KLK) 5, KLK7 and elastase 2), KLK5 is a potential key initiator of the proteolytic cascade. To address the role of KLK5 in the disease, we have generated a transgenic murine model overexpressing human KLK5 in the skin (Tg*KLK5*). The transgene was highly expressed in the granular layer of Tg*KLK5* epidermis and was associated with increased proteolytic activity. *In situ* zymography showed enhanced KLK5 activity, as well as elevated KLK7 and elastase 2 activities, two proteases which have been proposed to be activated by KLK5. Tg*KLK5* mice show low weight at birth with a growth delay. They show a skin barrier defect with increased transepidermal water loss. The *stratum corneum* is detached from the underlying epidermis as a result of desmosomal cleavage. Whiskers and hairs are abnormal and reduced. Additionally, Tg*KLK5* mice displayed hallmarks of cutaneous inflammation and allergy including high expression of IL-1β, TNF-α and the pro-Th2 cytokine TSLP with mast cells and eosinophils infiltration. Overtime, persistent scratching leads to alopecic, erosive and crusty skin lesions with lymph nodes hyperplasia. Inflammation was also detected at the systemic level, with high serum IgE and TSLP levels. CD4 T cells from lymph nodes showed increased secretion of Th2 and Th17 cytokines. These results help deciphering the role of KLK5 in the NS phenotype and provide a useful model for the development of new therapeutic strategies.

3032W

Unravelling the molecular mechanisms of pathogenesis of Incontinentia pigmenti Mendelian disorder. *F. Fusco[1], MI. Conte[1], A. Pescatore[1], E. Esposito[1], M. Paciolla[1], MG. Miano[1], MB. Lioi[2], JP. Bonnefont[3], S. Hadj-Rabia[3], C. Bodemer[3], G. Royer[3], A. Smahi[3], J. Steffann[3], M.V. Ursini[1].* 1) Dept Human Molecular Genetics, IGB-ABT-CNR, Naples, Italy; 2) University of Basilicata, Potenza, Italy; 3) Hôpital Necker-Enfants Malades, Paris, France.

Incontinentia Pigmenti (IP, OMIM308300) is a X-linked dominant neuroectodermal disease associated with skin defects and with extracutaneous manifestations at variable frequency. IP is caused by mutations in NEMO/IKBKG gene, that partially overlaps the G6PD gene and partially maps in LowCopyRepeat(LCR1). The duplicated copy (LCR2),located in Xq28 in opposite direction to LCR1,contains the nonfunctional copy of NEMO, pseudoNEMO. We will present a review of past and present findings on the NEMO mutations by exploring with high accuracy both genomic context in which the mutations occur and the functional effects on the mutated proteins. Our task is to understand mechanisms of mutagenesis and of pathogenesis that chacterize the IP disease in order to apply these knowledges in diagnosis of disease and eventually, in its therapeutic correction. To these ends, we have compiled a collection of published and novel NEMO mutations. Data catalogued show that 45% are single base-pair substitutions, 42% indels,12% deletions, and 1%complex rearrangements. Moreover, most altered nucleotide sequences result in a premature stop-codon and 10% generate missense mutations. Then, we produced key discoveries on the affected NEMO functions, analysing missense mutations in further detail. NEMO encodes the regulatory subunit of the IKK complex responsible of the activation of NF-kB transcription factor in multiple signalling pathways controlling immunity, cell survival, differentiation and proliferation. Inflammatory cytokines (TNFalfa, IL1 etc) probably function through conformational change in the kinase complex resulting from either ubiquitination of NEMO or NEMO binding to ubiquitin. One of the E3 ligase responsible for non-degradative polyubiquitination of NEMO is TRAF6, which participates in several signalling pathways controlling immunity, osteoclastogenesis, skin development, and brain functions. We studied the IP-associated E57K NEMO mutant found in a mild form of IP, showing an impaired IL-1 signaling, and we established that a fragment encompassing 57-69aa of NEMO is responsible for its binding to the CC domain of TRAF6, required for NF-kB activation. This site appears to work in concert with NUB domain, mapped trough the analysis of the A323P-IP mutant, which binds to the polyubiquitinated chains of TRAF6. These evidences suggest a dual mode of TRAF6 recognition that may allow the design of new inhibitors of NF-kB activation, restricted to TRAF6-dependent stimuli.

3033T

Clinical whole exome sequencing identifies mutations in UPK3A in individuals with renal adysplasia. *M.R. Bekheirnia[1,3], Z. Niu[1], D.A. Scott[1], L. Potocki[1], P. Lurix[1], J.W. Belmont[1], D.J. Lamb[1,2,3], Y. Yang[1], C.M. Eng[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Center for Reproductive Medicine, Baylor College of Medicine, Houston, TX; 3) Scott Department of Urology, Baylor College of Medicine, Houston, TX.

Congenital Anomalies of Kidney and Urinary Tract (CAKUT) are the most important causes of pediatrics renal failure. While missense mutations in UPK3A gene are rare cause of renal adysplasia, nonsense mutations were not previously reported to be associated with the renal phenotype. Clinical whole exome sequencing is performed in the Medical Genetics Laboratory at Baylor College of Medicine as a clinical test. This study reports our observation of UPK3A mutations during the clinical sign-out of the requested cases. ACMG standards are used for classification of the genetic variants. We have identified a heterozygous (c.545G>A, p.W182X) nonsense mutations in UPK3A in 3 patients with a clinical diagnosis of either unilateral renal agenesis (2 patients) or renal dysplasia (1 patient). Both patients with unilateral renal agenesis inherited the p.W182X variant from an apparently unaffected mother. A parental sample was not available for analysis of the third patient who had renal dysplasia. All 3 patients had other organ involvement. We also identified the p.W182X variant in 4 more individuals who apparently did not have any kidney phenotype. Twelve missense, 2 splice site (c.571+1G>A) and one indel variants were also detected. Among the 12 patients with missense variants, one had echogenic kidneys together with prematurity, dysmorphic features and developmental delay. Of the 2 patients with a splice site variant (c.571+1G>A), 1 had echogenic kidneys and gonadal agenesis on pelvic ultrasound. While there were 2 earlier reports of missense mutations in UPK3A found in association with renal adysplasia, this is the first report of other types of mutations, such as nonsense variants, that are likely to be related to the renal adysplasia phenotype. Thus, a possible loss of function mechanism resulting from a mutation in UPK3A may cause renal adysplasia. Although the nonsense variants display a higher penetrance than the missense variants, in 2/3 of affected cases the variants were inherited from apparently unaffected parents. The role of the uroplakin gene mutations in CAKUT phenotype clearly requires further investigation. Acknowledgement: Supported in part by K12 DK0083014, the Multidisciplinary K12 Urologic Research (KURe) Career Development Program to DJL (MRB is a KURe Scholar).

3034F

Molecular Etiology of the Mayer-Rokitansky-Küster-Hauser (MRKH) Syndrome. *A. Ekici[1], C. Büttner[1], P. Strissel[2], P. Oppelt[2], S. Renner[2], M. Beckmann[2], R. Strick[2].* 1) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 2) Department of Gynecology and Obstetrics, Laboratory for Molecular Medicine, University-Clinic Erlangen, Germany.

The Mayer-Rokitansky-Küster-Hauser (MRKH) syndrome describes women with congenital aplasia of the uterus, cervix and the upper two-thirds of the vagina due to absent paramesonephric (Müllerian) ducts. Women with MRKH have a normal female karyotype (46, XX), normal female external development, internal normally regressed mesonephric (Wolffian) ducts and have correctly timed pubarche and thelarche. The incidence of the MRKH syndrome is one in 4,500 female newborns. MRKH patients represent the second most frequent cause of primary amenorrhea after Turner syndrome and commonly have associated malformations, like skeletal and renal abnormalities. The MRKH syndrome occurs isolated as a genital malformation (type 1), but also with associated malformations (especially malformations of the kidneys and urinary tract (type 2) and MURCS association (Müllerian renal and cervicothoracic somite abnormalities; partial or complete absence of the spinal column, ribs, or arms; asymmetric or improperly developed ribs or arms). To date the genetic or molecular etiology of the MRKH syndrome is completely unknown. In the present project we describe an integrative analysis of three high-throughput analysis data sets from two pairs of discordant monozygotic twins and their family members. First we performed copy number variation analysis with high density SNP microarrays. Second we performed exome sequencing of these sample set, which have not yet been performed according to the literature. In addition, we sequenced the whole methylome of all samples to round out the picture. Even though we did not find any distinct differential variants between affected and non-affected subjects if we analyzed the single data category, we could identify differential methylation pattern in seven loci with candidate genes if we applied a combined genetic-epigenetic disease model for the data analysis. We believe that the methods in the present application will help to unravel the genetic or molecular causes of MRKH and be of clinical relevance in order to help clinicians for an accurate molecular diagnosis and council patients and their families in order to understand the syndrome.

3035W

Variable phenotype in individuals with a *MUC1* mutation causing medullary cystic kidney disease type1. *S. Kmoch[1], A.J. Bleyer[2], K. Kidd[2], K. Hodanova[1], M. Zivna[1], P. Vyletal[1], H. Hartmannova[1], V. Stranecky[1].* 1) Inst Inherited Metabolic Dis, Charles Univ 1st Faculty, Prague 2, Czech Republic; 2) Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC 27157.

A cytosine insertion in the VNTR of the *MUC1* gene was recently identified as the most common cause of medullary cystic kidney disease type 1. The phenotypic expression of *MUC1* mutations have not been well characterized. Methods: Genotyping and evaluation of clinical characteristics were performed on families with a history of autosomal dominant interstitial kidney disease. Results: Twenty-five families were identified with *MUC1* mutation. Of 179 family members undergoing mutational analysis, a mutation was identified in 95 individuals, and 84 individuals did not have a mutation. There were 110 individuals identified as historically affected. Individuals with a *MUC1* mutation suffered from chronic kidney failure with a widely variable age of onset of end-stage kidney disease, ranging from 16 years to greater than 80 years.

Age Range	<20	20 to <40	40 to <60	60 to <80	>=80
Percent starting dialysis	1	40	38	20	1

There appeared to be two types of affected families. In 13 families, the median age of kidney failure was approximately 30, and family members started dialysis between 18 and 49. In the other 12 families, median age of onset ranged from 41 to 68, with age of kidney failure ranging from 22 to 81. Ultrasounds of 38 individuals showed no medullary cysts. While *MUC1* is expressed in many tissues, there were no clinical manifestations of the *MUC1* mutation detected in the breasts, skin, respiratory or gastrointestinal tract. Conclusions: Individuals *MUC1* mutations have highly variable expression, with some individuals proceeding to dialysis in their teenage years, while others remain dialysis-free at greater than 80 years. While *MUC1* is expressed in many tissues, it only results in clinical abnormalities in the kidney. Gene-gene or gene-environment interactions likely contribute to phenotypic variability.

3036T

Unbiased Next Generation Sequencing analysis confirms the existence of autosomal dominant Alport syndrome. *C. Fallerini[1], L. Dosa[1, 2], D. Giachino[3], R. Tita[1], M. Baldassarri[1], D. Del Prete[4], S. Feriozzi[5], G. Gai[6], M. Clementi[7], A. La Manna[8], N. Miglietti[9], R. Mancini[2], G. Mandrile[3], R. Artuso[1], G.M. Ghiggeri[10], G. Piaggio[10], F. Brancati[11], L. Diano[11], E. Frate[12], A.R. Pinciaroli[13], M. Giani[14], P. Castorina[14], E. Bresin[15], F. Mari[1,2], M. Bruttini[1,2], M. De Marchi[3], F. Ariani[1], A. Renieri[1,2].* 1) Medical Genetics, University of Siena, Siena, Italy; 2) Medical Genetics, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 3) San Luigi Medical Genetics, University of Torino, Torino, Italy; 4) Department of Medicine, Nephrology Clinic, University of Padova, Padova, Italy; 5) Nephrology and Dialysis Unit, Hospital Belcolle, Viterbo, Italy; 6) S.C. Medical Genetics A.O. City of Health and Science, Torino, Italy; 7) Clinical Genetics Unit, Department of Pediatrics, University of Padova, Padova, Italy; 8) Department of Pediatrics, Second University of Napoli, Napoli, Italy; 9) Nephrology DH, Pediatric Clinic of the University of Brescia, Brescia, Italy; 10) Nephrology Dialysis and Transplantation Unit, Institute 'G Gaslini', Genova, Italy; 11) Medical Genetics, University Hospital 'Tor Vergata', Roma, Italy; 12) Medical Genetics, Azienda Unita Locale Socio-Sanitaria n° 9, Treviso, Italy; 13) Nephrology and Dialysis S.O.C., Department of General Medicine and Medical Specialties 'Giuseppe Spada', A.O. Pugliese-Ciaccio, Catanzaro, Italy; 14) U.O.C. PEDIATRIC NEPHROLOGY AND U.O.C. Pediatric Nephrology and Hemodialysis, Ospedale Maggiore Policilinico, Milano, Italy; 15) Clinical Research Center for Rare Diseases 'Aldo e Cele Daccò', Ranica, Italy.

The story of the mode of transmission of Alport syndrome (ATS) was very troubled. In 1927, the disease was hypothesized as autosomal dominant condition with male segregation distortion. In 1990, the discovery of mutations in COL4A5 gene on the X chromosome completely changed this orientation and by that time ATS was considered an X-linked semidominant condition. Later on, an autosomal recessive form due to either COL4A3 or COL4A4 mutations was identified and it was calculated to account for about 10% of cases. A dominant form was testified more recently by the description of some large pedigrees but the real existence of this form is still questioned by many and its exact prevalence is unknown. The introduction of Next Generation Sequencing allowed us to perform an unbiased simultaneous analysis of the 3 genes(454 GS Junior, Roche) in a large cohort of Italian families (90) with clinical suspicion of ATS. In 48 of them a mutation was identified (53%). The subsequent segregation analysis in 161 family members allowed to clarify that the prevalent form was indeed the X-linked one (65%) and the rarer form the recessive one (4%). Most interestingly, in a relevant percentage of cases corresponding to 31% (15 families) the mode of transmission was autosomal dominant. In these cases either a COL4A3 or a COL4A4 heterozygous mutation segregating with the disease in each generation was found. The autosomal dominant form must therefore be seriously taken into account in all pedigrees in which patients are present in subsequent generations. The use of an unbiased one step approach was crucial for the rapid diagnosis in 20 cases where clinical data and family history would not have been sufficient to split patients into a precise mode of inheritance and select the specific gene. Moreover, in one case classified as likely X-linked, we found two COL4A3 mutations in compound heterozygosity. The autosomal recessive form is less represented in this cohort and this may be due to the low inbreeding rate of the population studied. In conclusion, this study illustrates how NGS can assist clinicians in making the correct diagnosis with significant reduction of turnaround time and positive consequences on patient management. The availability of a rapid and cost effective test encourages the molecular diagnosis also in oligosymptomatic children and helps their treatment delaying renal failure and improving life expectancy in a time-dependent manner.

3037F

A novel truncating mutation in SOX18 dramatically impairs renal function in the hypotrichosis-lymphedema-telangiectasia syndrome. *S. Moalem[1], P. Brouillard[2], D. Kuypers[3], E. Legius[4], E. Harvey[5], M. Francois[6], M. Vikkula[7], C. Chitayat[1,8].* 1) Department of Pediatrics, Division of Clinical and Metabolic Genetics, University of Toronto, Toronto, Ontario, Canada; 2) Laboratory of Human Molecular Genetics, de Duve Institute Université catholique de Louvain, Brussels, Belgium; 3) Department of Nephrology and Renal Transplantation, University Hospital Gasthuisberg, Leuven, Belgium; 4) Department of Pediatrics, University Hospital Gasthuisberg, Leuven, Belgium; 5) Department of Pediatrics, Division of Nephrology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 6) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; 7) Walloon Excellence in Life Sciences and Biotechnology (WELBIO), Belgium; 8) Department of Obstetrics and Gynecology, The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada.

SOX18 mutations in humans have been previously associated with both recessive and dominant Hypotrichosis-Lymphedema-Telangiectasia Syndrome (HLTS) and have been identified in the HMG domain. We report two unrelated families with affected individuals carrying the same *SOX18* mutation: a living child and his stillborn brother from Canada, as well as a living Belgian boy. The two living index cases were diagnosed with HLTS and DNA analysis for the *SOX18* gene showed the same heterozygous C-to-A mutation, resulting in a premature truncation of the protein. In both families, parental DNA showed no detectable mutation. At young ages, the boy of the first family presented with renal failure and severe hypertension, while that of the second family was diagnosed with membranoproliferative glomerulonephritis and hypertension. With time, both boys had a slow progressive deterioration in their renal function, culminating in the need for kidney transplantation. The *SOX18* gene is well known to play a role in the formation of blood and lymphatic vessels, but those with homozygous mutations in this gene do not develop renal failure. Here, we report for the first time two independent cases of renal failure associated with the identical heterozygous mutation in *SOX18*, and thus propose that this disorder be named Hypotrichosis-Lymphedema-Telangiectasia-Renal Syndrome or HLTRS.

3038W

Exome sequencing for diagnosis of disorders of sex development. *E. Vilain[1], V.A. Arboleda[1], H. Lee[2], A. Eskin[1], S.F. Nelson[1,2], R.M. Baxter[1].* 1) Human Genetics, UCLA School of Medicine, Los Angeles, CA; 2) Pathology and Laboratory Medicine, UCLA School of Medicine, Los Angeles, CA.

Disorders of sex development (DSD) are congenital conditions with discrepancies between the chromosomal, gonadal and phenotypic sex of the individual. DSDs have historically been difficult to diagnose and cause great stress to the patients and their families. Development of high throughput genomic technologies are expanding our knowledge of the underlying mechanism of DSDs and opening new avenues for clinical diagnosis. Exome sequencing is poised to become the most effective method to screen all the known genes involved in DSD. We used exome sequencing to analyze 25 intractable cases of 46, XY DSD where all available clinical testing had been unable to generate a definitive genetic diagnosis. Data were analyzed using a gene list containing all genes with well-documented involvement in human DSD, and genes with less strong association with human sexual development (eg OMIM description contains the word 'sex'), or where data exists in animal models. This gene list approach identified a potential genetic diagnosis in 10 cases for a diagnostic yield of 40%. In five cases we identified a known variant or a type of variant considered highly likely to be damaging to the protein. Two individuals were found to have a known variant in *MAP3K1*, and one patient had a missense variant in *NR5A1* that had been associated only with hypospadias while our patient had a more severe phenotype. We also identified a homozygous premature stop codon in *LHCGR*, and a homozygous splice site variant in *STAR* that both fit well with the patients' phenotypes. In a further five cases we identified likely causative, novel, missense variants in known DSD genes. These variants had not been reported in databases of asymptomatic individuals, e.g. the Exome Variant Server, and were predicted to be damaging by in silico prediction algorithms (SIFT, PolyPhen and Condel). We found likely causative variants in *MAP3K1* (in addition to the two cases with a known variant), *WT1* (2 cases with different variants), *CHD7*, and *DHH* that have not previously been reported in human DSD cases. These data contribute further to our understanding of disease causing variants, expand genotype/phenotype correlations, and could be useful for future clinical surveillance. As the list of genes involved in sex determination and differentiation continues to expand we propose that clinical exome sequencing is the best first line diagnostic genetic test for individuals with a DSD.

3039T

RNA-seq gene expression profiling identifies MEK1/2 and SHP2 as positive regulators of chondrocyte terminal differentiation. *ME. Bowen[1], UM. Ayturk[1], W. Yang[2], ML. Warman[1].* 1) Orthopaedic Research Laboratories, Children's Hospital Boston, Boston, MA 02115; 2) Department of Orthopaedics, Brown University, Providence, RI 02903.

The genetic disorder, Metachondromatosis (MC), is caused by heterozygous loss-of-function mutations in *PTPN11*, which encodes the phosphatase SHP2. MC patients develop benign cartilage tumors, resembling exostoses and enchondromas, that likely arise following somatic 'second hit' mutations in *PTPN11*. In a normal growth plate, chondrocytes first proliferate, then undergo hypertrophy, and are subsequently replaced by bone. Since the enchondroma-like lesions in MC patients occur near the growth plate, they may arise from a cluster of growth plate chondrocytes that failed to undergo terminal differentiation and be replaced by bone. Thus, we chose to investigate the role of *PTPN11* in chondrocyte terminal differentiation. We used RNA-seq to identify genes differentially expressed after genetic inactivation of *Ptpn11* in primary murine chondrocytes induced to undergo terminal differentiation in pellet cultures. Furthermore, since SHP2 has been shown to positively regulate the RAF/MEK/ERK pathway in other cell types, we treated chondrocyte pellets with U0126, an inhibitor of MEK1/2, to determine the role of this pathway in chondrocyte terminal differentiation. We found that U0126-treated pellets had increased levels of transcripts associated with proliferative, pre-hypertrophic and early-hypertrophic chondrocytes, but had decreased levels of transcripts associated with late-hypertrophic chondrocytes, including multiple genes encoding proteases that degrade cartilage matrix components. Similar, but less substantial, changes in gene expression were observed after deletion of *Ptpn11*. We were also able to identify a set of MEK1/2-dependent or SHP2-dependent genes whose expression increased over time during the maturation of wild-type chondrocytes in pellet cultures. Many of these genes had not previously been associated with chondrocyte maturation, and thus represent potential novel regulators of terminal differentiation. In summary, our results suggest that SHP2 and the RAF/MEK/ERK pathway may negatively regulate the onset of hypertrophy, but positively regulate transition from early-hypertrophic chondrocytes to terminally differentiated chondrocytes. These data are consistent with the hypothesis that enchondroma-like lesions in MC patients arise due to a failure of chondrocyte terminal differentiation.

3040F

Absence of CyPB directly affects collagen folding and glycosylation, and indirectly affects helical hydroxylation by LH1, altering bone cross-link patterns. *W.A. Cabral[1], I. Perdivara[2], M.A. Weis[3], M. Terajima[4], A.R. Blissett[1], W. Chang[1], E.N. Makareeva[5], E.L. Mertz[5], S. Leikin[5], K.B. Tomer[2], D.R. Eyre[3], M. Yamauchi[4], J.C. Marini[1].* 1) Bone & Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD; 2) Laboratory of Structural Biology, NIEHS, NIH, Research Triangle Park, NC; 3) Orthopaedic Research Laboratories, University of Washington, Seattle, WA; 4) North Carolina Oral Health Institute, University of North Carolina, Chapel Hill, NC; 5) Section on Physical Biochemistry, NICHD, NIH, Bethesda, MD.

Cyclophilin B (CyPB), encoded by Ppib, is an ER-resident peptidyl-prolyl cis-trans isomerase (PPIase) that occurs both independently and as a component of the collagen prolyl 3-hydroxylation complex. CyPB is proposed to be the major PPIase catalyzing the rate-limiting step in collagen folding. Mutations in Ppib cause recessively inherited osteogenesis imperfecta type IX, a moderately severe to lethal bone dysplasia. To investigate the role of CyPB in collagen folding and modification, we generated Ppib-/- mice that recapitulate the OI phenotype. KO mice are small, with reduced femoral aBMD, BV/TV, and MOI. Ppib transcripts are absent in skin, fibroblasts, femora and calvarial osteoblasts, and CyPB is absent from KO OB and FB on western analysis. Only residual (2-11%) collagen 3-hydroxylation is detectable in KO cells and tissues. Collagen folds more slowly in the absence of CyPB, supporting its rate-limiting role in folding. However, total lysyl hydroxylation in Ppib-/- FB and OB collagen is normal, while increased diglycosylation accounts for collagen overmodification and delayed electrophoretic migration. Detailed analysis of bone and OB type I collagen by mass spectrometry revealed site-specific alterations of lysine hydroxylation by LH1. Most importantly, hydroxylation of crosslink residue K87 is reduced by 15-20% in OB collagen. In KO bone, there is dramatic underhydroxylation of helical crosslink residue K87 in both the α1(I) and α2(I) chains, decreased by 40% and 30% compared to WT, respectively. The altered lysine modification causes striking changes in the type and quantity of collagen crosslinks in KO bone. Unhydroxylated forms of divalent (HLNL) and trivalent (LP) crosslinks are increased 3- to 5-fold, leading to increased total crosslinks and decreased DHLNL/HLNL and HP/LP ratios. Altered crosslink patterns contribute to decreased collagen deposition into matrix by 70-80% and directly undermine bone strength. Finally, these findings suggest a novel role for CyPB in supporting the activity of specific collagen lysyl hydroxylase(s), especially LH1, by an unknown mechanism. Thus, CyPB not only facilitates collagen folding directly but also indirectly regulates collagen hydroxylation, glycosylation, crosslinking and fibrillogenesis.

3041W

Use of whole exome sequence analysis to identify TLE4 as a causal variant in a family with congenital kyphoscoliosis and prolonged patency of the anterior fontanelle. *P. Giampietro[1], A. Stoddard[2], D. Sweetser[3], C. Raggio[4], R. Blank[2], M. Stephan[5], K. Rasmussen[6], K. Gill[1], S. Sund[1], R. Lorier[2], A. Turner[2], U. Broeckel[2].* 1) University of Wisconsin-Madison, Madison, WI; 2) Medical College of Wisconsin, Milwaukee, WI; 3) Massachusetts General Hospital, Boston, MA; 4) Hospital for Special Surgery, New York, NY; 5) Madigan Healthcare System, Tacoma, WA; 6) Marshfield Clinic, Marshfield, WI.

Congenital vertebral malformations (CVM) represent defects in formation and segmentation of somites and have an estimated incidence of 0.13-0.50 per 1000 live births. Extreme genetic heterogeneity and the rarity of large families with CVM limit the ability to identify mutations in patterning genes associated with CVM by traditional genetic study designs. We therefore used whole exome sequencing (WES) to study a kindred in which a man and his two daughters suffered from kyphoscoliosis, vertebral body hypoplasia, short lumbar spinal pedicles, sacral and coccygeal hypoplasia and prolonged anterior fontanelle patency. WES was performed using Agilent SureSelect hybridization-based exome capture methodology. Results were filtered to excude all except heterozygous nonsynonymous coding variants with minor allele frequency < 1%. This strategy yielded 21 candidates, with *TLE4* harboring a c.A1318G:p.T440A variant. The functional significance of the *TLE4* variant was demonstrated by a mouse knockout of Tle4, the murine homolog. Tle-/- mice displayed a profound impairment of bone formation, while Tle+/- mice displayed a similar phenotype to the affected humans, including shortened vertebral pedicles and delayed mineralization of the skull. *TLE4* encodes a transcription factor in a gene family that participates in the determination of hematological cell fate in response to Notch signaling, a critical signaling pathway in somitogenesis.The Gro/TLE gene family was not known to involved in skeletal development before the identification of *TLE4* as a strong candidate gene in this family and the simultaneous characterization of Tle4 knockout mice. The *Gro/TLE* gene family serves as downstream effectors of Notch signaling, as well as inhibitors of Wntβ-catenin signaling. Disruption of these pathways is seen in certain families with CVMs. In addition, the *Gro/TLE* gene family is also known to interact with *RUNX* family members, with RUNX2 being required for calcification. These findings highlight the potential of whole exome sequence analysis, particularly when coupled with functional studies in appropriate models, to identify novel genetic determinants of uncommon disorders. While further work remains to be done to fully define the mechanisms by which *TLE4* mutations produce the observed phenotypes, these data nevertheless demonstrate the ability of a deep sequencing strategy to discover previously unknown skeletal genes.

3042T

Exome sequencing reveals INPPL1 mutations in Opsismodysplasia.
C. HUBER[1], E.A. FAQEIH[2], D. BARTHOLDI[3], C. BOLE-FEYSOT[4], Z. BOROCHOWITZ[5], D.P. CAVALCANTI[6], A. FRIGO[1], P. NITSCHKE[7], J. ROUME[8], H.G. SANTOS[9], S.A. SHALEV[10], A. SUPERTI-FURGA[11], A.L. DELEZOIDE[12], K.M. GIRISHA[13], M. WRIGHT[14], M. LE MERRER[1], A. MUNNICH[1], V. CORMIER-DAIRE[1]. 1) Département de Génétique, Unité INSERM U781, Université Paris Descartes-Sorbonne Paris Cité, Fondation Imagine, Hôpital Necker Enfants Malades, Paris, 75015, France; 2) Medical Genetics, King Fahad medical city, Children's Hospital, Pediatric deparment P.O. Box 59046 Riyadh 11525, Kingdom of Saudi Arabia; 3) Institute of Medical Genetics, University of Zurich, Schorenstrasse 16, CH 8603 Schwerzenbach, Switzerland; 4) Plateforme de génomique, Fondation IMAGINE, Paris, 75015, France; 5) The Simon Winter Institute for Human Genetics, Bnai-Zion Medical Center, Rappaport Faculty of Medicine and Research Institute, Technion-Israeli Institute of Technology, Haifa, 31048 Israel; 6) Grupo de Displasias Esqueléticas, Depto Genética Médica, FCM,Universidade Estadual de Campinas (UNICAMP), Campinas, SP 13081-970, Brasil; 7) Plateforme de Bioinformatique, Université Paris Descartes, Paris, 75015, France; 8) Unité de Génétique Médicale, Service de Cytogénétique et de Biologie de la Reproduction, CHI Poissy - St Germain - en- Laye, 78100, France; 9) Department of Medical Genetics, Serviço de Genética Médica Hospital Universitário S. Maria, 1649-035 Lisboa, Portugal; 10) Ha'Emek Medical Center, Afula 18101, and the Rapapport Faculty of medecine, Technio, Haifa 31096, Israel; 11) University of Lausanne, Centre Hospitalier Universitaire Vaudois (CHUV) 1011 Lausanne, Switzerland; 12) Service de Biologie de Développement, Université Paris Diderot, Hôpital, Robert Debré, Paris, 75019, France; 13) Department of Medical Genetics and Pediatrics, Kasturba Medical College, Manipal 576104, Udupi district, Karnataka, India; 14) Institute of Human genetics, International Center for lofe, Newcastle upon Tyne, NE1 3BZ, United Kingdom.

Opsismodysplasia (OPS) is a severe autosomal-recessive chondrodysplasia, belonging to the spondylodysplastic dysplasia group. The main clinical and radiological features are y pre- and postnatal micromelia, extremely short hands and feet, severe platyspondyly, squared metacarpals, delayed skeletal ossification, and metaphyseal cupping. In order to identify mutations causing OPS, a total of 19 cases from 12 unrelated families were included in this study. We performed exome sequencing in three cases from three unrelated families and only one gene was found to harbor mutations in all three cases, namely INPPL1 (inositol polyphosphate phosphatase like 1). Screening INPPL1 in the remaining cases identified a total of 16 distinct mutations. Among the 19 cases, prenatal findings led to early termination of pregnancies (especially in recurrent sibs) in 10/19 cases and hygroma, short long bones, short extremities and narrow thorax were consistently observed. Four children died early (stillborn 30 WG- 15 months of age). The five remaining cases (3 to 19 years old), had normal cognitive development, severe short stature (below- 4 SDS), lower limb deformity and severe scoliosis with atlanto-axial instability. Most mutations (8/16) resulted in premature stop codons, located in the catalytic domain, 5-phosphatase. INPPL1 belongs to the inositol-1,4,5-trisphosphate 5-phosphatase family, that govern a plethora of cellular functions by regulating the levels of specific phosphoinositides. The role of INPLPL1 in bone and cartilage is yet unknown. However, histological studies of the femoral growth plate performed in the prenatal cases from 2 families, show similar disorganization of the growth plate with, absence of columnar arrangement of proliferative cells and a reduced hypertrophic zone with a small number of hypertrophic chondrocytes. Our finding supports a key and specific role of INPPL1 in endochondral ossification.

3043F

Novel mutations in the LRP5 gene in patients with osteoporosis-pseudoglioma syndrome.
M. Pekkinen[1], G. Grigelioniene[2], L. Akin[3], K. Shah[4], K. Karaer[5], S. Kurtoglu[3], A.V. Ekbote[4], E. Sagsak[6], J. Söderholm[1], S. Vallius[1], S. Danda[4], E. Åström[2], O. Mäkitie[1]. 1) Folkhälsan Inst Genetics, Folkhälsan Inst Genetics, Univ Helsink, Helsinki, University of Helsinki, Finland; 2) Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; 3) Erciyes University, Faculty of Medicine, Department of Pediatric Endocrinology, Turkey; 4) Department of Clinical Genetics, Christian Medical College and Hospital Vellore, India; 5) Intergen, Genetic Diagnosis Research and Application Center, Ankara, Turkey; 6) Dr.Sami Ulus Children's Hospital, Department of Pediatric Endocrinology, Ankara,Turkey.

Background: Osteoporosis-pseudoglioma syndrome (OPPG) is a rare autosomal recessive disorder with congenital or early-onset blindness and severe osteoporosis. OPPG is caused by biallelic mutations in the low-density lipoprotein receptor-related protein 5 (LRP5) gene. We present six novel LRP5 mutations and the resulting phenotypes in four consanguineous Indian and Turkish families. Methods: Peripheral blood samples were obtained from the affected probands and their parents. DNA was extracted using standard procedures. The LRP5 gene was analyzed by direct sequencing after PCR amplification. The observed sequence changes were compared with reference database and 200 control samples. Results: Altogether seven patients from three Turkish and one Indian families were included in the study; in all families the parents were related. All patients had severe osteoporosis with peripheral and vertebral fractures, and congenital or early-onset blindness. In addition, three patients had delayed mental development. DNA sequencing demonstrated in each of the four probands homozygous LRP5 mutations. All mutations were novel and located in exons 1, 2, 3 and 12 of the LRP5 gene; they were regarded deleterious based on prediction programs. In one family the affected child had homozygous sequence variations A3A, A4A, P5L and a deletion of the rs72555376 microsatellite; the healthy parents had none of these changes. In three families with altogether six affected children we found homozygous missense mutations D116N, P197R and I882N; the parents were heterozygous for the changes. Mutations P197R in exon 3 and I882N in exon 12 were associated with mental retardation. Conclusions: We report six novel LRP5 mutations in Turkish and Indian patients with OPPG. All mutations regardless of the location resulted in similar ocular and skeletal phenotype. Mutations in exons 3 and 12 resulted also in developmental delay. The deletion of rs72555376, which causes shortening of a poly-leucine stretch in exon 1, has previously been described in heterozygous state by Chung et al. 2008; the disease mechanism and origin of the homozygous deletion together with 3 other sequence changes in our patient remains to be elucidated in further studies.

3044W

Altered Osteoblast Function underlies phenotype of type V Osteogenesis Imperfecta. *A. Reich[1], A.S. Bae[1], A.M. Barnes[1], W.A. Cabral[1], D. Chitayat[2], S.C. Hill[3], J.C. Marini[1].* 1) Bone and Extracellular Matrix Branch, NICHD, NIH, Bethesda, MD; 2) 2The Hospital for Sick Children, and the Department of Obstetrics and Gynecology, the Prenatal Diagnosis and Medical Genetics Program, Toronto, Ontario, Canada; 3) 3Diagnostic Radiology Department, Clinical Center, NIH, Bethesda, MD.

Osteogenesis imperfecta (OI) is a genetically heterogeneous disorder characterized by bone fragility. Type V OI is the only OI type with dominant inheritance not caused by mutations in type I collagen; furthermore, steady-state collagen from type V osteoblasts has normal gel mobility. Affected individuals have characteristic skeletal findings, including hypertrophic callus and ossification of the interosseus membrane, as well as mesh-like lamellation on bone histology. Type V OI is caused by a unique heterozygous mutation in IFITM5 (c.-14C>T), which encodes Bril, a transmembrane protein expressed in osteoblasts. The mutation generates a start codon, adding five residues to the Bril N-terminus. However, the mechanism of type V OI and its relationship with type I collagen is unknown. We identified 8 patients with the IFITM5 (c.-14C>T) mutation, with variable phenotypic expression. Osteoblasts from type V OI patients were differentiated with osteogenic media in culture over 21 days. Sequencing of cDNA from these cells verified expression of mutant IFITM5 transcripts, consistent with a dominant mechanism. In both control and type V OI differentiated osteoblasts, Bril expression increases substantially and comparably with BMP2 stimulation. Bril protein is moderately decreased in treated mutant cells. Osteocalcin (BGLAP2) expression, a marker of late osteoblast differentiation, was increased in type V osteoblasts vs control during days 10-15 of differentiation, and relatively augmented by addition of BMP2 to cultures. Mineralization, assayed by alizarin red deposition, was also increased in type V osteoblasts during differentiation, vs control. In contrast, type V OI osteoblasts had less than half the COL1A1 expression of control during differentiation. Matrix deposited by proband osteoblasts in long-term culture displayed normal amounts of collagen when examined by confocal microscopy, but with a more sheet-like organization than in control. Finally, comparison of osteoblasts transcripts from two type V patients and control on osteogenic arrays revealed common sets of significantly altered transcripts for mineralization and growth factors, collagens and BMPs. Thus, our data support a proposed role for regulation of Bril by BMP2. The increased mineralization and collagen-related defect demonstrated in type V OI osteoblasts during differentiation may underlie the overactive tissue calcification and hypertrophic callus formation seen in patients.

3045T

Whole Exome Sequencing and functional follow up in 114 cases of non-motile ciliopathies (Jeune-Asphyxiating Thoracic Dysplasia (JATD), Short-Rib-Polydactyly-Syndrome (SRPS), Bardet-Biedl-Syndrome (BBS), Joubert Syndrome (JS) and others). *M. Schmidts[1], E. Chanudet[2], A.M. McInerney-Leo[3], C. Cortes[4], V. Plagnol[5], F. Lescai[2], S. Christou[1], H. Jungbluth[6], G. Haliloglu[7], H. Kayserili[8], N. Elcioglu[9], B. Tuyusz[10], M.E. Hurles[11], P.J. Scambler[1], A. Zankl[3,12], M.A. Brown[3], E. Duncan[3], U. K10K[13], C. Wicking[4], P.L. Beales[1,2], H.M. Mitchison[1].* 1) Molecular Medicine Unit, Institute of Child Health, University College London (UCL), London, UK; 2) Centre for Translational Genomics-GOSgene, Institute of Child Health, UCL London, UK; 3) The University of Queensland Diamantina Institute, Translational Research Institute, Woolloongabba, Queensland, Australia; 4) Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia; 5) Department of Genetics, Environment and Evolution, UCL Genetics Institute (UGI), University College London, London, UK; 6) Department of Paediatric Neurology, Evelina Children's Hospital, Guy's and St Thomas' National Health Service (NHS) Foundation Trust, London, UK; 7) Department of Pediatric Neurology, Hacettepe University Children's Hospital, Ankara, Turkey; 8) Medical Genetics Department, Medical Faculty, Istanbul University, Istanbul, Turkey; 9) Department of Pediatric Genetics, Marmara University Hospital, Istanbul, Turkey; 10) Department of Pediatrics, Division of Genetics, Cerrahpasa Faculty of Medicine, Istanbul University, Istanbul, Turkey; 11) Wellcome Trust Sanger Institute, Hinxton, UK; 12) The University of Queensland, UQ Centre for Clinical Research, Herston, Queensland, Australia; 13) www.uk10k.org.

Background: Ciliopathies are complex develop-mental disorders resulting from mutations in genes encoding ciliary proteins. Although individually rare, together they represent a significant disease burden. Phenotypic as well as genetic overlap, together with extensive genetic heterogeneity, has hampered genetic diagnosis in the past but develop-ment of next generation sequencing techniques such as whole exome sequencing (WES) offers new dia-gnostic tools.**Methods:** We investigated 114 ciliopathy cases (61 JATD and SRPS, 37 BBS, 7 JS and 9 other ciliopathy cases) by WES. **Results:** We identified the disease-causing gene in 2/3 of all cases, with 50% of JATD and Joubert cases and up to 90% of BBS cases found to be caused by mutations in known disease causing genes. Copy number variations were detcted in approximately 10% of solved cases from the WES data.WES results also led to revision of the initial clinical diagnosis in 10% of the cases and new phenotype-genotype associations came to light. However, the proportion of cases given a genetic diagnosis was much lower in the more ambiguous phenotypic groups. In contrast to BBS cases, in whom we frequently identified two loss-of-function alleles per proband, all JATD cases carried at least one missense allele, consistent with the hypothesis that JATD patients are hypomorphic for the disease while BBS represents a true 'null' phenotype. 25% of the BBS cases carried >2 mutations, but the third mutation was always a missense allele except in one case. We were able to identify several new genes causing JATD, including 2 genes encoding WD40-repeat containing proteins (new intra-flagellar transport (IFT) dynein components previously undescribed in mammals) in addition to mutations in genes encoding other better understood IFT-subcomplex B components. Functional follow-up studies of these new genes using zebrafish and cell culture model systems are currently under way. **Summary:** Our findings in this large non-motile ciliopathy cohort demonstrate that WES is highly efficient for the genetic diagnosis of heterogeneous recessive disorders and that this diagnosis is facilitated by deep pheno-typing. A surprisingly large number of cases are caused by mutations in known genes, but we also identified several new ciliopathy genes. Compared to NGS gene panel sequencing, WES offers additional opportunities to identify new genes previously not associated with the condition investigated.

3046F

Haplotype Analysis Supports a "Founder" for the Balkan OPG Mutation Causing Juvenile Paget's Disease. *S. Mumm[1,2], K. Geczi[1], M. Huskey[1], D. Naot[3], S. Polyzos[4], T. Cundy[3], W. Van Hul[5], P. Singhellakis[6], M.P. Whyte[1,2].* 1) Washington University School of Medicine, St. Louis, MO, USA; 2) Shriners Hospital for Children, St. Louis, MO, USA; 3) University of Auckland, New Zealand; 4) Aristotle University of Thessaliniki, Greece; 5) University of Antwerp, Belgium; 6) Athens University, Greece.

Juvenile Paget's disease (JPD), a rare autosomal recessive disorder, features extremely rapid bone turnover causing skeletal pain, fracture, and deformity in early childhood. Deafness and retinopathy leading to blindness can follow. Most JPD is due to homozygous loss-of-function mutations in the TNFRSF11B gene encoding osteoprotegerin (OPG) -- the decoy receptor that prevents RANKL binding to its cognate receptor RANK, a major stimulus element for osteoclastogenesis. A variety of mutations in TNFRSF11B, typically transmitted by apparent 'founders' in various geographic locations worldwide, cause JPD. The severity of 'OPG deficiency JPD' seems to vary depending on mutation type and how much of the OPG coding region is disrupted. In two Greek men and another Croatian man who are seemingly unrelated and manifest the mildest form of JPD, a unique homozygous deletion/insertion mutation (966_969delTGACinsCTT) has been reported. This frame-shift deletes 79 carboxy terminal amino acids from the OPG monomer, including a cysteine necessary for homodimerization. When recently referred 2 additional unrelated Greek JPD patients homozygous for this 'Balkan mutation', we studied the likelihood that all of these patients are in fact related by a distant 'founder' in whom the mutation arose many generations ago. To test this hypothesis, we performed haplotype analysis of the 5 Balkan JPD patients and 2 control JPD patients from other geographic regions using SNPs within and surrounding TNFRSF11B. We developed primer sets for 13 informative SNPs, and then PCR-amplified and sequenced these SNPs using leukocyte-derived genomic DNA from each patient. All 5 individuals shared a homozygous common core haplotype of 4 SNPs, whereas each of the 'control' JPD patients had a unique haplotype. Two of the four Greek patients shared a common haplotype of 10 homozygous SNPs, indicating closest kinship. A different Greek patient shared only 5 of these SNPs suggesting more distant relationship. The Croatian JPD patient shared only 4 common markers suggesting he is the most remotely related and consistent with his geographic distance. Hence, all Balkan JPD patients share the common OPG mutation (966_969delTGACinsCTT) and common haplotypes indicating relationship to a genetic founder. Further analysis can determine how many generations ago this founder mutation arose.

3047W

Exome analysis on tubular aggregate myopathy. *Y. Endo[1,2], K. Motomura[1], Y.K. Hayashi[1,2], S. Noguchi[2], I. Nonaka[2], M. Mori-Yoshimura[3], Y. Oya[3], I. Nishino[1,2].* 1) Department of Clinical Development, Translational Medical Center, National Center of Neurology and Psychiatry (NCNP), Kodaira, Japan; 2) Department of Neuromuscular Research, National Institute of Neuroscience, NCNP; 3) Department of Neurology, National Center Hospital, NCNP.

Tubular aggregates (TAs) are a unique structure in muscle fibers, which is derived from sarcoplasmic reticulum, and are seen in certain types of muscle diseases including periodic paralysis, congenital myasthenic syndromes (CMS) and tubular aggregate myopathy (TAM). TAM is defined as unclassified hereditary myopathy, pathologically characterized by the presence of TAs, although causative relationship has not been well established. The objective of this study is to identify pathogenic mutations of TAM and to clarify the clinical phenotypes. We performed whole-exome sequencing in 13 individuals with TAM but with no known genetic defects, and identified pathogenic mutations in *GFPT1* in two individuals. Both individuals had novel homozygous missense mutations, c.41G>A (p.Arg14Gln) and c.723_724insG (p.G241fs), respectively. *GFPT1* was recently reported as a causative gene for limb-girdle CMS (LG-CMS). However, the clinical presentations of the two affected individuals were slightly different from LG-CMS. Individual 1 exhibited slowly progressive limb-girdle muscle weakness with no fluctuation therefore the clinical diagnosis was limb-girdle muscle dystrophy. Individual 2 exhibited limb-girdle and facial muscle weakness from infancy. Therefore, the clinical diagnosis was congenital myopathy. Muscle biopsy showed myopathic change with TAs although the first biopsy of individual 2 performed at age 8 years showed only minimal change without TA. Repetitive stimulation performed in individual 1 after genetic analysis demonstrated the decremental responses diagnostic of myasthenia. Acetylcholine-esterase inhibitor therapy was dramatically effective, improving his status from wheelchair-bound to ambulant without support. *GFPT1*-associated LG-CMS should be considered when affected individuals have limb-girdle muscle weakness and TAs.

3048T

Novel *COL11A1* mutations in Stickler syndrome detected by next-generation sequencing. *F. Acke[1,2], F. Malfait[2], O. Vanakker[2], W. Steyaert[2], K. De Leeneer[2], I. Dhooge[1], A. De Paepe[2], E. De Leenheer[1], P. Coucke[2].* 1) Department of Otorhinolaryngology, Ghent University Hospital, Ghent, Belgium; 2) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium.

Stickler syndrome is a connective tissue disorder with considerable phenotypic and genotypic variability, characterized by distinctive facial abnormalities, ocular problems, hearing loss, and joint problems. Up to now, mutations in 5 different collagen genes have been associated with the disease. Mutations in the *COL11A1* gene may result in Stickler syndrome type 2, which can be differentiated clinically by a 'beaded' vitreous and a more severe hearing loss compared to the more prevalent type 1, caused by mutations in *COL2A1*. In literature, only a few dozen *COL11A1* mutations have been reported. We selected 33 unrelated *COL2A1* mutation-negative patients, based on their clinical features, for molecular *COL11A1* analysis, by means of next-generation sequencing technology (MiSeq, Illumina). The bioinformatic pipeline included the CLC bio Workbench 6.0 followed by an in-house developed software package for variant interpretation. Assays lacking sufficient coverage as well as the mutations identified were verified by Sanger sequencing. Out of the 33 selected probands, disease-causing mutations could be identified in 14 independent patients. Half of these mutations are novel, of which most are missense mutations (including 2 glycine substitutions). The other half are previously reported mutations, mostly splice-site alterations. Moreover, additional SNPs were identified. All molecularly confirmed patients had a clinical presentation compatible with Stickler syndrome type 2. In conclusion, *COL11A1* screening in selected Stickler syndrome patients, based on meticulous clinical evaluation, is rewarding. For large genes such as *COL11A1*, next-generation sequencing proves to be an efficient and cost-effective molecular tool.

3049F

Mutation analysis of COL1A1 and COL1A2 genes in Indian patients with Osteogenesis Imperfecta. *J. Stephen[1], A. Shukla[1], A. Dalal[2], G. Katta[3], N. Gupta[4], M. Kabra[4], P. Dabadghao[5], S. Phadke[1].* 1) Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Science, Lucknow, India; 2) Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 3) Department of Medical Genetics, Kasturba Medical college, Manipal, India; 4) All India Institute of Medical Science, New Delhi, India; 5) Endocrinology, Sanjay Gandhi Post Graduate Institute of Medical Science, Lucknow, India.

Osteogenesis Imperfecta (OI) is a condition of decreased bone density of heterogeneous etiology. Most of the cases are inherited in autosomal dominant fashion and are caused by mutations in COL1A1 or COL1A2 genes. These two genes are very large so they have been rarely analyzed systematically in Indian patients with OI. We have selected 35 Indian patients who were clinically diagnosed for OI and sequenced all exons of both the genes. Mutations in COL1A1 gene were identified in 14 cases (six novel) and mutations in COL1A2 were identified in eleven cases (seven novel). Total 55 polymorphisms have been identified in both the genes; out of them eight were novel in the coding region and twelve were novel in the non coding region. No mutation has been detected in ten patients; out of them six were from consanguineous families with one or two similarly affected siblings suggesting autosomal recessive genes. Excluding consanguineous families, mutation has been identified in 25 out of 29 families which accounts for about 86 percent of mutation detection rate in these genes in Indian patients with OI from non consanguineous families. Out of identified mutations, 56 percent were in COL1A1 and 44 percent were in COL1A2 which is similar to the reported rate worldwide.

3050W

WNT1 Mutations in Early-onset Osteoporosis and Osteogenesis Imperfecta. *P. Campeau[1], C. Laine[2,3], K.S. Joeng[1], R. Kiviranta[4,5], K. Tarkkonen[4], M. Grover[1], J.T. Lu[6,7], M. Pekkinen[2], M. Wessman[2,8], T.J. Heino[9], V. Nieminen-Pihala[4], M. Aronen[2], T. Laine[10], H. Kröger[11], W.G. Cole[12], A.E. Lehesjoki[2,13,14], L. Nevarez[15], D. Krakow[16,17], C.J.R. Curry[18], D.H. Cohn[15,16], R.A. Gibbs[1,6], O. Mäkitie[10,19], B.H. Lee[1,20]*. 1) Molec & Human Gen, Baylor College Med, Houston, TX; 2) Folkhälsan Institute of Genetics, Helsinki, FINLAND; 3) Department of Endocrinology, Sahlgrenska University Hospital, Gothenburg, SWEDEN; 4) Department of Medical Biochemistry and Genetics and Department of Medicine, University of Turku, Turku, FINLAND; 5) Department of Medicine, Turku University Hospital, Turku, FINLAND;; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA;; 7) Department of Structural and Computational Biology & Molecular Biophysics, Baylor College of Medicine, Houston, TX, USA; 8) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, FINLAND;; 9) Department of Cell Biology and Anatomy, University of Turku, Turku, FINLAND; 10) Department of Pediatric Orthopedic Surgery, Sahlgrenska University Hospital, Gothenburg, SWEDEN;; 11) Bone and Cartilage Research Unit, University of Eastern Finland and Kuopio University Hospital, Kuopio, FINLAND; 12) Division of Pediatric Surgery, University of Alberta, Edmonton, CANADA;; 13) Haartman Institute, Department of Medical Genetics and Research Program's Unit, Molecular Medicine, University of Helsinki, Helsinki, FINLAND;; 14) Neuroscience Center, University of Helsinki, Helsinki, FINLAND;; 15) Department of Molecular, Cell and Developmental Biology, University of California-Los Angeles, CA, USA;; 16) Department of Orthopaedic Surgery, University of California-Los Angeles, CA, USA;; 17) Department of Human Genetics, University of California-Los Angeles, CA, USA;; 18) University of California San Francisco/Genetic Medicine Central California, Fresno, CA, USA; 19) Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, FINLAND;; 20) Howard Hughes Medical Institute, Houston, TX, USA;.

This report identifies human skeletal diseases associated with mutations in *WNT1*. In ten family members with dominantly inherited early-onset osteoporosis, a heterozygous missense variation c.652T>G (p.Cys218Gly) in *WNT1* segregated with the disease, and a homozygous nonsense mutation (c.884C>A, p.Ser295*) was identified in two siblings with recessive osteogenesis imperfecta. *In vitro*, aberrant forms of WNT1 protein showed impaired capacity to induce canonical WNT signaling, their target genes, and mineralization. *Wnt1* was clearly expressed in bone marrow, especially in B cell lineage and hematopoietic progenitors; lineage tracing identified expression in a subset of osteocytes, suggesting altered cross-talk of WNT signaling between hematopoietic and osteoblastic lineage cells in these diseases.

3051T

Identification of novel _SHOX_ target genes in the developing limb using a transgenic mouse model. *G.A. Rappold[1], A. Glaser[1], K. Kleinschmidt[2], I. Scholl[1], R. Röth[1], L. Li[3], N. Gretz[4], G. Mechtersheimer[5], M. Karperien[6], A. Marchini[1,7], W. Richter[2], K.U. Beiser[1]*. 1) Department of Human Molecular Genetics, Institute of Human Genetics, Heidelberg University Hospital, Heidelberg, Germany; 2) Division of Experimental Orthopaedics, Orthopaedic University Hospital, Heidelberg, Germany; 3) Institute of Pathology, University of Göttingen, Göttingen, Germany; 4) Medical Research Center (ZMF), Medical Faculty Mannheim at Heidelberg University, Mannheim, Germany; 5) Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany; 6) Department of Developmental Bioengineering, University of Twente, Enschede, The Netherlands; 7) German Cancer Research Center (DKFZ), Heidelberg, Germany.

Deficiency of the human short stature homeobox-containing gene (*SHOX*) has been identified in several disorders characterized by reduced height and skeletal anomalies such as Turner, Léri-Weill and Langer syndrome as well as idiopathic short stature. SHOX acts as a transcription factor during limb development and is expressed mainly in prehypertrophic and hypertrophic chondrocytes of the growth plates. Although highly conserved in vertebrates, rodents lack a *SHOX* orthologue. This offers the unique opportunity to analyze the effects of human *SHOX* expression in transgenic mice. We have generated a mouse expressing the human *SHOXa* cDNA under the control of a murine *Col2a1* promoter and enhancer. *SHOX* and marker gene expression as well as skeletal phenotypes were characterized in two transgenic lines. No significant skeletal anomalies were found in transgenic compared to wildtype mice. Quantitative and *in situ* hybridization analyses revealed that the *Col2a1-SHOX* transgene, however, affected extracellular matrix gene expression during early limb development, suggesting a role for *SHOX* in growth plate assembly and extracellular matrix composition during long bone development. For instance, we could show that the connective tissue growth factor gene *Ctgf*, a gene involved in chondrogenic and angiogenic differentiation, among other genes is transcriptionally regulated by SHOX in transgenic mice. This finding was confirmed in human NHDF and U2OS cells and chicken micromass culture, demonstrating the value of the *SHOX*-transgenic mouse for the characterization of SHOX-dependent genes and pathways in early limb development.

3052F

HIBCH deficiency in a patient with phenotypic characteristics of mitochondrial disorders. *M.S. Reuter[1], J.O. Sass[2], A.B. Ekici[1], I. Schanze[3], T. Leis[4], S. Uebe[1], R. Abou Jamra[1], A. Reis[1]*. 1) Institute of Human Genetics, Erlangen, Germany; 2) Clinical Chemistry and Biochemistry, University Children's Clinic Zurich, Switzerland; 3) Institute of Human Genetics, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany; 4) Department of Pediatrics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.

The catabolic pathway of the amino acid valine is dependent on a specific 3-hydroxyisobutyryl-CoA hydrolase (HIBCH), a nuclear-encoded mitochondrial protein. Only two patients with HIBCH deficiency and biallelic mutations in the HIBCH gene have been reported in the literature. Both of them demonstrated delayed motor development, muscular hypotonia, and early-onset deterioration of neurological functions. One of the patients additionally manifested various congenital malformations. The phenotype was attributed to an accumulation of methacrylyl-CoA, a highly reactive compound, which readily undergoes reactions with free sulfhydryl groups. We report a third patient, first child of healthy consanguineous parents, with a homozygous one-base pair insertion resulting in a premature stop codon (c.1129_1130insT, p.K377_S378delinsX) in the HIBCH gene (NM_014362). The mutation was detected by homozygosity mapping and whole-exome sequencing on a SOLID 5500xl platform. The patient displays a variety of symptoms reminiscent of mitochondrial dysfunction; like severe muscular hypotonia, intellectual disability, Leigh-like hyperintensities of the basal ganglia in MRI, progressive brain atrophy, repeatedly elevated blood lactate, optic nerve atrophy, seizures, and respiratory chain complex 1 deficiency with borderline depletion of mitochondrial DNA in native muscle tissue. There were no organ or brain malformations. Metabolic screening for aminoacidopathies and organic acidurias yielded normal results. The phenotype of the patient further elucidates the clinical spectrum of HIBCH deficiency. Physical malformations do not appear to be a consistent feature. Instead, the clinical features described here suggest that mitochondrial dysfunction, presumably caused by an accumulation of toxic valine metabolites, contributes to the clinical phenotype. We suggest that HIBCH deficiency should be considered as a differential diagnosis in patients with suspected mitochondrial disorders, particularly with regard to the potential treatability of amino acid metabolism disturbances.

3053W

Discovery of a novel genetic disorder and increased diagnostic rate using Next Generation Sequencing in heterogeneous Ataxias. *A.H. Németh[1,2,3], A.C. Kwasniewska[1,3], S. Lise[3], R. Parolin Schnekenberg[3,4], E.B.E. Becker[5], K.D. Bera[5], M.E. Shanks[3, 16], L. Gregory[3], D. Buck[3], M.Z. Cader[1,5], K. Talbot[1,5], R. de Silva[6], N. Fletcher[7], R. Hastings[8], S. Jayawant[9], P.J. Morrison[10], P. Worth[11], M. Taylor[12], J. Tolmie[13], M. O'Regan[14], R. Valentine[15], E. Packham[16], J. Evans[16], P. Clouston[16], A. Seller[16], J. Ragoussis[3,17], UK Ataxia Consortium.* 1) Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom; 2) Department of Clinical Genetics, Churchill Hospital, Oxford University Hospitals NHS Trust, Oxford, OX3 7LJ; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, OX3 7BN; 4) School of Medicine, Universidade Positivo, Curitiba, Brazil; 5) Department of Physiology, Anatomy and Genetics, MRC Functional Genomics Unit, University of Oxford, OX1 3QX; 6) Department of Neurology, Essex Centre for Neurological Sciences, Queen's Hospital, Romford; 7) Walton Centre NHS Foundation Trust, Liverpool, L9 7LJ; 8) Department of Clinical Genetics, St Michael's Hospital, Bristol, BS2 8EG; 9) Department of Paediatrics, Oxford University Hospitals NHS Trust, Oxford, OX3 7LJ; 10) School of Medicine, Dentistry and Biomedical Sciences, Queens University, Belfast; 11) Department of Neurology, Norfolk and Norwich University Hospital, Norwich; 12) School of Cancer Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT; 13) Department of Clinical Genetics, Southern General Hospital, Glasgow G51 4TF; 14) Fraser of Allander Neurosciences Unit, Royal Hospital for Sick Children, Glasgow G3 8SJ; 15) Thames Valley Dementia and Neurodegenerative Diseases Network, Oxford; 16) Oxford Regional Molecular Genetics Laboratories, Oxford University Hospitals NHS Trust; 17) McGill University and Genome Quebec Innovation Centre, McGill University, Montreal, Canada.

Ataxias are highly heterogeneous disorders caused by mutations in many genes. The diagnostic process is often long and many patients undergo multiple investigations without ever reaching a conclusive molecular diagnosis. The advent of massively parallel, Next Generation Sequencing (NGS) promises to revolutionise genetic testing and shorten the diagnostic process for some patients. We performed a pilot study using heterogeneous ataxias as a model neurogenetic disorder to assess the introduction of NGS into clinical practice. 58 known human ataxia genes were captured followed by NGS in 50 patients who had been extensively investigated and were refractory to diagnosis. All cases had been tested for Spinocerebellar ataxia 1-3,6,7, FRDA and had multiple other biochemical, genetic and invasive tests. Pathogenicity was assessed using a bioinformatics pipeline and novel variants were validated using functional experiments. The overall detection rate in our heterogeneous cohort was 18% and varied from 8.3% in those with an adult onset progressive disorder to 40% in those with a childhood or adolescent onset progressive disorder. The highest detection rate was in those with an adolescent onset and a family history (75%). The majority of cases with detectable mutations had a childhood onset but most are now adults, reflecting the long delay in diagnosis. The delays were caused by the lack of available clinical testing, presence of atypical phenotypes, use of indirect testing and presence of unknown genetic disorders. In the cases in whom we made an eventual diagnosis, the delay was 3-35 years (mean 18.1 years). Alignment and coverage metrics indicated that the capture and sequencing was highly efficient. Our pathogenicity interpretation pathway predicted 13 different mutations in 8 different genes: *PRKCG*, *TTBK2*, *SETX*, *SPTBN2*, *SACS*, *MRE11*, *KCNC3* and *DARS2* of which 9 were novel including one causing a newly described recessive ataxia syndrome, Spectrin-Associated autosomal recessive cerebellar ataxia type 1 (SPARCA1), which is allelic to SCA5. Targeted capture followed by NGS was efficient, cost-effective, and enabled a molecular diagnosis in many refractory cases. A specific challenge of NGS data is pathogenicity interpretation, but functional analysis confirmed the pathogenicity of novel variants showing that our pipeline was robust. Our results have broad implications for clinical neurology practice and the approach to diagnostic testing.

3054T

Shouting won't help: a combined strategy (Next Generation Sequencing and Linkage analysis) to identify Hereditary Hearing Loss (HHL) genes in affected families. *D. Vozzi, G. Girotto, F. Faletra, E. Rubinato, A. Morgan, D. Vuckovic, S. Lenarduzzi, M. Morgutti, P. Gasparini.* 1Med Genet, IRCCS-Burlo Garofolo Children Hospital, Via dell'Istria 65/1, Trieste, Italy.

Nonsyndromic HHL is a common disorder accounting for at least 60% of prelingual deafness. Despite the presence of some common HHL genes (GJB2, GJB6 and MTRNR1) still there is a need to 1) search for new causative mutations/genes and 2) develop new diagnostic approaches to overcome the problem of genetic heterogeneity. As regards to the first point a combined strategy based on linkage analysis (LA) followed by whole-exome sequencing (WES) was developed. Data analysis pipeline is based on data quality evaluation, hg19 reference reads mapping by BWA, variants calling and quality filtering by GATK. 'In silico' analysis of variants was done using a series of tools (PolyPhen-2, MutationTester, etc.). Using this approach we analyzed 6 Italian and 5 Qatari families affected by sensorineural, moderate to severe HHL. In a consanguineous family from Qatar, LA identified a region of 40 Mb on chromosome 5q13 (LOD score 3.8). WES data reported a causative mutation (c.7873 t>g leading to p.*2625Gluext*11) in BDP1 gene. The mutation disrupts the termination codon of the transcript resulting in an elongation of 11 residues of the protein. Immunohistochemistry in the mouse inner ear showed a clear expression of Bdp1 gene. In an Italian dominant family, LA showed a LOD of 3.3 on chromosome 12q24, where WES data analysis identified a new missense mutation (c.1057G>C; pG353R) in P2X2 gene, A 3D model of this protein suggests that the substitution of the hydrophobic Gly, with a charged residue as the Arg, it is expected to destabilize the protein folding. Finally, an Italian pedigree resembling a Y-linked form of HHL was analysed by WES identifying a missense mutation (c.A206T:p.D69V) in a gene (YYY). In silico prediction analysis reported this mutation as disease causing. YYY, a homolog of a HHL gene, is expressed in multiple cochlea cell types. Functional studies are now in progress to confirm the pathogenetic effect of the variation identified. As regards the second point, a Targeted Re-Sequencing (TRS) protocol has been developed based on Ion Torrent™ technology. It analyzes 96 HHL related genes ensuring 92% coverage of the target region (a total of 411.420 bp). The protocol has been tested and validated in many cases and is currently used in our diagnostic service. These findings definitely increase our knowledge of new HHL genes and may suggest new targets for hearing impairment treatment and prevention.

3055F

Genetic etiology study of the non-syndromic deafness in Chinese Hans by targeted next-generation sequencing. *T. Yang[1, 2], H. Wu[1, 2], Y. Chai[1, 2].* 1) Dept of Otolaryngology - Head And Neck Surgery, Shanghai Jiaotong University School of Medicine, Xinhua Hospital, Shanghai, China; 2) Ear Institute, Shanghai Jiaotong University School of Medicine, Shanghai, China.

Although over 60 non-syndromic deafness genes have been identified to date, the etiologic contribution of most deafness genes remained elusive. In this study, we addressed this issue by targeted next-generation sequencing of a large cohort of non-syndromic deaf probands. Probands with mutations in commonly screened deafness genes *GJB2*, *SLC26A4* and *MT-RNR1* were pre-excluded by Sanger sequencing. The remaining 125 deaf probands proceeded through targeted exon capturing of 79 known deafness genes and Illumina HiSeq2000 sequencing. Bi-allelic mutations in 15 less commonly screened deafness genes were identified in 28 deaf probands, with mutations in *MYO15A*, *GPR98*, *TMC1*, *USH2A* and *PCDH15* being relatively more frequent (≥3 probands each). Dominant mutations in *MYO6*, *TECTA*, *POU4F3* and *COCH* were identified in 4 deaf families. A mitochondrial *MTTS1* mutation was identified in one maternally inherited deaf family. No pathogenic mutations were identified in three dominant deaf families and two consanguineous families. We concluded that mutations in the less commonly screened deafness genes were heterogeneous and contributed to a significant percentage (17.4%) of causes for non-syndromic deafness. Targeted next-generation sequencing provided a comprehensive and efficient diagnosis for known deafness genes. Complementary to linkage analysis or whole-exome sequencing of deaf families, pre-exclusion of known deafness genes by this strategy may facilitate the discovery of novel deafness genes.

3056W

Exome sequencing identifies NFS1 deficiency in a novel Fe-S cluster disease, infantile mitochondrial complex II/III deficiency. *S.M.K. Farhan[1,2], J. Wang[1], J.F. Robinson[1], P. Lahiry[1], V.M. Siu[5,6], C. Prasad[5,6], J.B. Kronick[4], D.A. Ramsay[3], C.A. Rupar[5,6], R.A. Hegele[1].* 1) Department of Biochemistry, Western University, London, Ontario, Canada; 2) Robarts Research Institute, Western University, London, Ontario, Canada; 3) Department of Pathology, London Health Sciences Centre, Western University, London, Ontario; 4) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, Ontario; 5) Medical Genetics Program, Department of Pediatrics, London Health Sciences Centre, London, Ontario; 6) Children's Health Research Institute, London Ontario.

Iron-sulfur (Fe-S) clusters are a class of highly conserved and ubiquitous prosthetic groups with unique chemical properties that allow the proteins that contain them, Fe-S proteins, to assist in various key biochemical pathways. Mutations in Fe-S proteins often disrupt Fe-S cluster assembly leading to a spectrum of severe disorders such as Friedreich's ataxia or ISCU myopathy. Herein, we describe infantile mitochondrial complex II/III deficiency, a novel autosomal recessive mitochondrial disease characterized by lactic acidemia, hypotonia, respiratory chain complex II and III deficiency, multisystem organ failure and abnormal mitochondria. We applied autozygosity mapping to identify a homozygous region on chromosome 20 containing 453 genes. Next through whole-exome sequencing, we identified a novel variant within the autozygous region: c.215G>A, p.Arg72Gln in NFS1, a highly conserved cysteine desulfurase involved in the Fe-S cluster assembly machinery and is essential for the maturation of other Fe-S proteins such as ISCU, ISD11 and FXN. NFS1 p.Arg72Gln co-segregates with disease status in the family and was consistently predicted to be damaging by multiple in silico analyses. The exceptionally low minor allele frequency from the community of origin, from geographically and ethnically matched healthy control subjects coupled with reduced NFS1 expression in patient fibroblast cells as demonstrated by independent transfections, allowed us to converge on NFS1 p.Arg72Gln as the likely causative mutation in the affected kindred. We thus describe the first disease in man likely caused by deficiency in NFS1. Our results further demonstrate the importance of NFS1 expression in human physiology.

3057T

Reducing the cost of whole-exome sequencing of parent-affected offspring trios by joint Bayesian variant calling. *B. Hilbush[1], J. Blue-Smith[1], S. Lombardi[1], R. Littin[2], L. Trigg[2], A. Jackson[2], D. Ware[2], J.G. Cleary[2], F.M. De La Vega[1].* 1) Real Time Genomics, Inc., San Bruno, CA; 2) Real Time Genomics, Inc., Hamilton, NZ.

Whole-exome sequencing (WES) has been a successful and cost-effective strategy to identify disease-causing mutations of Mendelian highly penetrant diseases in research and clinical settings. The power to identify disease mutations is much higher by sequencing parent-affected offspring trios rather than the proband alone as genotypes of the parents remove irrelevant variants, aids in the removal of false positives shown as Mendelian inheritance errors (MIEs), and permits identifying recessive patterns. However, the cost of trio-based sequencing has limited its adoption in both clinical and research settings resulting in unsolvable cases. To circumvent this barrier, here we demonstrate a novel informatics approach to substantially reduce the cost of WES of trios leveraging the shared pedigree data allowing for full coverage sequencing of the proband and half coverage of the parents while maintaining high sensitivity and accuracy. Key to our results is the use of a Bayesian caller that simultaneously looks-up the alignment data of all family members when examining a site, scoring genotypes with priors based on Mendelian segregation patterns. To validate the approach, we obtained sequence data for a CEPH trio (Illumina 100bp paired-end, Agilent SureSelect v4+UTR) at a sequencing depth of ~60X coverage for NA12878 and at both 60X coverage and 30X coverage for the parents. We called variants on the full and half parental coverage trios and analyzed sensitivity and call set metrics for each of the individuals. We found that even with reduced parental coverage, we maintained 97.6% sensitivity to detect in NA12878 sites present in the 1000 genomes project OMNI array data across the targeted regions, and could call 98.6% of the offspring variant sites in both parents. Genotype concordance between full and half coverage for the child was 99.79% while both parents were above 99.4%. We also detected 78 de novo candidates and a MIE rate of 0.09%, a rate that is approximately 100X lower than that obtained with a singleton caller for NA12878. Our results show that our approach allows for a rational allocation of sequencing capacity in pedigree studies, resulting in cost savings with little deterioration of call quality. In the case of parent affected-offspring trios, this saving effectively result in sequencing three exomes for the cost of two, greatly empowering disease and clinical studies.

3058F

Systematic identification of causal mutations in Mendelian disorders using exome sequence data. *B. Thomas[1,2], M. Lek[1,2], N. Clarke[3], L. Waddell[3], M.A. Depristo[2], M.J. Daly[1,2], J. Laporte[4], J.J. Dowling[5], C.G. Bonnemann[6], K.N. North[3], D.G. MacArthur[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute of Harvard and MIT, Boston, MA, USA; 3) Institute for Neuroscience and Muscle Research, Sydney, NSW, Australia; 4) Department of Translational Medicine, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U964, CNRS UMR7104, Université de Strasbourg, Collège de France, 67404 Illkirch, France; 5) Department of Paediatrics, University of Michigan Medical Centre, Ann Arbor, MI 48109-2200, USA; 6) Neuromuscular and Neurogenetic Disorders of Childhood Section, NINDS, National Institutes of Health, Bethesda, MD, USA.

Exome sequencing has proven to be a powerful and cost-effective approach for the identification of causal mutations in many patients suffering from rare, severe Mendelian diseases. However, exome analysis unambiguously identifies a causal mutation in only 30-50% of sequenced families, indicating much work remains to be done to increase the yield of causal variants from sequencing-based approaches. Causal mutations can be missed by current exome sequencing approaches for a variety of reasons. Conversely, there may be multiple gene candidates that require further information to prioritize for functional studies. We describe the development of an integrated pipeline for the identification of causal variants from exome data and its application to a cohort of severe, undiagnosed muscle disease patients. Our online application called xBrowse enables the intuitive analysis of family-based exome data, permitting researchers and clinicians to rapidly explore the effects of altering inheritance modes and function/quality filters on the identification of potential causal mutations. Through xBrowse, our collaborators have early access to gene-based RNA expression data across various human tissues, a disease gene-centric protein interaction networks and a large reference panel of over 50,000 exomes. We have applied this integrated approach to exome data over 250 individuals consisting of families and probands affected by a range of neuromuscular diseases. We describe the detection of novel sequence variants with strong evidence for causality in these patients, and provide case studies indicating the value of tissue expression data, protein interaction networks, large reference panels for the prioritization of disease-associated mutations.

3059W

Mutations in PDGFRB and NOTCH3 as Causes of Autosomal Dominant Infantile Myofibromatosis. *L. Tian[1], J. Martignetti[2], L. Dong[1], M. Ramirez[2], O. Camacho[2], C. Camacho-Vanegas[2], Y. Guo[1], D. Zand[3], A. Bernstein[4], S. Masur[4], C. Kim[1], F. Otieno[1], C. Hou[1], N. Abdel-Magid[1], B. Tweddale[1], D. Metry[5], J. Fournet[6], E. Papp[7], E. McPherson[8], C. Zabel[8], G. Vaksmann[9], C. Morisot[9], B. Keating[1], P. Sleiman[1], J. Cleveland[10], D. Everman[11], D. Cho[12], Z. Li[12], E. Zackai[13], H. Hakonarson[1].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Mount Sinai School of Medicine, New York, NY 10029, USA; 3) Division of Genetics and Metabolism , Children's National Medical Center, 111 Michigan Avenue, NW Washington, DC 20010, USA; 4) Ophthalmology, Mount Sinai School of Medicine, New York, NY 10029, USA; 5) Department of Dermatology, Texas Children's Hospital, Baylor College of Medicine, Houston, TX 77030, USA; 6) HÔpital Necker-Enfants-Malades, 149 Rue de Sèvres, 75743 Paris Cedex 15, France; 7) Program in Genetics and Genome Biology, Hospital for Sick Children Research Institute, Toronto, Ontario, Canada; 8) Marshfield Clinic, 1000 North Oak Avenue , Marshfield, WI 54449, USA; 9) Department of Pediatric Cardiology, Cardiologic Hospital, 59037 Lille Cedex, France; 10) Carolinas HealthCare System, Matthews, NC 28105, USA; 11) Greenwood Genetic Center, 106 Gregor Mendel Circle, Greenwood, SC 29646, USA; 12) Department of Chemistry & Biochemistry, University of the Sciences in Philadelphia, Philadelphia, PA 19104; 13) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Infantile myofibromatosis (IM) is a disorder of mesenchymal proliferation characterized by the development of non-metastasizing tumors in the skin, muscle, bone, and viscera. Occurrence within families across multiple generations is suggestive of an autosomal dominant (AD) inheritance pattern, but recessive (AR) modes of inheritance have also been proposed. In a recent publication, we reported results from a whole-exome sequencing (WES) study in members of 9 unrelated families clinically diagnosed with AD IM to identify the genetic origin of the disorder. We have shown that In eight of the families we identified one of two disease-causing mutations, c.1978C>A (p.Pro660Thr) and c.1681C>T (p.Arg561Cys), in PDGFRB. one family did not have either of these PDGFRB mutations but all affecteds had a c.4556T>C (p.Leu1519Pro) mutation in NOTCH3. In this study, we carried out protein structural homology modeling of PDGFRB and functional studies of NOTCH3 to understand how the mutations in PDGFRB cause IM. We applied novel structural modeling tools to the construction of protein and protein-complex models designed to help elucidate biological mechanisms. We performed bioinformatics network analysis and results have shown the crosstalk between PDGFRB and NOTCH pathways, which may offer new opportunities to identify mutations in other genes which result in IM and is a necessary first step toward understanding the mechanisms of both tumor growth and regression and its targeted treatment.

3060T

Mild case of Unverricht-Lundborg disease presenting as juvenile myoclonic epilepsy. *E. Andermann[1,3,4,5,7], D. Amrom[1,3,4,7], F. Andermann[2,3,4,6,7], A.-L. Lehesjoki[8].* 1) Neurogenetics Unit; 2) Epilepsy Service and Seizure Clinic; 3) Montreal Neurological Hospital and Institute, Montreal, Quebec, Canada; 4) Department of Neurology & Neurosurgery; 5) Department of Human Genetics; 6) Department of Pediatrics; 7) McGill University, Montreal, Quebec, Canada; 8) Folhalsan Institute of Genetics and Neuroscience Centre, University of Helsinki.

Unverricht-Lundborg disease (ULD) (EPM1) is a progressive myoclonus epilepsy with variable severity and course, and variable degrees of cognitive deterioration. The onset is usually between 6 and 15 years of age with myoclonus and generalized tonic-clonic seizures. With good antiepileptic management, the patients can now survive into their 50's and 60's. The gene for this disorder was identified in 1996 as cystatin B (CSTB), a cysteine protease inhibitor. The most common mutation is a dodecamer repeat, although rare point mutations have also been described. A 30-year-old female patient had one generalized tonic-clonic seizure during sleep at the age of 11 years, and onset of myoclonic jerks on awakening at around the same time, which are well-controlled with valproic acid. She carries a clinical diagnosis of juvenile myoclonic epilepsy (JME). The patient was born prematurely at 25 weeks'gestation weighing 750 grams. Her developmental milestones were normal and she works as a highschool teacher for behaviourally challenged children. The parents were not known to be consanguineous, although they originated from two small towns 35 kms apart. Four siblings of the maternal grandmother were diagnosed clinically with ULD and were known to us; three sisters died in their 20's and 30's, and one brother died at age 65. A third degree cousin of the mother had epilepsy and died at 18 years of age. A paternal uncle had a single generalized tonic-clonic seizure at the age of 7 years. A distant cousin of the paternal great grandmother was also said to have progressive myoclonic epilepsy. The patient and her partner presented for preconceptional genetic counseling. Carrier screening for the CSTB gene was carried out for both the patient and her partner, employing PCR for the dodecamer repeat, and sequencing of the CSTB gene to rule out point mutations. CSTB testing in the patient surprisingly revealed that she was a compound heterozygote for two mutations: an expansion of the dodecamer repeat and a splice site c.67-1G>C mutation in intron 1, predicting a deletion of the downstream exon 2 with in-frame deletion of 34 aminoacids (p.delV23_K56). This is the second most common mutation underlying EPM1. The husband has no potential EPM1-causing sequence alterations. Although ULD is often confused with JME in the early stages of the disease, it is rare to find patients with ULD at age 30 who are as well controlled and high-functioning as this patient.

3061F

Whole exome sequencing combined with homozygosity mapping in a family with mental retardation, muscle weakness, and abnormal movement. *E. Jaberi[1], B. Farham[2], G.A. Shahidi[3], M. Rohani[3], I. Safari[1], B. Klotzle[4], E. Elahi[1,5].* 1) School of Biology, College of Science, University of Tehran, Tehran, Iran; 2) Iran University of Science and Technology, Tehran, Iran; 3) Dept. of Neurology, Tehran University of Medical Sciences, Tehran, Iran; 4) Illumina, San Diego, California, USA; 5) Dept. of Biotechnology, College of Science, University of Tehran, Tehran, Iran.

We performed whole genome homozygosity mapping in a consanguineous Iranian family with three affected and five unaffected individuals with mental retardation, muscle weakness, and abnormal movement using high density SNP chips. The pattern of inheritance was autosomal recessive. Their symptoms started from childhood with learning difficulty and evaluate as moderate mentally retarded. In addition to mental retardation and muscle weakness, they had severe kyphscoliosis and underwent spinal surgery. All patients had normal sensory and cerebellar examination. Brain MRIs revealed mild cerebral, cerebellar and brainstem atrophy without any significant structural abnormality. Electromyography and nerve conduction studies showed normal sensory and motor action potentials but neurogenic MUAPs in distal upper and lower extremity muscles. Disease status in the family linked to a homozygous region of 13 Mb on chromosome 6. We captured exomes of one affected individual from the family and performed sequencing analysis by a second-generation sequencer with a mean coverage of 30× and sufficient depth to call variants at ~97% of each targeted exome. The genetic variants in the homozygous region were filtered against the 1000 Genomes Project, HapMap, and the dbSNP131 database. After annotation and functional expectation, one gene with a novel splice site disruption mutation was found to be candidates for the disease segregated in the family. The candidate gene has not been reported for any disease before. The observed variation segregated with disease status in the pedigree, and the mutation had not seen in 400 control individuals. According to the sequences of cDNAs of the 3 affected individuals compared to controls, the mutation causes early stop codon which results in synthesis of a truncated protein. Further functional studies are being performed to validate the relation of the candidate gene with the disease phenotype in the family.

3062W

Next Generation Sequencing Defines New Gene(s) Involved in the Enlarged Vestibular Aqueducts and Pendred Syndrome. *F. Alasti, M. Hildebrand, T. Yang, RJH. Smith.* Molecular Otolaryngology and Renal Research Laboratories, Department of Otolaryngology, University of Iowa, Iowa City, 52252 IA. USA.

In patients with sensorineural hearing loss (SNHL), enlarged vestibular aqueduct (EVA), cochlear hypoplasia and Mondini (both EVA and cochlear hypoplasia) have been considered as the most common radiological finding. Pendred Syndrome (PS) is characterized by SNHL, temporal bone anomalies, and thyroid goiter. It is believed that non-syndromic DFNB4 SNHL, EVA, Mondini dysplasia and PS are parts of the same disease spectrum. Biallelic mutations in SLC26A4, is the main cause of the PS-Mondini-EVA-DFNB4 disease spectrum. Pendrin, the anion exchanger encoded by SLC26A4, likely helps maintaining the fluid level and ionic composition of endolymph, which is highly important during development of the shape of the bony structures of inner ear. However, mutations in cis- and trans-regulatory elements of SLC26A4 (transcriptional regulator FOXI1) as well as the potassium channel KCNJ10 have also been reported in carriers of SLC26A4 mutations. Our goal is to study more about the possible physiological mechanisms of ionic homeostasis of the endolymph. We believe that identification of new gene(s) involved in this disease can open new insights into the pathophysiological mechanisms of SLC26A4-related SNHL. We hypothesized that using targeted-sequence capture followed by massively parallel sequencing on our cohort of PS-Mondini-EVA-DFNB4 patients carrying 0 or 1 mutation in SLC26A4; we can likely identify additional gene(s) associated with this phenotype. To address this issue, we designed RNA baits and developed a platform (called EVASeq) by targeting the coding exons of the best candidate genes extracted from our previous murine microarray expression data. Our strategy was screening our samples for the EVASeq platform, as well as for the OtoSCOPE (a platform targetting all the known HL-causing genes) panel. Then, the pedigrees with multiple affected individuals, in which all the genes in both EVASeq and OtoSCOPE panels were excluded, were considered for Whole Exome Sequencing (WES). Sequencing the candidate gene(s) we found through WES in a large cohort of patients with the same clincal data (in order to find a second family), was a strong confirmation. Planning for the relevant functional assays for the gene(s) we found within our WES families is our current approach. The result of this study most likely will provide the first comprehensive insight into the identification of new genetic components involved in this PS-Mondini-EVA-DFNB4 disease spectrum.

3063T

Haploinsufficiency of *GJB5* identified via exome sequencing causes a novel form of cutis laxa. *M. Dasouki[1,2], J. Roberts[3], K. Gonzalez[4], W. Zeng[4], M. Butler[3], A. Belousov[5], I. Saadi[5].* 1) Dept. of Neurology, Univ Kansas Med Ctr, Kansas City, KS; 2) Dept. of Genetics, King Faisal Specialists Hospital & Research Center, Riyadh, Saudi Arabia; 3) Dept of Psychiatry, Univ Kansas Med Ctr, Kansas City, KS; 4) Ambry Genetics, Aliso Viejo, CA, KS; 5) Dept. of Anatomy & Cell Biology, Univ Kansas Med Ctr, Kansas City, KS.

Cutis laxa is a clinically and molecularly heterogeneous genodermatosis. Mutations in five connexin genes (*GJA1*, *GJB2*, *GJB3*, *GJB4*, and *GJB6*) have been associated with a range of heritable skin disorders. Human *GJB5*, *GJB4* and *GJA4* gene cluster maps within a 55 kb genomic region at chr.1p34.3. GJB5 is a novel gene, which is highly expressed in skin and placenta. *Gjb5* null mice have reduced viability. So far, two large microdeletions encompassing GJB5 had been reported. A 4.8 Mb deletion was found in a patient with hypothyroidism, low birth weight, short neck, microtia, micrognathia, strabismus and speech delay while a 2.1 Mb microdeletion was reported in an unrelated child with mental retardation. We ascertained a 36 year old Caucasian female who originally presented with chronic arthralgia, osteopenia, cutis laxa and heart palpitations. Routine clinical evaluations showed persistent large joint hypermobility, significant cutis laxa and left ventricular dysfunction. She had several unsuccessful surgical corrections of her lax skin. Magnetic resonance angiography of her chest was normal. She was treated with beta blockade and intravenous bisphosphonate. Exome sequencing of her gDNA revealed a deleterious frame shift mutation [c.37delG; p.V13fsx26] in *GJB5*. Two unrelated additional variants [c.3193G>A, p.V10645I; c.96818G>A, p.R32273H] were identified in *TTN* gene, which is known to be associated with cardiomyopathies. These *TTN* variants provide a plausible explanation for the ventricular dysfunction in our patient. However, this is the first report of an apparently deleterious mutation in *GJB5* causing cutis laxa. Functional studies using Hela cell scratch assays and patch clamp analyses are underway to characterize the *GJB5* mutation.

3064F

Identification of mutations causing congenital anomalies of the kidney and urinary tract through targeted sequencing. *N. Nicolaou[1], I.J. Nijman[1], A.M. van Eerde[1], G. Monroe[1], E.M. Bongers[2], E. Cuppen[1], K.Y. Renkema[1], N.V. Knoers[1].* 1) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, Netherlands; 2) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Congenital anomalies of the kidney and urinary tract (CAKUT) are developmental disorders that involve a spectrum of renal structural malformations. They occur in 1 out of 500 live-births and comprise the major cause of end-stage renal failure in childhood. There are 15 genes reported in the Human Gene Mutation Database with mutations responsible for CAKUT. These genes are very important for renal development as they play a role in kidney cell differentiation. However, a large proportion of cases remain unexplained. Our aim is to identify novel genetic defects involved in the etiology of CAKUT. We hypothesize that CAKUT arise from molecular defects during kidney development, and suggest that the unknown CAKUT candidate genes are also renal developmental genes. Therefore, rare mutations in any of these genes could underline the genetic heterogeneity of CAKUT. We employ a targeted next generation sequencing approach in 500 Dutch CAKUT patients. Enrichment methods allow us to capture and analyze the protein-coding regions of 208 candidate genes and 36 miRNAs, collectively referred to as the CAKUTome, in parallel. All genes selected were previously shown to have experimental evidence for their role in human isolated or syndromic CAKUT or were shown to be involved in disrupted nephrogenesis from previous studies in transgenic animals and cell models. For data analysis we used our in-house mapping and variant calling methodology. Sequencing results from the first 60 patients with severe kidney dysplasia showed that the average read depth is 210X. Approximately 200 variants were called per individual. Further variant prioritization, based on variant frequency data from dbSNP and Genome of the Netherlands, evolutionary conservation, and in-silico predictions, resulted in a list of potentially pathogenic mutations that were further validated by Sanger sequencing. Here we show that our targeted sequencing approach and variant prioritization method is efficient in identifying gene mutations in a large cohort of sporadic CAKUT cases. Previously reported causal mutations in known CAKUT genes were successfully identified in our cohort. Interestingly, the majority of the novel and promising pathogenic variants that we identified were unique for each patient. Hence, CAKUT might be even more heterogeneous in their etiology than expected. Functional studies to test the impact of novel mutations on protein function and kidney development are currently on-going.

3065W

***De novo* germline mutations in *MYL3* and *DYNC1H1* genes discovered by exome sequencing in an infant with congential joint contractures.** *D.W. Sant[1], R.L. Margraf[1], J. Durtschi[1], T.M. Newcomb[3], J.M. Opitz[2,4,5,6], J.C. Carey[5,6], H. Zhou[2], B.E. Katz[7], K.V. Voelkerding[1,2], K.J. Swoboda[3,6].* 1) ARUP Laboratories, Salt Lake City, UT; 2) Department of Pathology, University of Utah, Salt Lake City, UT; 3) Department of Neurology, University of Utah, Salt Lake City, UT; 4) Department of Human Genetics, University of Utah, Salt Lake City, UT; 5) Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT; 6) Department of Pediatrics (Division of Medical Genetics), University of Utah, Salt Lake City, UT; 7) Private Practice of Pediatrics and Pediatric Neurology (retired), Twin Falls, ID.

This report presents a family with one daughter affected by a previously unclassified disease, characterized by congenital joint contractures. Parents and four siblings are unaffected. To identify the potential causal variants, exome sequencing was performed on the proband, one sister, and both parents. A heuristic filtering approach identified several rare mutations, yielding the *de novo* mutations in the genes *MYL3* and *DYNC1H1* that were potentially correlated with the disorder and predicted inheritance. Sanger sequencing showed that the mutations were both present in the affected daughter, but not in either of the parents or siblings, showing that both mutations arose *de novo*. The *de novo* mutation in *MYL3* is a heterozygous, single base change from A to T on at position chr3:46899880. This variant causes a single-base amino acid change from tyrosine to asparagine (p.Y185N). This base is highly conserved between humans and mice to zebrafish and fruitfly. *MYL3* encodes the myosin light chain 3, an alkali light chain that is necessary for contraction of cardiac muscle and slow skeletal muscle. Mutations in *MYL3* can cause autosomal dominant disorders and have been associated with hypertrophic cardiomyopathy, Duchenne muscular dystrophy, and familial hypertrophic cardiomyopathy-8. The *de novo* mutation in *DYNC1H1* is a heterozygous, single base change from G to A at position chr14:102446717. This variant causes a single-base amino acid change from arginine to glutamine (p.R264Q). This base is highly conserved from humans to fungi. *DYNC1H1* codes for a cytoplasmic dynein protein that is important for cell movement and ATPase activity. Mutations in *DYNC1H1* have been associated with autosomal dominant lower extremity spinal muscular atrophy. In conclusion, we discovered two *de novo* heterozygous mutations in two genes with known heterogeneous and dominant phenotypes. Future work will determine which of these two mutations is likely causative or if both are contributing to cause this novel disorder.

3066T

ZNF259 IS A CANDIDATE GENE FOR ALOPECIA-PRIMORDIAL DWARFISM-RENAL SYNDROME. *A.C Smith[1], C.L Clericuzio[2], A. Ahmed[1], S.L Sawyer[1], D.E Bulman[1], K.M Boycott[1] FORGE Canada Consortium.* 1) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) University of New Mexico Health Science Center, Albuquerque, NM.

Alopecia-primordial dwarfism-renal syndrome (APDRS) is a distinctive pattern of malformation characterized by congenital alopecia, pre- and postnatal dwarfism and hypoplastic kidneys. This syndrome was observed in 5 Hispanic New Mexican children from four families. All had symmetric intrauterine growth restriction (~ 3 SD below the mean) and progressively severe postnatal growth restriction (~8 SD below the mean). Skin color is very pale with a porcelain-like appearance. The craniofacial appearance is distinctive with prominent forehead, deep-set and up-sloping eyes, small nose, downturned corners of the mouth and full cheeks. One child had cleft palate, two had shunted hydrocephalus and both males had genital hypoplasia. All had severe global delays and sensorineural hearing loss. Four children died at ages 10, 12, 20 and 31 months, from uremia and/or sepsis. The surviving patient is a 16 yo girl with renal failure, an unusual central fat distribution and type1 diabetes. Whole exome sequencing performed on the surviving patient and her unaffected parents yielded a homozygous missense mutation in the patient at a highly conserved locus of the ZNF259 gene, both her parents were carriers. Sanger sequencing of unaffected parents and siblings of a deceased patient demonstrated that these parents and one sibling were also heterozygous carriers of the mutation, although no patient DNA is available. ZNF259 encodes a zinc finger protein with diverse cellular localization and functions. It binds to the inactive form of the EGF receptor in quiescent cells and is then released upon activation. It has been demonstrated to bind to eukaryotic translation elongation factor 1A in proliferating cells, form a multiprotein complex with SMN protein and accumulate in subnuclear structures. Reduction of ZNF259 expression in mammalian cells by siRNA knockdown causes widespread phenotypes, including defects in transcription, prevention of DNA synthesis and accumulation of cells in the G1 and G2 phases of cell cycle. It has been hypothesized the profound inhibition of transcription and cell cycle arrest is due to a defect in pre-mRNA splicing, since cells lacking ZNF259 have been shown to have a deficiency in subcellular localization of snRNP's. Functional ZNF259 assays on patient fibroblast cells, as well as validating the mutation in other patients presenting with clinical features of APDRS are needed to confirm the role of ZNF259 in the pathogenesis of the disorder.

3067F

Novel MSX1 Mutations in Japanese Tooth Agenesis Patients. *S. Yamaguchi[1,2], J. Machida[2,3], M. Kamamoto[2,4], M. Kimura[2,5], A. Shibata[2,5], T. Tatematsu[2,5], H. Miyachi[2], Y. Higashi[5], P. Jezewski[6], A. Nakayama[7], K. Shimozato[2], Y. Tokita[2,5].* 1) Department of Dentistry and Oral Surgery, Aichi Children's Health and Medical Center, Obu, Aichi, Japan; 2) Department of Maxillofacial Surgery, School of Dentistry, Aichi-Gakuin University, Nagoya, Japan; 3) Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota, Japan; 4) Department of Oral and Maxillofacial Surgery, Japanese Red Cross Society Himeji Hospital, Himeji, Japan; 5) Department of Perinatology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan; 6) Department of Periodontology, University of Alabama at Birmingham School of Dentistry, Institute of Oral Health Research, Birmingham, USA; 7) Department of Embryology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan.

Since *MSX1* and *PAX9* are linked to the pathogenesis of non syndromic oligodontia and hypodontia, we performed a detailed mutation analysis of the *MSX1* and *PAX9* genes sampled from Japanese tooth agenesis cases. We identified two novel *MSX1* missense mutations with an amino acid substitution in homeodomain, Thr174Ile(T174I) from a sporadic hypodontia case and Leu205Arg (L205R) from a familial oligodontia case. Both Thr174 and Leu205 in MSX1 are highly conserved amino acids with in homeodomain among wide range of species. To define a possible role of these variants in the pathogenesis, we performed several functional analyses; Western blotting, immunocytochemistry, luciferase reporter assay and electrophoretic mobility shift assay (EMSA). Although the gene products are stable and capable of normal nuclear localization in transfected cells, these variants lose suppression activity on *myoD*-promoter in cells with differentiated condition. As the result of EMSA, we clarified that the DNA binding ability of both mutants was abolished by the mutations. Our findings argue in favor of a causative role of the T174I and L205R in the *MSX1* gene in tooth agenesis and suggest that these amino acid substitutions at the homeodomain of MSX1 may influence cell proliferation and differentiation resulting in lesser tooth germ formation *in vivo*.

3068W

Whole exome sequencing identifies variants causing different monogenic diseases in one nuclear family. *Y. Li[1], E. Lausch[2], K.O. Schwab[2], N. van der Werf-Grohmann[2], T. Velten[2], D. Lütjohann[3], P.V. Lorini[2], U. Matysiak-Scholze[2], B. Zabel[2], A. Köttgen[1,4].* 1) Department of Internal Medicine IV, University Hospital Freiburg, 79110 Freiburg, Germany; 2) Paediatric Genetics Division, Centre for Paediatrics and Adolescent Medicine, University Hospital Freiburg, 79106 Freiburg, Germany; 3) Institute of Clinical Chemistry and Clinical Pharmacology, University Clinics of Bonn, 53127 Bonn, Germany; 4) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Purpose: Whole exome sequencing has greatly facilitated the identification of mutations causing single-gene disorders, and has the potential to implicate genes previously not known to cause monogenic diseases. The purpose of this study was to search for causal mutations in two children from a consanguineous marriage. Both children are affected by hypomagnesemia and congenital hypothyroidism; one sibling also displayed hyperlipidemia. Methods: Whole exome sequencing was carried out using the Agilent SureSelect Exome Enrichment kit and the 5500 SOLiD system. Lifescope software was used for alignment and variant calling. Variants were tested for reproducibility using GATK. Quality metrics such as coverage, number of variants, percentage of known variants and the Ti/Tv ratio were evaluated. Results: Across individuals, mean coverage of the target region ranged from 40x to 63x, and target coverage at >20x ranged from 83% to 91%. After variant filtering (<1% population frequency; stop, splice, frameshift or missense variants; autosomal recessive mode of inheritance; and region of homozygosity obtained homozygosity mapping), 19 candidate variants were identified. One of these is the c.2667+1G>A mutation in TRPM6, which encodes a magnesium transporter expressed in intestine and kidney. This splice variant leads to skipping of exon 19 and is a known cause of autosomal recessive hypomagnesemia with secondary hypocalcemia (MIM #602014). In the child with hyperlipidemia, we identified a homozygous c.C1336T mutation in ABCG5, a sterol transporter. This known mutation introduces a premature stop codon and causes autosomal recessive sitosterolemia (MIM #210250). The other family members are heterozygous for this mutation. Biochemical analyses confirmed the diagnosis of sitosterolemia with elevated serum plant sterols. Mutations in both TRPM6 and ABCG5 were confirmed by Sanger sequencing. We are currently following up additional variants that fulfilled the filtering criteria as candidates for the thyroid phenotype. Both children are treated with thyroxine since shortly after birth and develop normally. Conclusion: Different mutations causing single-gene disorders cluster in offspring from consanguineous marriages, and can give rise to complex clinical pictures. Treatments for all three single-gene disorders are available; the lipid-lowering treatment can be adapted for the specific diagnosis.

3069T

Isolated X-linked Hypertrophic Cardiomyopathy with Restrictive Physiology Caused by a Novel Mutation of the Four-and-a-Half LIM Domain 1 Gene. *H. Hartmannova[1], M. Kubanek[2], M. Sramko[2], L. Piherova[1], L. Noskova[1], K. Hodanova[1], V. Stranecky[1], A. Pristoupilova[1], J. Sovova[1], T. Marek[2], J. Maluskova[3], P. Ridzon[4], J. Kautzner[2], H. Hulkova[1], S. Kmoch[1].* 1) Institute of Inherited Metabolic Disorders, Charles University, 1.st Medical Faculty, Prague 2, Czech Republic; 2) Department of Cardiology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic,; 3) Department of Pathology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic; 4) Department of Neurology, Thomayer's Hospital, Prague, Czech Republic.

Hypertrophic cardiomyopathy (HCM) with restrictive hemodynamic pattern has been associated with mutations of myofilament genes, severe exercise intolerance and poor prognosis. However, molecular etiology of this disease remains unexplained in a large proportion of cases. We performed exome sequencing in a Czech family with three males affected by non-obstructive HCM with restrictive phenotype. A novel frameshift mutation c.599_600insT of the four-and-a-half LIM Domain 1 (FHL1) gene was identified in these individuals. In accordance with expected X-linked recessive inheritance the mutation co-segregated in heterozygous females with abnormal electrocardiogram and in one case also with apical HCM. FHL1 is transcribed into three alternatively spliced mRNA isoforms FHL1A, FHL1B and FHL1C, encoding FHL1A, FHL1B and FHL1C proteins, respectively. The insertion does not affect translation of FHL1C but encodes for a frameshifts in translation of FHL1A and FHL1B, that are in both cases followed soon thereafter by a novel stop codon predicting proteosynthesis of one sequentially identical truncated form of the FHL1A containing a neo-peptide composed of 32 aminoacids (200 HRCGGPVLLRGLLQELCGQEVC WMQEPHHWVW-232) on the C-terminal end, (p.F200fs32X). The mutation does not affect transcription, splicing and stability of FHL1A mRNA and results into production of a truncated FHL1 protein, which is contrary to heart tissue homogenates not detectable in frozen tissue sections of myocardial biopsy of affected males. This suggests that p.F200fs32X mutation results in proteosynthesis of truncated neo-protein, which has, either due to loss of LIM 3 and LIM4 domains and/or presence of the neo-peptide on its C-terminal end, altered structural properties limiting function and immunodetection of the mutant protein in the native state. Although skeletal muscle involvement is a common finding in FHL1-related diseases, we could exclude myopathy in all mutation carriers. Our findings thus support the view, that mutations leading to reduced amounts of FHL1 are less deleterious (or even benign, as in our case) for skeletal muscle than missense mutations exerting their pathogenic effects through misfolding, self-aggregation and co-aggregation of FHL1-binding partners.

3070F

Mutations in PIK3R1 cause SHORT syndrome. *D. Dyment[1], A. Smith[2], D. Alcantara[3], J.A. Schwartzentruber[4], L. Basel-Vanagaite[5], C.J. Curry[6], I.K. Temple[7], W. Reardon[8], S. Mansour[9], M.R. Haq[10], R. Gilbert[10], O.J. Lehmann[11], M.R. Vanstone[2], C.L. Beaulieu[2], The. FORGE CANDADA CONSORTIUM[12], J. Majewski[4], D.E. Bulman[2], M. O'Driscoll[3], K.M. Boycott[1,2], A.M. Innes[13,14].* 1) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 3) Genome Damage and Stability Centre, University of Sussex, Brighton, UK; 4) 4McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada; 5) 5Department of Pediatric Genetics, Schneider Children's Medical Center of Israel, Petah-Tikva, Israel; 6) Genetic Medicine Central California, Fresno, CA, USA and the Department of Pediatrics, University of California San Francisco, CA, USA; 7) 7Faculty of Medicine, University of Southampton and the Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, UK; 8) Our lady's Hospital for Sick Children, Crumlin, Dublin 12, Ireland; 9) SW Thames Regional Genetics Service, St. George's Hospital Medical School, London, UK; 10) Department of Paediatric Nephrology, Southampton Children's Hospital, University Hospital Southampton University Hospitals NHS Foundation Trust, Southampton, UK; 11) Department of Ophthalmology, University of Alberta, Edmonton, AB, Canada; 12) Steering Committee Membership is listed at end of poster; 13) Department of Medical Genetics, University of Calgary, Calgary, AB, Canada; 14) Alberta Children's Hospital Research Institute for Child and Maternal Health, University of Calgary, Calgary, AB, Canada.

SHORT syndrome is a rare, multisystem disease characterized by short stature, anterior chamber eye anomalies, characteristic facial features, lipodystrophy, hernias, hyperextensibility and delayed dentition. As part of the FORGE Canada Consortium we studied individuals with clinical features of SHORT syndrome to identify the genetic etiology of this rare disease. Whole exome sequencing in an affected child-unaffected parents trio identified a de novo frameshift insertion in exon 14 of PIK3R1, c.1906_1907insC, p.Asn636Thrfs*18. Heterozygous mutations in exon 14 of PIK3R1were subsequently identified by Sanger sequencing in three additional affected individuals and two affected family members. One of these mutations, c.1945C>T, p.Arg649Trp, was confirmed to be a de novo mutation in one affected individual and was also identified and shown to segregate with the phenotype in an unrelated family. The other mutation was a de novo truncating mutation, c.1971T>G, p.Tyr657*, identified in another affected individual. PIK3R1 is involved in the PI-3K signaling cascade and, as such, plays an important role in cell growth, proliferation and cell survival. Functional studies on the fibroblasts cells with the PIK3R1, c.1906_1907insC mutation showed decreased phosphorylation of downstream S6 target of the PI-3K-AKT-mTOR pathway. Our findings show that PIK3R1 mutations are the major cause of SHORT syndrome and suggest the molecular mechanism of disease may involve down-regulation the PI-3K-AKT-mTOR pathway.

3071W

Whole Exome Sequencing of Permanent Neonatal Diabetes Patients. *H.M. Highland[1], H. Ye[2], J.E. Below[1], V.P. Paz[2], D. Muzny[3], R.A. Gibbs[3], E. Boerwinkle[1,3], C.L. Hanis[1], G.I. Bell[2], L.H. Philipson[2], S.A.W. Greeley[2].* 1) Human Genetics Center, The University of Texas Health Science Center at Houston, P.O. Box 20186, Houston, TX 77225; 2) Section of Adult and Pediatric Endocrinology, Diabetes, & Metabolism, The University of Chicago, 5841 South Maryland Avenue, MC 1027, Chicago, IL 60637; 3) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030.

Introduction: Permanent Neonatal Diabetes Mellitus (PNDM) is characterized by hyperglycemia onset within the first six months of life that does not resolve. While most PNDM patients were historically treated with insulin therapy, those with mutations in *KCNJ11* or *ABCC8* usually have an excellent response to oral sulfonylurea treatment, highlighting the critical importance of a genetic diagnosis to guide personalized treatment. PNDM is a monogenic disease with marked genetic heterogeneity in a case series. It is most commonly caused by mutations in *KCNJ11*, *INS*, *ABCC8*, *EIF2AK3*, *FOXP3*, *GATA6*, *GCK*, *PDX1*, *RFX6*, *PTF1A*, *NEUROD1*, *NEUROG3*, *GLIS3*, and other very rare causes. Mutations in many of these genes result in numerous syndromic features in addition to PNDM. **Methods:** We undertook deep whole exome sequencing of 18 singleton patients with PNDM, four parent-offspring trios with unaffected parents, and one multiplex family. With the exception of one positive control, these patients had been screened for mutations in PNDM genes consistent with their clinical presentation. Three patients also have trisomy 21. We have investigated three hypotheses of the relationship between PNDM and trisomy 21. 1) PNDM and trisomy 21 are unrelated conditions. 2) A disomic homozygous variant on chromosome 21 is sufficient to cause recessive disease. 3) Increased dosage of a gene on chromosome 21, such as a micro RNA, interacts with an uncommon variant elsewhere in the genome to cause permanent neonatal diabetes. **Results:** Of the 22 independent cases, we have identified mutations likely to be the basis of PNDM in 11 of the patients. These mutations have been found in known PNDM genes (*FOXP3*, *INS*, *KCNJ11*, *EIF2AK3*, and *IER3IP1*). The patients with mutations in *FOXP3* presented without all the characteristics typical of mutations in that gene. For the trisomy 21 patients, viable candidates have emerged under each model that require further biological validation. **Conclusions:** We've found broader phenotypic variation in patients with mutations in known PNDM genes than has previously been established. Our ability to find variants under each proposed relationship between trisomy 21 and PNDM points out both the promises and challenges of next generations sequencing for Mendelian diseases.

3072T

Pseudoxanthoma elasticum in a United States Veteran. *S. Konda, W. Chen, J. Toro.* Dermatology Department, Veterans Affairs Medical Center, Washington, DC.

Pseudoxanthoma elasticum (PXE) is an autosomal recessive disorder with ectopic mineralization of the elastic fibers of the reticular dermis of the skin, Bruch's membrane of the eye, and the cardiovascular system. The prevalence is estimated to be around one in 50,000 with a carrier frequency of ~1:150-300. Inactivating mutations in the ABCC6 gene, which encodes MRP6, an ABC-cassette efflux transporter protein in hepatocytes and renal cells that has been implicated in the pathogenesis of PXE. Over 300 distinct mutations in the ABCC6 gene have been described. A 67-year-old Caucasian man with a history of Hodgkin's disease, hypercholesterolemia, basal cell carcinomas, and actinic keratoses presented with a lifelong history of yellow cobblestone lesions and lax skin localized to his neck, axillary vaults, and antecubital and popliteal fossae. He had a history of angioid streaks with retinal hemorrhages and associated decreased visual acuity for which he is followed at the NIH. He denied any history of myocardial infarctions, mitral valve prolapse, gastrointestinal bleeding, elastosis serpiginosa perforans, oral penicillamine therapy, or family history of similar lesions. Dermatologic exam showed numerous yellow papules and reticulated yellow plaques with a cobblestone appearance and redundant skin localized to his neck and axillary vaults with subtle involvement of his antecubital and popliteal fossae. A skin biopsy obtained from a plaque localized to his left axillary vault demonstrated clumped fragmented elastosis in the reticular dermis consistent with PXE. The fragmented elastic fibers stained strongly with Verhoeff-Van Gieson and Von Kossa stains. His skin lesions are asymptomatic. Genetic genetic testing for mutations in ABCC6 showed the patient is heterozygous for two mutations: IVS21±1 G>T and R1141X. He is currently undergoing treatment for his angioid streaks with bevacizumab, a humanized monoclonal antibody that inhibits vascular endothelial growth factor A (VGEF-A). Some studies advocate patients may benefit from limiting dietary calcium and phosphorous. He is enrolled in a clinical trial to evaluate if dietary magnesium supplementation can modulate the course of his disease.

3073F

A *de novo* mutation in *TLE3*, encoding the transducin-like enhancer protein 3, is associated with a rare subtype of frontonasal dysplasia.
Y. Xi[1], C.L. Beaulieu[1], A.C. Smith[1], J. Schwartzentruber[2], F.P. Favaro[3], D.B. Bulman[1], M.A. Guion-Almeida[3], A. Richieri-Costa[3], A.M. Innes[4], K.M. Boycott[1] FORGE Canada Consortium. 1) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada; 3) Department of Clinical Genetics, Hospital of Rehabilitation of Craniofacial Anomalies, University of Sao Paulo (HRAC-USP), Bauru, SP, Brazil; 4) Department of Medical Genetics, University of Calgary, Calgary, AB, Canada.

Frontonasal dysplasia is the hallmark of several syndromes involving the frontonasal process. One of the syndromes within this spectrum is characterized by midline facial clefting, agenesis of the corpus callosum, basal encephalocele, pituitary and ocular abnormalities. These features are very close to Sakoda syndrome (OMIM # 610871). All cases reported to date have been sporadic. Here employing whole-exome sequencing on a patient and his parents, we identified a *de novo* mutation in the *TLE3* gene (NM_ 005078:exon15:c.A1517G:p.K506R), encoding the transducin-like enhancer protein 3. Analysis of a further three patients with a similar clinical presentation did not identify any mutations in this gene and suggests this may be a genetically heterogeneous developmental syndrome. TLE3 belongs to the Gro/TLE protein family, members in which are broadly expressed during development and have essential functions in many developmental pathways including Notch and Wnt signalling (Jennings *et al* Genome Biol 2008). TLE family proteins perform their transcriptional repression functions by interacting with a diverse profile of transcription factors (including Hes, Runx and Pax) rather than binding DNA directly. The C-terminal WD repeat domain in TLE proteins is highly conserved and responsible for their interactions with transcription factors (Carvalho *et al* Mol Endocrinol 2010). In our study, the mutation in TLE3 (p.K506R) lies in the first WD repeat and this mutation could change its binding efficiency to various transcription factors during frontonasal development. Recently, TLE proteins were also shown to biochemically interact with EphrinB1, mutations in which cause craniofrontonasal dysplasia; the WD domain plays important role in regulation of this interaction (Kamata *et al* BMB Reports 2011). Thus *TLE3* is a novel candidate gene for this subtype of frontonasal dysplasia. Future experiments such as immunoprecipitation and Western blotting will be performed to dissect the binding capacities of wt and mutant TLE3 with different transcription factors and EphrinB1. Morpholino knockdown on zebrafish *tle3a* and *tle3b* (*TLE3* homologs in zebrafish) will also be performed to characterize their functions during craniofrontonasal development.

3074W

Minimal evidence for oligogenic inheritance in Joubert syndrome based on the first 20 genes.
R. Bachmann-Gagescu[1], I. Phelps[2], B. O'Roak[3], C. Isabella[2], D. O'Day[2], M. Kircher[3], J. Dempsey[2], I. Glass[2], D. Witten[4], G. Cooper[5], J. Shendure[3], D. Doherty[2]. 1) Institue for Molecular Life Sciences, University of Zurich, Zurich, Switzerland; 2) 2Dept of Pediatrics, University of Washington, Seattle, USA; 3) Dept of Genome Sciences, University of Washington, Seattle, USA; 4) 4Dept of Biostatistics, University of Washington, Seattle, USA; 5) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama.

Joubert syndrome (JS) is a ciliopathy characterized by a distinctive hindbrain malformation, ataxia and cognitive dysfunction. It is typically inherited in an autosomal recessive manner with biallelic mutations in one of 20 genes shown to be associated with this disorder. Despite the rapid progress in gene identification and the large number of genes already identified, the underlying cause remains undetermined in almost 50% of subjects with JS, because they lack biallelic mutations in any of the 20 known JS genes. However, heterozygous variants in two or more JS-associated genes have been reported in multiple patients, raising the possibility that JS could be caused by oligogenic inheritance in some patients. If oligogenic inheritance were a common mechanism for JS, we hypothesize that patients without biallelic mutations in a single causal gene will carry heterozygous rare, deleterious variants (RDVs) in multiple genes. To test this hypothesis, we used a novel, highly efficient, molecular inversion probe-based capture technology to sequence the 20 known JS-associated genes in 362 patients with JS. Using a recessive model where biallelic RDVs in any of the 20 known JS-associated genes were considered causal, we were able to determine the cause in less than two-thirds of our subjects. The total number of RDVs across all JS-associated genes was lower in subjects without an identified cause compared to those with an identified cause. Moreover, the proportion of subjects that carried ≥2 heterozygous RDVs was not significantly different between the groups with and without biallelic RDVs in a causal gene. Our data provide no support for the hypothesis that a significant proportion of JS cases are due to oligogenicity (i.e. heterozygous RDVs in multiple genes). The etiology underlying JS in the group still lacking an identified cause will most likely be due to a second unidentified RDV in a known JS-gene with an already identified heterozygous RDV, or to variants in genes yet to be implicated in JS. True oligogenic inheritance may still occur in JS; however, it would have to be quite rare or involve unidentified RDVs in the currently known JS-genes, RDVs in as yet undiscovered JS-genes, or common variants in either group of genes. Future work will focus on identifying the remaining genetic causes of JS and determining the role of additional heterozygous RDVs as modifier variants in expression of the phenotype.

3075T

Exome Sequencing Identifies Germline Mutations in *SPAG1* as a Cause of Primary Ciliary Dyskinesia Associated with Defective Outer and Inner Dynein Arms. *M.A. Zariwala[1], L.E. Ostrowski[2], N.T. Loges[3], T. Hurd[4,5], M.W. Leigh[6], L. Huang[2], W.E. Wolf[2], J.L. Carson[6], M.J. Hazucha[2], W. Yin[2], S.D. Davis[6,7], S.D. Dell[8], T.W. Ferkol[9], S.D. Sagel[10], K.N. Olivier[11], C. Jahnke[3], H. Olbrich[3], C. Werner[3], H.Y. Gee[4,12], E.A. Otto[4], J. Halbritter[4,12], E.H. Turner[13], A.P. Lewis[13], M.J. Bamshad[13,14], D.A. Nickerson[13], F. Hildebrandt[4,12,15], J. Shendure[13], H. Omran[3], M.R. Knowles[2], Genetic Disorders of Mucociliary Clearance Consortium (GDMCC).* 1) Department of Pathology & Laboratory Medicine, UNC School of Medicine, Chapel Hill, NC 27599, USA; 2) Department of Medicine, UNC School of Medicine, Chapel Hill, NC 27599, USA; 3) Department of General Pediatrics and Adolescent Medicine, University Hospital Muenster, 48149 Muenster, Germany; 4) Department of Pediatrics, University of Michigan, Ann Arbor, Michigan 48109, USA; 5) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburg, Edinburgh EH4 2XU, UK; 6) Department of Pediatrics, UNC School of Medicine, Chapel Hill, NC 27599, USA; 7) Section of Pediatric Pulmonology, Allergy and Sleep Medicine, James Whitcomb Ruket Riley Hospital for Children, Indiana University School of Medicine, Indianapolis, IN 46202, USA; 8) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON, M5G1X8, Canada; 9) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA; 10) Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO 80045, USA; 11) Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA; 12) Division of Nephrology, Harvard Medical School, Boston Children's Hospital, Boston, MA 02114, USA; 13) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 14) Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98195, USA; 15) Howard Hughes Medical Institute, Chevy Chase MD 20815, USA.

Primary ciliary dyskinesia (PCD) is a genetically heterogeneous, autosomal recessive disorder, characterized by oto-sino-pulmonary disease and situs abnormalities. PCD-causing mutations have been identified in 19 genes, but collectively they account for only ~65% of all PCD. To identify mutations in additional genes that cause PCD, we performed exome sequencing on 3 unrelated probands with ciliary outer and inner dynein arm (ODA+IDA) defects. Mutations in *SPAG1* were identified in one family with three affected siblings. Further screening of *SPAG1* in 98 unrelated affected individuals (62 with ODA+IDA defects; 35 with ODA defects; one without available ciliary ultrastructure) revealed biallelic loss-of-function mutations in 11 additional individuals (including one sib-pair). All 14 affected individuals with *SPAG1* mutations had a characteristic PCD phenotype, including 9 with situs abnormalities. Additionally, all individuals with mutations who had defined ciliary ultrastructure had ODA+IDA defects. SPAG1 was present in human airway epithelial cell lysates, but was not present with isolated axonemes, indicating that it likely plays a role in the cytoplasmic assembly/trafficking of the axonemal dynein arms. Together, these results demonstrate that mutations in *SPAG1* cause PCD with ciliary ODA+IDA defects, and that exome sequencing is useful to identify genetic causes of heterogeneous recessive disorders. This abstract was funded by 5U54HL096458-06 (NIH/ORDR/NHLBI), 5R01HL071798 (NIH/NHLBI), 5R01HL094976 (NIH/NHLBI), RC2 HL-102923 (NIH/NHLBI), 5R21HG004749 (NIH/NHGRI), UL1 TR000083 (NIH/NCATS), UL1 TR000154 (NIH/NCATS), DFG Om6/4, IZKF Om2/009/12, BESTCILIA (EU), SYSCILIA (EU) and Howard Hughes Medical Institute.

3076F

A Genetic Snapshot of Risk Alleles in the Jewish Population. *D. Zielinski, Y. Erlich.* Whitehead Institute for Biomedical Research, Cambridge, MA.

Carrier screening has been instrumental in preventing rare genetic disorders. As a result of various bottlenecks and isolation, risk alleles for certain disorders are found at a higher frequency in the Jewish population, which has particularly benefited from prenatal screening. Despite great success, including the prevention of fatal disorders in some communities, carrier screening does not cover the entire allele spectrum, resulting in false negatives and thus residual risk. Current methods to ascertain new risk alleles rely on sequencing affected children. We are taking a proactive approach to uncover risk alleles through complete sequencing of the entire coding regions of more than 150 genes documented by the Israeli Ministry of Health that are associated with genetic disorders in Jews of various ethnic backgrounds. The aims of this study are to uncover novel alleles in known disease-associated genes and generate more data on disease burden, including carrier frequencies of known founder mutations as well as minor mutations.

Studying rare variations is a needle in a haystack problem, as large cohorts have to be sequenced in order to trap the variations and gain statistical power. However, sample preparation techniques have not kept pace with the exponential growth in high throughput sequencing technology and rely on tedious, individual barcoding. We have developed a scalable solution, called DNA Sudoku, to reduce sample preparation time and costs and increase throughput. After several successful pilot studies, we have employed the pipeline to take a genetic snapshot of risk alleles for genetic disorders in the Jewish population, focusing our efforts on a relatively small cohort of healthy, ethnically matched individuals, in order to create a comprehensive risk catalog and further improve carrier screening.

3077W

Exome sequencing identifies *de novo* and post-zygotic mutations in *GATA6* associated with congenital diaphragmatic hernia. *J. Bennett[1], L. Yu[2], J. Wynn[2], Y.H. Cheung[2], Y. Shen[3], G.B. Mychaliska[4], T.M. Crombleholme[5], K.S. Azarow[6], F.Y. Lim[7], D.H. Chung[8], D. Potoka[9], B.W. Warner[10], B. Bucher[10], C. Stolar[11], G. Aspelund[11], M.S. Arkovitz[12], H. Mefford[1], W.K. Chung[2], The University of Washington Center for Mendelian Genomics.* 1) Division of Genetic Medicine, University of Washington, Seattle, WA; 2) Division of Molecular Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY; 3) Department of Biomedical Informatics, Columbia University Medical Center, New York, NY; 4) Department of Surgery, University of Michigan Health System, Ann Arbor, MI; 5) Colorado Fetal Care Center, Division of Pediatric General, Thoracic, and Fetal Surgery, Children's Hospital Colorado and the University of Colorado School of Medicine, Aurora, CO; 6) Department of Pediatric Surgery, University of Nebraska College of Medicine, Omaha, NE; 7) Division of Pediatric General, Thoracic, and Fetal Surgery, Center for Molecular Fetal Therapy, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 8) Department of Pediatric Surgery, Vanderbilt University Medical Center, Vanderbilt Children's Hospital, Nashville, TN; 9) Department of Pediatric Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA; 10) Division of Pediatric Surgery, Washington University School of Medicine, St. Louis, MO; 11) Division of Pediatric Surgery, Department of Surgery, Columbia University Medical Center, New York, NY; 12) Division of Pediatric Surgery, Tel Hashomer Medical Center, Tel Hashomer, Israel.

Congenital diaphragmatic hernia (CDH) is a relatively common and severe birth defect that affects 1 in 3,000 births. It is characterized by herniation of abdominal viscera into the chest cavity through the incomplete formation of diaphragm. Lung hypoplasia is frequently associated with CDH and contributes significantly to its morbidity and mortality. Although chromosomal anomalies and mutations in several genes have been implicated in the etiology of CDH, the cause for most patients is unknown. We used whole exome sequencing (WES) in two families with CDH and congenital heart disease, and identified mutations in *GATA6* in both. In the first family, we identified a *de novo* missense mutation (c.1366C>T; p.R456C) in a sporadic CDH patient with tetralogy of Fallot. In the second, a nonsense mutation (c.712G>T; p.G238X) was identified in two siblings with CDH and a large ventriculoseptal defect. The G238X mutation was inherited from their mother, who was clinically affected with congenital absence of the pericardium, patent ductus arteriosus, and gastrointestinal malrotation, but without clinical manifestations of CDH. Deep sequencing of blood and saliva derived DNA suggested somatic mosaicism as an explanation for her milder phenotype, with only approximately 15% mutant alleles. To determine the frequency of *GATA6* mutations in CDH, we used targeted capture and sequencing to screen a cohort of 357 patients with CDH, with and without congenital heart disease. We identified one additional *de novo* mutation (c.1071delG, p.V358Cfs34X) in a patient with CDH and atrioseptal defect. Mutations in *GATA6* have been previously associated with pancreatic agenesis and congenital heart disease, and mutations in its paralogue, *GATA4*, have recently been implicated in CDH as well. We conclude that, in addition to the heart and the pancreas, *GATA6* is involved in development of two additional organs, the diaphragm and the pericardium, which share a common embryonic origin. In addition we have shown that both *de novo* and post-zygotic mutations can contribute to the development of common birth defects. Our data suggests that WES provides an unbiased powerful tool to identify genes causing CDH.

3078T

Whole-Exome Sequencing identified a homozygous *FNBP4* mutation in a family with a condition similar to Microphthalmia with Limb Anomalies. *Y. Kondo[1], E. Koshimizu[1], A. Megarbane[2], H. Hamanoue[3], I. Okada[1], K. Nishiyama[1], H. Kodera[1], S. Miyatake[1], Y. Tsurusaki[1], M. Nakashima[1], H. Doi[1], N. Miyake[1], H. Saitsu[1], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Medical Genetics Unit, St. Joseph University, Beirut, Lebanon; 3) Department of Obstetrics and Gynecology, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Microphthalmia with limb anomalies (MLA), also known as Waardenburg anophthalmia syndrome or ophthalmoacromelic syndrome, is a rare autosomal recessive disorder. Recently, we and others successfully identified *SMOC1* as the causative gene for MLA. However, there are several MLA families without *SMOC1* abnormality, suggesting locus heterogeneity in MLA. We aimed to identify a pathogenic mutation in one Lebanese family having an MLA-like condition without *SMOC1* mutation by whole-exome sequencing (WES) combined with homozygosity mapping. A c.683C>T (p.Thr228Met) in *FNBP4* was found as a primary candidate, drawing the attention that FNBP4 and SMOC1 may potentially modulate BMP signaling.

3079F

Depletion of exome sequencing reads reveals a 14kb homozygous deletion in a patient with PHARC syndrome. *T. Harel[1], C. Gonzaga-Jauregui[1], T.O. Crawford[2], M. Koenig[3], R.A. Gibbs[1,4], J.R. Lupski[1,5,6], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 2) Departments of Neurology and Pediatrics, Johns Hopkins University, Baltimore, Maryland; 3) Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC), CNRS-INSERM-Universite de Strasbourg, 67404 Illkirch, France; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030; 6) Texas Children's Hospital, Houston, TX, 77030.

Whole-exome sequencing (WES) has proved extremely useful in identifying simple nucleotide variation (SNV), i.e., single base pair mutations and small indels, responsible for highly penetrant Mendelian phenotypes. However, assessment of copy number variation (CNV), small deletions and duplications, responsible or contributing to Mendelian phenotypes from WES data is still a task in progress as bioinformatic algorithms are being developed and improved. We performed exome sequencing in a patient from a Qatari consanguineous family that presented with polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract (PHARC). Analysis of the WES data reads and mapping in the vicinity of the ABHD12 gene, encoding abhydrolase domain-containing protein 12 and associated with PHARC syndrome, showed complete absence of sequencing reads in a 14kb region on chromosome 20, encompassing exon 1 of ABHD12. Additionally, there was absence of heterozygosity (AOH) surrounding the proposed deletion region. Subsequent confirmation by PCR of the deletion breakpoint confirmed autosomal recessive segregation of the deletion with the phenotype in the family. Moreover, Sanger sequencing of the breakpoint revealed this deletion to be identical to one reported previously by Fiskerstrand et al. (2010) in a family from the United Arab Emirates. This finding suggests that this deletion might have occurred as a single event in a common ancestor's haplotype and that it might currently be segregating in the population of this region. The deletion breakpoint maps within a region of microhomology and high identity between a pair of flanking Alu elements. Interestingly, WES also identified a homozygous rare and predicted deleterious GDAP1 variant (p.A336S) in the proband. GDAP1 is a known and well characterized neuropathy-associated gene. Segregation of this variant was confirmed in two of the three affected siblings of the Qatari family but was not present in the third, nor in the Emirati patient. The contribution of this mutation to the neuropathy phenotype in these patients need be considered. In summary, we show the importance of considering CNV and SNV contribution and interplay in Mendelian phenotypes. Meticulous analysis of WES data can reveal copy number variation of direct clinical relevance, providing additional utility of exome sequencing in identifying the genetic basis of rare genetic disorders.

3080W

From monogenic to oligogenic: strategies to uncover oligogenic modes of inheritance in individuals with intellectual disability. *W.D. Jones[1], M. Van Kogelenberg[1], D. King[1], T. Fitzgerald[1], H.V. Firth[1,2], M.E. Hurles[1], J.C. Barrett[1], Deciphering Developmental Disorders Study.* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Cambridge University Department of Medical Genetics, Addenbrooke's Hospital, Cambridge CB2 2QQ.

Phenotypes that segregate in a Mendelian manner have long been thought to result from alterations in a single gene or unique loci. However there is recent evidence to suggest that oligogenic modes of inheritance may underlie intellectual disability in some individuals. The Deciphering Developmental Disorders (DDD) Project is a study of individuals with developmental disorders with recruitment from 24 Regional Genetics Services across the UK and Republic of Ireland. Individuals first undergo detailed clinical phenotyping at their local Genetics centre. This is followed by exome-array Comparative Genomic Hybridisation (CGH) of the proband and exome sequencing of the proband and both parents. The combining of data from array CGH analysis and trio exome sequencing affords a powerful platform to look for evidence of oligogenic modes of inheritance in individuals with intellectual disability. A control data set known as the 'imaginary siblings' will be generated to carry out this analysis, consisting of variants present in the parents not inherited by the proband. The computational and statistical methods employed in the analysis for oligogenic models will be presented alongside preliminary results.

3081T

Integrative genomics reveals that distal cis-regulatory mutations cause isolated pancreatic agenesis. *M.N. Weedon[1], I. Cebola[2,3], A. Patch[1], S.E. Flanagan[1], E. De Franco[1], R. Caswell[1], S.A. Rodríguez-Seguí[2,3,4], C. Shaw-Smith[1], C. Cho[5], H. Lango Allen[1], J.A.L Houghton[1], C.L. Roth[6], R. Chen[7], K. Hussain[8,9], P. Marsh[10], L. Vallier[5], A. Murray[1], S. Ellard[1], J. Ferrer[2,3,11], A.T. Hattersley[1].* 1) University of Exeter Medical School, UK; 2) Institut d'Investigacions Biomèdiques August Pi I Sunyer, Spain; 3) CIBER de Diabetes y Enfermedades Metabólicas, Spain; 4) Ciudad Universitaria, Argentina; 5) Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute, Cambridge, UK; 6) Seattle Children's Hospital Research Institute, USA; 7) School of Biomedical Science, King's College London, UK; 8) London Centre for Paediatric Endocrinology and Metabolism, London, UK; 9) University College London, London, UK; 10) School of Medicine, King's College London, London, UK; 11) Imperial College, London, UK.

The contribution of cis-regulatory mutations to human disease remains poorly understood. Whole genome sequencing can identify all non-coding variants, yet discrimination of causal regulatory mutations represents a formidable challenge.

We performed homozygosity mapping and whole genome sequencing on probands from two consanguineous families with non-syndromic pancreatic agenesis. We first looked for recessive coding mutations in the exomes of these patients, but no causal mutations were identified.

To search for non-coding disease-causing mutations, we annotated the 6,024 rare (<1% frequency in the 1000 genomes project) or novel homozygous variants from these patients using epigenome maps from human pancreatic progenitor cells. Only one variant, a novel SNV, occurred in a functionally annotated region and was shared by both unrelated patients. This variant occurs several tens of kb from PTF1A in a short (~400bp) region of conservation. Follow-up sequencing identified four different recessive point mutations and an 8kb deletion spanning the functional element in 9 of 12 non-syndromic pancreatic agenesis cases.

Experiments including 3C and reporter assays clearly demonstrate that this previously uncharacterised non-coding element acts as a developmental enhancer of PTF1A in human pancreatic progenitor cells. The 6 mutations prevent enhancer activity by abolishing transcription factor binding or deleting the enhancer sequence.

Integrating genome sequencing and epigenomics in a disease-relevant cell type can uncover novel non-coding elements underlying human development and disease.

3082F

Clinical Exome Sequencing identifies de novo mutations in the *MLL* gene causing Atypical Wiedemann-Steiner Syndrome in two unrelated Individuals. *F. Quintero-rivera[1], S.P. Strom[1], J. Mann[4], R. Lozano[5], H. Lee[1], N. Dorrani[3], O. O'Lague[4], N. Mans[5], J.I. Deignan[1], E. Vilain[2,3], S.F. Nelson[1,2,3], W.W. Grody[1,2,3].* 1) Deparment of Pathology and Laboratory Medicine, UCLA School of Medicine, Los Angeles, CA; 2) Deparment of Human Genetics, UCLA School of Medicine, Los Angeles, CA; 3) Deparment of Pediatrics, UCLA School of Medicine, Los Angeles, CA; 4) Kaiser Permanente, Fresno Medical Center; 5) Deparment of Pediatrics, University Of California Davis Medical Center, Sacramento, CA.

Wiedemann-Steiner Syndrome (WSS) is characterized by mild to moderate developmental delay, dysmorphic facial features, and hypertrichosis cubiti (excessive hair on the elbows) [MIM#: 605130]. Here we report two unrelated patients for whom Exome sequencing of parent-proband trios was performed clinically at the UCLA Molecular Diagnostics Laboratories. For patient 1, clinical features at 9 years of age, include developmental delay, craniofacial and multiple minor anomalies: down-slanting palpebral fissures, thick eyebrows and hair, premature dental eruption, fifth finger clinodactyly, tapered fingers, and astigmatism. Patient 2 presented at 1 year of age with intrauterine growth retardation, developmental delay, significant central hypotonia, Rt. microphthalmia and small palpebral fissure, prominent nasal bridge, micrognathia, Lt. hand 3-4 syndactyly, mild pectus excavatum, and small sacral hair tuft. In patient 1, a de novo missense (p.Cys1448Arg) variant was identified in the *MLL* gene (RefSeq transcript ID: NM_001197104.1). In patient 2, a de novo splice site mutation (c.4086+1G>A) was identified in the same gene. *MLL* - Myeloid/Lymphoid Or Mixed Lineage Leukemia Gene - encodes a broadly expressed DNA-binding protein that methylates histone H3K4 and positively regulates expression of target genes, including genes in the HOX and WNT pathways. It is frequently translocated in hematological malignancies, including ALL, AML, and MLL. Mice carrying a hemizygous knockout allele of *MLL* are small at birth and have reduced growth, indicating that dosage of the gene is critical to development. However, the molecular mechanism of *MLL*-related WSS has not been established. Based on clinical and molecular genomic findings, both patients may have novel presentations of WSS. As the hallmark hypertrichosis cubiti was not initially appreciated in either case, this syndrome was not suspected clinically. This report expands the phenotypic spectrum of clinical phenotypes and de novo mutations of the *MLL* gene associated with WSS.

3083W

Follow-up of diagnostic exome sequencing in persons with severe intellectual disability; Re-analysis of data, recurrency screening, CNV detection and genome sequencing. *L. Vissers[1], H. IJntema[1], J. Hehir-Kwa[1], C. Gilissen[1], J. de Ligt[1], R. Leach[2], R. Tearle[2], W. Nillesen[1], B. van Bon[1], M. Willemsen[1], H. Scheffer[1], H. Brunner[1], B. de Vries[1], T. Kleefstra[1], J. Veltman[1].* 1) Dept Human Gen 855, UMC Nijmegen, Nijmegen, Netherlands; 2) Complete Genomics, Inc. 2071 Stierlin Court Mountain View, CA 94043.

In 2012 we described the use of exome sequencing in the diagnostic workup of patients with unexplained severe intellectual disability (ID). In this initial study we detected 16 pathogenic mutations in 100 patients, mostly involving de novo germline point mutations and insertion-deletion events. Here we describe our ongoing genetic studies in this cohort. One of the follow-up studies is aimed at identifying recurrent de novo germline mutations in 21 candidate genes by using various approaches, including targeted next generations sequencing in a large ID cohort. So far these efforts have revealed additional mutations in 7 of our candidate ID genes. Detailed clinical analysis of the patients with mutations showed phenotypic overlap, confirming that these genes cause ID when mutated. A second follow-up study involves the re-analysis of the existing exome data of these patients and their parents using a new software program, Lifescope, which is more sensitive in detecting variants than the program originally used. This analysis provided a positive molecular diagnosis in 5 patients by the identification of de novo mutations in known ID genes. A third follow-up study, which is currently ongoing, is aimed at the identification of CNVs based on read-depth information of exome data using CoNIFER, a software tool that we recently validated in a series of patients with pathogenic CNVs. Finally, we performed whole genome sequencing in 50 of the unexplained ID patients and their parents. Preliminary results of 22 trios confirm the presence of 2 de novo exonic mutations detected only by re-analysis of the original exome sequencing data. In addition whole genome sequencing in these trios revealed 5 de novo exonic mutations in novel candidate ID genes as well as one de novo mutation in the UTR of a known ID gene. While much of this work is still ongoing, it does already demonstrate that the initial diagnostic yield of 16% by exome sequencing is an underestimation; at this moment the number is already at 26%. This yield can be substantially higher by further improving variant detection on exome sequencing data and by follow-up of the many candidate genes in larger clinical cohorts. In addition, whole genome sequencing may not only give a more complete picture of variation present in the exome but will also provide us insight into possible causative mutations in the non-coding part of the genome.

3084T

Mutation analysis of the RBPJ gene in Adams-Oliver syndrome patients. *W. Wuyts, F. de Vogel, E. Van Hul.* Dept Med Gen, Univ Antwerp, Antwerp, Belgium.

Introduction Adams-Oliver syndrome (AOS)is a rare genetic disorder characterized by aplasia cutis congenita (ACC) and terminal transverse limb defects. Mutations in the RBPJ and ARHGAP31 genes have been described as a cause of autosomal dominant AOS, while mutations in the DOCK6 and EOGT genes were shown to be responsible for the autosomal recessive form. We performed mutation analysis of the RBPJ gene in a cohort of AOS patients to determine the relative contribution of RBPJ to AOS. **Materials and Methods** A cohort of 58 patients with clinical characteristics of AOS were included in the study, including familial and sporadic cases. Mutation analysis of the RBPJ gene was performed by sequence analysis of all coding exons. **Results and discussion** In three families an RBPJ missense variant was detected, including one previously reported pathogenic RBPJ mutation (p.Thr441Thr, p.Lys169Glu,p.Phe66Val). Three other intronic variants need further evaluation in a larger control population. In conclusion, mutations in the RBPJ gene seem only present in a limted number of AOS patients.

3085F

Application of targeted next-generation sequencing in the diagnosis of pediatric neurological disorders. *N. Okamoto[1], F. Miya[2], T. Tsunoda[2], M. Kato[3], S. Saitoh[4], M. Yamasaki[5], Y. Kanemura[6,7], K. Kosaki[8].* 1) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 2) Laboratory for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 3) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 4) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 5) Department of Pediatric Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 6) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 7) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 8) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

[Introduction] We have developed a next generation sequencing (NGS)-based mutation screening strategy. This system enabled us to screen 284 genes associated with pediatric neurological disorders. [Materials and Methods]Fifty patients with neurological disorders were included in the study. They are from one institution (Osaka Medical Center). Under the approval by our institutional ethics committee, patients were analyzed by the targeted exome sequence. To capture the target exonic DNAs, we used SureSelectXT Custom capture library (Agilent) for 1.6 Mb exons of neuronal genes capture. The sequence library was constructed using SureSelect XT Target Enrichment System for Illumina Paired-End Sequencing Library kit (Agilent). We performed sequencing the DNA with 101-bp paired-end by 24-multiplex per lane using Illumina HiSeq 2000 sequencer and obtained 1.90 + 0.32 (mean + s.d.) Gb sequence data. To extract disease causal mutation, we excluded known variants found in dbSNP, 1000 Genomes Project, ESP6500 and our control samples , and narrowed the candidates to nonsynonymous and nonsense SNVs. [Results] Several pathogenic mutations were identified. An early termination codon in the *DYRK1A* gene was found in a patient with severe intellectual disability (ID), absent speech, motor disturbance and visual disturbance (Patient 1). Heterozygous mutations in the *CASK* gene were found in two females with mental retardation and microcephaly with pontine and cerebellar hypoplasia (Patient 2,3). A patient with microcephaly and severe ID and showing similar conditions with Kleefstra syndrome had a mutation in the *MBD5* gene(Patient 4). A patient with cutis laxa had an *ALDH18A1* mutation (Patient 5). A female patient with Angelman syndrome like features had mutation in the *GABRD* gene (Patient 6). A patient with ID, ptosis and pachygyria of the frontal lobe had a mutation in the *ACTB* gene. He was compatible with Baraitser-Winter syndrome (patient7). [Discussion] Our NGS-based mutation screening strategy is sensitive and specific in detecting sequence variants in some neurological disorders. We propose that this NGS-based targeted sequencing method would be an alternative to current technologies for identifying the multiple genetic causes of neurological disorders.

3086W

Disruption of the Rac GTPase activator DOCK7 in epileptic enecephalopathy. *F.F. Hamdan[1], J.M. Capo-Chichi[1], B. Maranda[2], G.A. Rouleau[3], J.L. Michaud[1].* 1) Sainte Justine Hospital Research Center, Montreal, Quebec, Canada; 2) Division of Genetics, Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Quebec, Canada; 3) Montreal Neurological Institute, McGill University, Montreal, Qc, Canada.

The genetic basis of intellectual disability remains unexplained in the majority of cases. Here we used exome sequencing to study two sisters (aged 7 and 9 yrs) with severe intellectual disability and intractable epilepsy from a non-consanguineous French Canadian family. We found 186 rare amino acid-altering or splicing variants that were shared between the exomes of the 2 affected siblings. None of these was homozygous. *DOCK7* is the only gene containing compound heterozygous mutations (c.C3709T:p.R1237X and c.2510delA:p.D837fs; NM_001271999) shared by the sisters. Both mutations, which truncate DOCK7 upstream of its GEF domain, were absent from all public SNP databases (1000 Genomes, Exome Variant Server (EVS), dbSNP138), and from our in-house exome dataset (>1000) which includes at least 200 French Canadian individuals. Inspection of *DOCK7* for truncating or splicing variants in the exome variant server data set, revealed only 2 heterozygous truncating variants, in 2 different individuals, out of approximately 13000 sequenced alleles. This indicates that, despite its large size (2129 amino acid), *DOCK7* does not accumulate truncating mutations and, therefore, maybe essential in humans. DOCK7 encodes a guanine exchange factor (GEF) belonging to the DOCK180 super family whose members activate Rho GTPases. DOCK7 is enriched in the brain and has been shown to activate Rac and promote axon formation, axon myelination, and cortical neurogenesis. Based on the importance of DOCK7 in brain function and the severity and rarity of the identified mutations, our data suggests that disruption of *DOCK7* may be responsible for the epileptic encephalopathy observed in this family. Interestingly, DOCK7 is known to physically interact with the TSC1-TSC2 protein complex whose dysfunction causes Tuberous Sclerosis Complex, a multisystemic disorder with high incidence of ID and epilepsy. Identification of additional similarly affected families with *DOCK7* recessive deleterious mutations will be necessary to establish *DOCK7* as a neurodevelopmental disease gene. Collaborations towards this effort are welcome.

3087T

Whole genome sequencing, clinical interpretation, and deep brain stimulation in a severely mentally ill person. *G.J. Lyon[1,2,3], J. O'Rawe[1,2], R. Robison[3], E. Kiruluta[4], G. Higgins[5], M. Reese[4].* 1) Stanley Institute for Cognitive Genomics, One Bungtown Road, Cold Spring Harbor Laboratory, NY, USA, 11724; 2) Stony Brook University, 100 Nicolls Rd, Stony Brook, NY, USA, 11794; 3) Utah Foundation for Biomedical Research, E 3300 S, Salt Lake City, Salt Lake City, UT, USA, 84106; 4) Omicia Inc., 2200 Powell St., Emeryville, CA, USA, 94608; 5) AssureRx Health, Inc., 6030 S. Mason-Montgomery Road, Mason, Ohio 45040.

We report here the detailed phenotypic characterization, clinical-grade whole genome sequencing (WGS), and two year outcome of one man with severe obsessive compulsive disorder treated with deep brain stimulation (DBS) targeting the anterior limb of the internal capsule (ALIC). Since implantation, this man has reported steady improvement, highlighted by a drop in his Yale-Brown Obsessive Compulsive Scale score from ~35 to a score of ~25. A rechargeable Activa RC neurostimulator battery has been of major benefit in terms of facilitating a degree of stability and control over the stimulation. His psychiatric symptoms reliably worsen within hours of the battery becoming depleted, thus providing confirmatory evidence for the efficacy of DBS for OCD in this person. Whole genome sequencing in the CLIA-certified Illumina WGS lab revealed that he is a heterozygote for the p.Val66Met variant in BDNF, encoding a protein that is a member of the nerve growth factor family, and which has been found to predispose carriers to various psychiatric illnesses. He carries the p.Glu429Ala allele in methylenetetrahydrofolate reductase (MTHFR) and the p.Asp7Asn allele in ChAT, encoding choline O-acetyltransferase, which synthesizes the neurotransmitter acetylcholine, with both alleles having been shown to confer an elevated susceptibility to psychoses. We have discovered many other variants in his genome, including pharmacogenetic variants, and have archived and offered the clinical sequencing data to him, so that he and others can re-analyze his genome for years to come. To our knowledge, this is the first N=1 human study in the clinical neurosciences including 1) clinical-grade WGS with management of genetic results for a person with severe mental illness and 2) detailed neuropsychiatric phenotyping and individualized treatment with deep brain stimulation for his OCD. His WGS results and positive outcome with DBS for OCD is one example of individualized medicine in neuropsychiatry, including genomics-guided preventive efforts and brain-implantable devices.

3088F

Genes Make Sense: Seeking Causative Genes for Human Congenital General Anosmia. *A. Alkelai[1], T. Olender[1], P. Tatarskyy[1], V. Boyko[1], D. Oz-Levi[1], I. Keydar[1], R. Milgrom[1], E. Feldmesser[1], E. Ben-Asher[1], E.K. Ruzzo[2], D.B. Goldstein[2], E. Pras[3,4], D. Lancet[1].* 1) Molecular Genetics, The Weizmann Institute of Science, Rehovot, Israel; 2) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA; 3) The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Ramat Gan, Israel; 4) The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

The neuronal and cell-biological mechanisms underlying vertebrate olfaction have been studied in considerable detail, but there is a gap in our understanding of the genetic and molecular basis of chemosensory variability and related monogenic disorders. Congenital general anosmia (CGA) affects <0.1% of general population, and appears in isolated or syndromic forms. While for an appreciable number of syndromic CGA types a causative gene has been identified, knowledge about the genetic basis of the isolated CGA instances has been scarce. Our laboratory has samples from a large CGA cohort, including 66 families of Jewish origin (Feldmesser et al, Chem Senses 32:21-30, 2007). We performed whole-exome capture (SureSelect Human All Exon kit - Agilent Technologies, Santa Clara, CA) and next-generation sequencing (Hiseq 2000 sequencing system, Illumina, Inc. San Diego, CA) in 22 selected individuals from 7 of these families multiply affected with anosmia. Now we report results of the Affymetrix human SNP 6.0 array genotyping, linkage analysis, homozygosity mapping, CNV analysis and whole-exome sequencing in one of these families with X-linked/recessive/dominant with incomplete penetrance mode of inheritance. By using variety of methods we identified number of candidate regions harboring possible CGA pathogenic variants. The best candidate variants are located in such genes as *SPATA5L1* (compound heterozygous mutation), *CALML6* (heterozygous mutation) and *SEMA3A* (heterozygous mutation) and now are being further validated. All the implicated genes appear in our recently constructed general olfactory sensitivity database (GOSdb) (Keydar et al, Hum Mutat, 34: 32-41, 2013). The identification of specific pathogenic functional CGA variants will help to elucidate the molecular basis of general olfactory sensitivity.

3089W

Genetic mapping and whole exome sequencing to unravel the genetic basis of undiagnosed non-syndromic arthrogryposis multiplex congenita. *J. Melki[2,24], J. Maluenda[2], A. Camus[2], F. Nolent[2], L. Fontenas[2], K. Dieterich[2,3,23], J. Zhou[2], N. Monnier[4], P. Latour[5], D. Heron[6], T. Attie-Bitach[7], I. Desguerres[8], S. Lyonnet[7], C. Beneteau[9], C. Bauman[10], D. Bonneau[11], A. Goldenberg[12], C. Bellesme[13], C. Francannet[14], L. Rigonnot[15], S. Sigaudy[16], J. Lunardi[4], S. Odent[17,22], P.S. Jouk[3], M. Granier[18], D. Sternberg[19], I. Gut[20], M. Tawk[2], M. Gonzales[21], A. Laquerriere[1].* 1) Service d'Anatomie et de Cytologie Pathologiques, CHU de Rouen, 1 rue de Germont, 76031 Rouen; 2) UMR-788, Inserm and University Paris 11, 80 rue du Général Leclerc, 94276 Le Kremlin-Bicêtre; 3) Département de Génétique, CHU Grenoble, 38043 Grenoble; 4) Laboratoire de Biochimie et Génétique Moléculaire, CHU Grenoble 217XF , 38043 Grenoble; 5) Service de Neurobiologie, Hopitaux de Lyon, 59 boulevard Pinel, 69677 Bron; 6) Département de Génétique, CHU Pitié-Salpêtrière, 47-83 boulevard de l'Hôpital, 75013 Paris; 7) Département de Génétique, INSERM U781, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75743 Paris; 8) Unité de Neuropédiatrie, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75743 Paris; 9) Service de Génétique Médicale, CHU de Nantes, 9 Quai Moncousu, 44093 Nantes; 10) Département de Génétique, Hopital Robert Debré, 48 boulevard Serrurier, 75935 Paris; 11) Service de Génétique, CHU Angers, 4 rue Larrey, 49933 Angers; 12) Service de Génétique Clinique, CHU Rouen, 1 rue de Germont, 76031 Rouen; 13) Service de Neuropédiatrie, CHU Bicêtre, 80 rue du Général Leclerc, 94276 Le Kremlin-Bicêtre; 14) Unité de Génétique Médicale, CHU Clermont Ferrand, 1 place Lucie Aubrac, 63003 Clermont-Ferrand; 15) Service d'obstétrique, CH Sud-Francilien, 116 Boulevard Jean Jaures, 91108 Corbeil Essonnes; 16) Département de Génétique Médicale, CHU de La Timone, 13385 Marseille; 17) Service de Génétique Clinique, CHU Rennes, Hôpital Sud, 35203 Rennes; 18) Service de Médecine Néonatale, CH Sud-Francilien, 116 Boulevard Jean Jaures, 91108 Corbeil Essonnes; 19) Laboratoire de Génétique, Hopital Pitié-Salpétrière, 47 Boulevard de L'Hopital, 75651 Paris; 20) Centro Nacional de Análisis Genómico, University of Barcelona, Barcelona, 080028, Spain; 21) Service de Génétique et d'Embryologie Médicales, Hopital Trousseau, 26 avenue du Dr. A. Netter, 75571 Paris; 22) UMR6290 CNRS GPLD, Université Rennes1; 23) INSERM U836, Grenoble Institut des Neurosciences, Equipe Muscle et Pathologies, Chemin Fortune Ferrini, 38043 Grenoble; 24) Unité de Génétique Médicale, CH Sud-Francilien, 91108 Corbeil Essonnes.

Arthrogryposis multiplex congenita (AMC) is characterized by congenital contractures of at least two distinct joints of the body. The overall incidence is 1 in 3000 live births. Non-syndromic AMC is the direct consequence of fetal hypo/akinesia sequence which may lead, in addition to AMC, to pterygia, lung hypoplasia, diaphragmatic defect, or cleft palate. Non-syndromic AMCs include a large spectrum of diseases of motor neurons, neuromuscular junction, or skeletal muscle. The difficulty in establishing the genetic diagnosis in AMC patients is likely caused by the high locus heterogeneity and/or non-identified disease genes and the lack of suitable screening methods. In order to gain insight into the underlying cause of these diseases, whole genome scanning using SNP microarrays alone or combined with whole exome sequencing (WES) were performed in a cohort of 32 multiplex and/or consanguineous families with undiagnosed non-syndromic AMC. In these families, extensive clinical and genetic evaluation had not led to established diagnosis. We identified mutations in genes already known to be responsible for AMC or neuromuscular disorders (NMD) in 20 out of 32 (62%) families. The primary targets were skeletal muscle in 12 families, neuromuscular junctions in 4 families, axoglial in 3 families or other target in 1 family. Among known AMC or NMD genes, our study confirmed that mutations of SYNE-1 are responsible for AMC and revealed two AMC families carrying mutations of the TTN gene extending the clinical spectrum of TTN gene mutations. Pathogenic mutations of 4 new genes were identified. The mutations of these genes may be classified as causative (2 genes), highly candidate (1 gene) or possibly causative (1 gene). This is based on gene mutation type, RNA and morphological analysis of the neuromuscular system (including transmission electron microscopy) in patient tissue samples and the availability of mouse models or the generation and characterization of a zebrafish model. NGS technologies may be therefore regarded as the most promising approach to unravel undiagnosed AMC leading to a diagnosis in ~80% of cases as shown in the present study. This will be particularly useful during pregnancy when ultrasound examination reveals AMC or reduced fetal mobility but also when single joint contracture is associated with additional symptoms such as the ones described above. Establishing early diagnosis should provide accurate information on the prognosis.

3090T

Scientific advances from removing the financial barrier- results of the NGS pilot program at Boston Children's Hospital. *C.A. Brownstein[1,2,3], W.A. Wolf[1,3], T.W. Yu[1,3], L.M. Kunkel[1,3], C.A. Walsh[1,3], A.H. Beggs[1,2,3], D.M. Margulies[1,2,3].* 1) Genomics- Research Connection, Boston Children's Hospital, Boston, MA; 2) Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 3) Harvard Medical School, Boston, MA.

Lack of access to new technologies inhibits discoveries. In an effort to equalize access to next generation sequencing (NGS) for its researchers and clinicians, Boston Children's Hospital (BCH) awarded a number of pilot grants to provide whole genome and exome sequencing for patients and human research subjects. Emphasis was on assessing the impact of NGS on analysis of a mix of patients with phenotypes ranging from severely affected individuals with diagnostic dilemmas and for whom testing would have a likelihood of clinical impact, to subjects or cohorts of subjects with phenotypes that were likely to yield important new insights into the biology and treatment of disease. Awards were determined by multidisciplinary, cross-departmental committees, and sequencing was performed by the BCH Genetic Diagnostics Laboratory and LabCorp. Datasets were used to develop an in-house BCH NGS analysis pipeline optimized for human disease gene discovery from pedigrees and cohorts. In addition to providing sequencing data to clinicians and investigators, the project provided a single standard of oversight for the hospital, and evaluated the entire process from selecting patients for sequencing, obtaining their informed consent, generating and analyzing sequencing data, interpreting and reporting clinical results, integrating results into the medical record, seeking reimbursement from payors, and complying with governmental regulations. As of May 2013, the pilot has awarded over 600 exomes and genomes for 42 investigators at BCH (averaging 30 exomes per award), leading to gene discoveries, grants, and publications. A post-sequencing inventory has been performed for 3 researchers awarded grants. From the first award, pathogenic mutations of known genes were identified in 29 patients and 3 novel disease genes have been discovered with one paper published, and two more papers in preparation. From the second award, one paper has been submitted, 3 patients now have diagnoses, and two have had treatment changes as the result of their new diagnoses. From the third award, one grant has been written, and two papers are in preparation. In less than 18 months, results of this program have served to better understand disease, improve diagnosis, develop a robust and institutionally viable NGS pipeline, and ultimately improve treatment of our patients. Sequencing will now be offered to BCH researchers and clinicians at a reduced cost to maintain this pace of discovery.

3091F

Alu Yb8 insertion near SMN1 exon 7 as a rare cause of SMA. *S.M. Kirwin[1], K.M.B. Vinette[1], I.L. Gonzalez[1], S.L. Dugan[2], K.J. Swoboda[3], T.M. Newcomb[3], V.L. Funanage[1].* 1) Molecular Diagnostics Laboratory, Nemours/duPont Hospital for Children, Wilmington, DE; 2) Division of Medical Genetics, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 3) Pediatric Motor Disorders Research Program, Department of Neurology, University of Utah School of Medicine, SLC, UT.

Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder primarily affecting young children. Approximately 95% of cases of SMA are caused by homozygous deletion of exon 7 of the SMN1 gene, although up to 5% of cases can arise by heterozygous deletion of a single SMN1 copy paired with a subtle mutation elsewhere within the gene. We report a novel SMN1 gene mutation that involves insertion of an Alu element at an L1 endonuclease recognition site, coinciding with the intron 6 branch point. The proband originally presented with severe muscular atrophy, tongue fasciculations, and ventilator dependency, with no family history of SMA. MLPA testing at another lab had reported a single functional copy of SMN, and gene sequencing was requested at our molecular diagnostic laboratory. Mutations at the branch point of intron 6 have been reported to affect inclusion of exon 7 in the SMN product; however, the position of the insertion precludes detection by the MLPA assay used, as the primers are 3' to the inserted Alu. Our SMN quantitative assay uses primers within intron 6 and intron 7, and detected zero copies of SMN1, 1 copy of SMN2, and a product of an unusual size that did not appear to digest with DraI. This product was purified, sequenced, and revealed the presence of a normal SMN1 exon 7 preceded by a complete Alu element of the Yb8 family. Members of the Yb8 family of Alu elements are described as currently active in retroposition, and there are reports of de novo Alu Yb8 insertions causing disease. The 'A' tail found on the Alu in our patient's SMN1 gene was 40 bp in length. Recently inserted Alus are characterized by A-tails between 40 and 97bp. Parental testing revealed that the father carries one copy of SMN1, whereas the mother carries two copies of SMN1, one containing the inserted Alu. Since the first submission of the proband's sample for testing, an additional affected sibling was born, and was shown to carry only the same mutated SMN1 allele, thus confirming the pathogenicity of the Alu insertion in this family. Unique neuroimaging features present in both brothers and not typically seen in 5q SMA include generalized cortical atrophy, hypoplastic appearing optic nerves and possible CNS myelination abnormalities. The severe infantile phenotype in association with congenital weakness and clearly evident central nervous system involvement expand the phenotype previously associated with infantile onset 5q SMA.

3092W

Identification of a novel *GATA3* mutation in a Taiwanese family with idiopathic sensorineural hearing impairment by massively parallel sequencing. *Y.H. Lin[1,2], C.C. Wu[1,3], T.Y. Hsu[4], W.Y. Chiu[5], C.J. Hsu[1], P.L. Chen[2,3,5,6]*. 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 3) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 4) Department of Otolaryngology, E-DA Hospital, I-Shou University, Kaohsiung, Taiwan; 5) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 6) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan.

Hearing impairment is a genetically heterogeneous condition with > 100 deafness genes identified thus far. Recent studies have confirmed the utility of massively parallel sequencing (MPS) in addressing the genetically heterogeneous hereditary hearing impairment. In our previous study, we used Illumina HiSeq2000 to sequence 80 genes in 12 multiplex families, and identified 2 known mutations and 2 novel mutations as the causative mutations in 4 families. Recently, we further upgraded the screening panel to include 131 deafness genes, and applied the panel to another 12 unrelated multiplex families with idiopathic sensorineural hearing impairment. Criteria for data filtering included: allele frequencies <5%, both PolyPhen2 and SIFT scores >0.95, Sanger sequencing, segregation pattern, and evolutionary conservation of amino acid residues. In the proband of an autosomal dominant family with 9 affected members, we found a novel single nucleotide deletion mutation, c.149delT, in exon 2 of *GATA3*. If translated, c.149delT should result in a frameshift with a premature stop codon after a new amino acid sequence of 144 residues (p.Leu50fsX144). Heterozygosity for c.149delT was identified in all the 9 affected family members and 1 family member with normal hearing; the absence of hearing phenotypes in the latter might be attributed to his young age. *GATA3* haploinsufficiency is thought to be associated with HDR syndrome which is characterized by hypoparathyroidism, deafness, and renal anomalies. We examined the clinical features in 7 affected members in the family. All the 7 subjects revealed mild to moderate sensorineural hearing impairment. By contrast, except for 1 subject with bilateral parenchymal renal disease, the other 6 presented no signs of hypoparathyroidism or renal anomalies, indicating relatively mild phenotypes in this family. To our knowledge, the present study represents the first report to achieve genetic diagnosis before the clinical diagnosis of HDR syndrome is made. Compared to hearing impairment which is the most penetrant feature of HDR syndrome and usually develops at an earlier age, symptoms/signs of hypoparathyroidism and renal anomalies may remain clinically undetected at the time of assessment. Accordingly, genetic screening for multiple deafness genes with MPS might be helpful in identifying certain types of syndromic hearing loss such as HDR syndrome, contributing to earlier diagnosis and earlier treatment in the affected individuals.

3093T

DNM2 - related centronuclear myopathy found with the use of exome sequencing in a patient with clinical diagnosis of peripheral neuropathy. *D. Safka Brozkova[1], J. Haberlova[1], M. Gonzales[2], S. Zuchner[2], P. Seeman[1]*. 1) Charles University 2nd Medical School and University Hospital Motol, Prague, Czech Republic; 2) University of Miami Miller School of Medicine, Miami, Florida.

Centronuclear myopathy is a slowly progressive muscular disease, where autosomal dominant form is caused by mutations in DNM2 gene. Peripheral neuropathy is a neurological disease that is very heterogeneous and can be caused by mutations in more than 40 genes, therefore to find the right cause of the neuropathy can be difficult. A patient with healthy parents and 3 healthy siblings was affected with muscle weakness on upper and lower limbs and reduced facial expression. The electrophysiology examination showed the neurological lesion. First genetical examination was focused on deletion of the exon 7 and 8 in SMN1 gene and did not reveal any deletion. Subsequent Sanger sequencing of peripheral neuropathy associated genes - HSP22, SH3TC2, HSP27, MFN2, BSCL2, GDAP1 did not reveal any pathological variant. Since the Sanger sequencing of all neuropathy associated genes would be much expensive, we decided to perform the exome sequencing of the affected patient. According to a pedigree we were looking for the recessive or de novo variant. The filtering criteria for candidate variants were as follows: non synonymous coding or splice-site variant, not present in NHLBI EVS, GERP score >3 or PhastCons score > 0.6. No recessive variant was found. 43 variants were revealed with de novo mode of inheritance. After the detailed research, only one variant p.E368K, in DNM2 gene was already reported as the cause of centronuclear myopathy. Sanger sequencing confirmed the variant in affected patient, variant was missing in parents. The phenotype of the p.E368K mutation is in between the myopathy and neuropathy. The exome sequencing helped to refine the clinical diagnosis. Thanks to the already published variant, the exome seqencing of the parents was not needed. Support- IGA MHCZ NT 14348.

3094F

Comprehensive molecular diagnosis of 179 Leber congenital amaurosis and juvenile retinitis pigmentosa patients by targeted next-generation sequencing. *X. Wang[1,2], H. Wang[1,2], V. Sun[5], H. Tuan[1], V. Keser[5], K. Wang[2], H. Ren[5], I. Lopez[5], J.E. Zaneveld[1,2], S. Siddiqui[5], S. Bowles[1], A. Khan[5], J. Salvo[1,4], S.G. Jacobson[6], A. Iannaccone[7], F. Wang[1,2], D. Birch[8], J.R. Heckenlively[9], G.A. Fishman[10], E.I. Traboulsi[11], Y. Li[1,2], D. Wheaton[8], R.K. Koenekoop[5], R. Chen[1,2,3,4]*. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA; 3) Program in Developmental Biology, Baylor College of Medicine, Houston, USA; 4) Structural and Computational Biology & Molecular Biophysics Graduate Program, Baylor College of Medicine, Houston, USA; 5) McGill Ocular Genetics Laboratory (MOGL), Montreal Children's Hospital, Departments of Paediatric Surgery, Human Genetics and Ophthalmology, McGill University Health Center, Montreal, Quebec, Canada; 6) Scheie Eye Institute, Department of Ophthalmology, University of Pennsylvania, Philadelphia, USA; 7) Hamilton Eye Institute, University of Tennessee Health Science Center, Memphis, USA; 8) Retina Foundation of the Southwest and Dept. of Ophthalmology, University of Texas Southwestern Medical School, Dallas, USA; 9) Center for Retinal and Macular Degeneration, University of Michigan, Ann Arbor, USA; 10) The Chicago Lighthouse for the Blind and Visually Impaired, Chicago, USA; 11) Ophthalmology, Cleveland Clinic, Cleveland, USA.

Background: Leber congenital amaurosis (LCA) and juvenile retinitis pigmentosa (RP) are inherited retinal diseases that cause early-onset severe visual impairment. An accurate molecular diagnosis can refine the clinical diagnosis and allow gene-specific treatments. Methods: We developed a capture panel that enriches the exonic DNA of 163 known retinal disease genes. Using this panel, we performed targeted next-generation sequencing (NGS) for a large cohort of 179 unrelated and prescreened patients with the clinical diagnosis of LCA or juvenile RP. Systematic NGS data analysis, Sanger sequencing validation, and segregation analysis were utilized to identify the pathogenic mutations. Patients were re-visited to examine the potential phenotypic ambiguity at the time of initial diagnosis. Results: Pathogenic mutations for 72 patients (40 percent) were identified, including 45 novel mutations. Of these 72 patients, 58 carried mutations in known LCA or juvenile RP genes and exhibited corresponding phenotypes, while 14 carried mutations in retinal disease genes that were not consistent with their initial clinical diagnosis. We re-visited patients in the latter case and found that homozygous mutations in PRPH2 can cause LCA/juvenile RP. Guided by the molecular diagnosis, we re-classified the clinical diagnosis in 2 patients. Conclusions: We have identified a novel gene and a large number of novel mutations that are associated with LCA/juvenile RP. Our results highlight the importance of molecular diagnosis as an integral part of clinical diagnosis.

3095W

A Founder Mutation in *COL4A3* Causes Autosomal Recessive Alport Syndrome in the Ashkenazi Jewish Population. *B.D. Webb[1], T. Brandt[1], L. Liu[1], J. Liao[1], C. Jalas[2], A. Fedick[3], M.D. Linderman[4], G.A. Diaz[1], R. Kornreich[1], H. Trachtman[5], L. Mehta[1], L. Edelmann[1]*. 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Bonei Olam, Center for Rare Jewish Genetic Disorders, Brooklyn, NY; 3) Microbiology and Molecular Genetics, UMDNJ- Robert Wood Johnson Medical School, Piscataway, NJ; 4) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Pediatrics, New York University School of Medicine, New York, NY.

Alport syndrome is a genetically heterogeneous inherited progressive nephropathy that arises from mutations in the type IV collagen genes, *COL4A3*, *COL4A4*, and *COL4A5*. Symptoms may also include sensorineural hearing loss and ocular lesions. Here we determined the molecular basis of Alport syndrome in a non-consanguineous Ashkenazi Jewish family with multiple affected females by performing linkage analysis followed by targeted next generation sequencing. We identified a homozygous *COL4A3* mutation, c.40_63del, in affected individuals. Interestingly, we observed that the mutant alleles in the affected children were inherited from each parent on partially conserved haplotypes. Large-scale population screening of 2,017 Ashkenazi Jewish samples was next completed and we determined the carrier frequency of this mutation to be 1 in 183. These findings suggest that the *COL4A3* c.40_63del is a founder mutation and that this mutation may be a common cause of autosomal recessive Alport syndrome in the Ashkenazi Jewish population. Additionally, we have determined that heterozygous carriers of the *COL4A3* c.40_63del mutation in this family do not meet criteria for a diagnosis of Thin Basement Membrane Nephropathy, and we conclude after reviewing the literature that most patients heterozygous for this mutation will not develop benign familial hematuria.

3096T

De novo frameshift mutation in ASXL3 in a patient with global developmental delay, microcephaly, and craniofacial anomalies. *D.L. Dinwiddie[1,2,3,4,5,6], S.E. Soden[3,4,6], C.J. Saunders[3,4,5,6], N.A. Miller[3,4], E.G. Farrow[3,4,6], L.D. Smith[3,4,6], S.F. Kingsmore[3,4,5,6].* 1) Department of Pediatrics, University of New Mexico Health Sciences Center, Albuquerque, NM; 2) Clinical Translational Science Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico; 3) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, Missouri; 4) Department of Pediatrics, Children's Mercy Hospital, Kansas City, Missouri; 5) Department of Pathology, Children's Mercy Hospital, Kansas City, Missouri; 6) School of Medicine, University of Missouri-Kansas City, Kansas City, Missouri.

Currently, diagnosis of affected individuals with rare genetic disorders can be lengthy and costly, resulting in a diagnostic odyssey and in many patients a definitive molecular diagnosis is never achieved despite extensive clinical investigation. The recent advent and use of genomic medicine has resulted in a paradigm shift in the clinical molecular genetics of rare diseases and has provided insight into the causes of numerous rare genetic conditions. In particular, whole exome and genome sequencing of families has been particularly useful in discovering *de novo* germline mutations as the cause of both rare diseases and complex disorders. We present a six year old, African American female with microcephaly, autism, global developmental delay, and metopic craniosynostosis, who is nonverbal, and suffered intrauterine growth restriction. Exome sequencing of the patient and her two parents revealed a heterozygous two base pair *de novo* deletion, c.1897_1898delCA, p.Gln633ValfsX13 in the ASXL3 gene, predicted to result in a frameshift at codon 633 with substitution of a valine for a glutamine and introduction of a premature stop codon. To better understand the rare, nonsynonymous variant burden and impact of mutations in the ASXL3 gene, we examined the variants of a frequency of less than 1% in the ASXL3 gene in an internal variant database that contains clinical phenotype and variant information from more than 1,300 exomes. 29 variants of with a frequency of less than 1 percent were discovered. Characterization of these 29 rare variants revealed 19 that were predicted to deleterious by SIFT and 11 predicted to be probably or possibly damaging by PolyPhen2. Phenotypic evaluation of the samples with these rare, non-synonymous variants revealed zero patients with developmental delay, microcephaly, or other craniofacial anomalies, suggesting that these variants are unlikely to be pathogenic in a heterozygous state. In conclusion, we provide additional evidence that, indeed truncating and frameshifting mutations in the ASXL3 gene are the cause of a newly recognized disorder characterized by severe global developmental delay, small birth size, and craniofacial anomalies. Furthermore, we expand the knowledge about disease causing mutations and the genotype-phenotype relationships in the ASXL3 gene and provide evidence that rare, nonsynonymous, damaging mutations are not associated with developmental delay or microcephaly.

3097F

Unraveling disease genes causing autosomal recessive disorders in Qatari population by whole exome sequencing. *S. Fahiminiya[1], M. Almuriekhi[2], Z. Nawaz[3], KH. Abu Khadija[3], J. Majewski[1], T. Ben-Omran[2].* 1) Department of Human Genetics, Faculty of medicine, McGill University and Genome Quebec Innovation Center, Montreal, Quebec, Canada; 2) Section of Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Doha, Qatar; 3) Section of Cytogenetic, Department of Pathology and Laboratory Medicine, Hamad Medical Corporation, Doha, Qatar.

Background: Whole Exome Sequencing (WES) was applied as a molecular diagnostic tool to identify disease-causing mutations in autosomal recessive disorders (ARD) in Qatari population where alternative molecular diagnostic tools had failed to detect pathogenic variants in causal genes. ARD are usually severe and rare, resulting from the transmission of one defective allele by each parent, which occurs with higher rate in consanguineous families. The prevalence of consanguineous marriage is high in Qatar (47%) that results in a higher incidence of several ARD. To reduce the overall socio-economic burden of such diseases, the development of diagnostic tools and prevention strategies is a priority for our population. To achieve these goals, the genes causing human genetic diseases should be first discovered. Methods: Whole exome capturing, sequencing and bioinformatics analyses were performed using our standard protocols at Genome Quebec Innovation Center, Canada. Results: WES was performed on 26 consanguineous Qatari families. The mode of inheritance was assumed to be AR because of unaffected status of parents, consanguinity and having equally affected male and female children. In families with one or only male affected child, de novo and X-linked inheritance were also considered. This led to the identification of definitive causal mutations in 12 families: Hypophosphatemic rickets; Hurler syndrome; Glycogen storage disease; Noonan-Like syndrome; Seckel syndrome; Geleophysic dysplasia; Limb-girdle muscular dystrophy; Multiple Fractures; Metachromatic Leukodystrophy; Immunodeficiency and Juvenile onset cataract. For 6 other families, we identified several candidate (1-26) genes in which the validation or functional studies are in progress: mental retardation; CNS anomaly; eye anomalies; peripheral neuropathy; axonal peripheral neuropathy and Oro-facio-digital syndrome. The 8 remaining families are still inconclusive, but further bioinformatics analysis and comparisons with future samples may yet reveal pathogenic cause. Conclusion: Our study highlights that WES is a powerful molecular approach for discovery pathogenic gene mutations- especially when traditional molecular genetic screening has failed. Furthermore, with decreasing sequencing costs and improving analysis pipelines, we expect WES to be in widespread clinical use in the near future that will help us to decrease time and cost of diagnosis and to focus on appropriate treatment and supportive care.

3098W

Exome sequencing unveils novel disease-causing variation in a Charcot-Marie-Tooth disease cohort. *C. Gonzaga-Jauregui[1], T. Harel[1], D. Pehlivan[1], Y. Okamoto[1], W. Wiszniewski[1], D. Muzny[2], R.A. Gibbs[1, 2], J.R. Lupski[1, 2, 3], Centers for Mendelian Genomics.* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Charcot-Marie-Tooth (CMT) disease is the most common hereditary neuropathy affecting approximately 1/2500 individuals. CMT is a clinically heterogeneous distal symmetric polyneuopathy (DSP) with two major groups distinguished electrophysiologically: demyelinating CMT1 and axonal CMT2. Although the major cause of CMT, the 1.4 Mb duplication CNV of 17p11.2 that is responsible for ~70 percent of the CMT1 cases, is used extensively in clinical diagnosis and part of the evidenced based practice guidelines in the United States for evaluation of DSP, CMT shows extensive underlying genetic heterogeneity with about 50 loci identified or linked to date to the different subtypes of the disease. Furthermore, homozygous CMT1A duplication CNV and triplication CNV can portend a more severe clinical course while each conveys distinct genetic recurrence risk. Exome sequencing allows assessing all the coding variation in the 2 percent of the human diploid genome that we can interpret through our knowledge of the genetic code; however, personal exomes contain ~10,000 - 12,000 nonsynonymous variants and many of these are novel. Further, even in genetic conditions with known responsible genes, interpretation can be complicated by the presence of novel variants in more than one causative gene. We have performed exome sequencing at high coverage (> 100X) of a cohort of 39 patients with different clinical presentations of CMT, in whom the genetic cause had not previously been identified using a multitude of clinically available and research molecular genetic analyses. We have found the apparent causative mutations in several individuals, and potentially disease causing mutations in novel genes in additional subjects. We also show that affected individuals can have known or novel rare variants in multiple CMT genes, possibly contributing to the mutational load and phenotypic variation in the disease. These findings are consistent with the recently proposed Clan Genomics hypothesis [Cell (2011);147(1):32-43] which posits that new mutations (in both CNV and SNV) in patients or those that arose in recent ancestors in the family or clan, and novel combinations from the proband's parents, but NOT common/ancient alleles studied in populations by Genome-Wide Association Studies (GWAS), account for many medically actionable variants.

3099T

A Novel Missense Mutation in EDAR Underlies the Autosomal Recessive Hypohidrotic Ectodermal Dysplasia with Bilateral Amastia and Palmoplantar Hyperkeratosis. *H. Haghighi-Kakhki[1], A. Haghighi[2,3], P. Nikuei[4], N. Saleh-Gohari[5], S. baghestani[6], P. Krawitz[7], J. Hecht[8], S. Mundlos[8].* 1) Masshad Azad University, Mashhad, Iran; 2) Wellcome Trust Centre for Human Genetics, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Hormozgan Fertility and Infertility Research Center, Hormozgan University of Medical Sciences, Bandarabbas, Iran; 5) Genetic Department, Kerman University of Medical Sciences, Kerman, Iran; 6) Department of Dermatology, Hormozgan University of Medical Sciences, Bandar Abbas, Iran; 7) Institut für Medizinische Genetik und Humangenetik, Charité Universitätsmedizin Berlin, Berlin, Germany; 8) Max-Planck-Institut für molekulare Genetik, Ihnestr. 63-73, 14195 Berlin, Germany.

Ectodermal dysplasias (EDs) are a large group of heritable complex conditions with more than 200 members and common clinical characteristics of anomalies of the hair, teeth, nails, and sweat glands with or without involvement of other organs. Anhidrotic ectodermal dysplasia is a rare inherited disorder characterized by a congenital dysplasia of ectodermal structures and their accessory appendages manifested primarily by hypohidrosis, hypotrichosis and hypodontia. Autosomal recessive ectodermal dysplasia is caused by mutations in EDAR gene. Using whole exome sequencing, we identified novel compound homozygous missense mutations in EDAR, c.338G>A (p.C113Y) and c.1037C>T (p.T346M), causing the disease in an extended consanguineous kindred. The proband presented with sparse hair, absent sweating, hypodontia, and conical teeth. The unusual features of absence of breasts and palmoplantar keratoderma were also noted. The findings of our study expand the knowledge on genotype-phenotype correlations in EDAR and will have important implications for genetic screening and diagnosis. This is the first genetic study of anhidrotic ectodermal dysplasia in Persian population.

3100F

Whole-exome sequencing and genome-wide homozygosity mapping analysis of fetal autopsy tissue reveals a putative pathogenic frameshift mutation in *OBSL1*, consistent with a diagnosis of 3-M Syndrome. *A.C. Lionel[1,2], R.F. Wintle[1], S. Farrell[3], D. Cushing[3], T. Paton[1], T.L. Stockley[4], D.J. Stavropoulos[4], P.N. Ray[4], M. Szego[1,2,5], L. Lau[1], S.L. Pereira[1], C.R. Marshall[1,2], S.W. Scherer[1,2].* 1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) McLaughlin Centre and Department of Molecular Genetics, the University of Toronto, Ontario, Canada; 3) Genetics Department, Credit Valley Hospital, Mississauga, Ontario, Canada; 4) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 5) Centre for Clinical Ethics, the University of Toronto, Ontario, Canada.

We report here a consanguineous couple who have experienced three consecutive pregnancy losses following the fetal ultrasound finding of short limbs caused by retardation of long bone growth. Post-termination autopsies revealed some proximal limb shortening in two fetuses, and a spectrum of mildly dysmorphic features. Radiographs showed no evidence of skeletal dysplasia and karyotype was normal. High resolution microarray genotyping using the Illumina Infinium HumanOmni2.5-quad BeadChip was performed for the three fetuses and the parents using DNA from autopsy material and whole blood respectively. Copy number variation (CNV) analysis did not reveal any potentially pathogenic CNVs that were present in all three fetuses. Homozygosity mapping revealed two regions present in all three affected fetuses that reached genome-wide significance: one of 9.7 Mb at chromosome 2q34-q35, and another of 35.7 Mb at chromosome 8q13.2-q22.3. These regions were both heterozygous in the parents. Paired-end whole exome sequencing was performed on two of the affected fetuses on a Life Technologies SOLiD 5500xl platform with target enrichment using the Agilent SureSelect 50 Mb human all exon capture kit. We prioritized novel, nonsynonymous variants that were in the regions of shared homozygosity. The most plausible candidate was an insertion variant in the gene *OBSL1* on chromosome 2 (c.1273insA), resulting in a frameshift and premature stop (p.T425nfsX40). This variant was validated by Sanger sequencing, and showed segregation consistent with an autosomal recessive model of disease inheritance (i.e., heterozygous in both parents, homozygous in all three fetuses). This frameshift variant in *OBSL1* has been previously observed in patients with 3-M syndrome, an autosomal recessive disorder characterized by severe pre- and postnatal growth retardation. Prenatal diagnosis of 3-M syndrome based on ultrasound is findings is unreliable, given that intrauterine growth retardation has many causes and is not specific for this syndrome. Additionally, although growth retardation of long bones in utero has been reported, this finding is not always seen in children with 3-M syndrome. Exome- or whole-genome sequencing of additional 3-M syndrome families will help identify mutations in *OBSL1* and other genes. Our study provides novel insight into the early clinical manifestations of 3-M syndrome, and demonstrates the utility of exome sequencing as a tool for prenatal diagnosis.

3101W
Functional Evaluation of candidate mutations identified in whole exome sequneces of patients with undiagnosed diseases. *Y. Lu[1], P. Xie[1], E. Ruzzo[1], A. Need[1], V. Shashi[2], Y. Jiang[2], X. Zhu[1], D. Goldstein[1].* 1) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina, USA; 2) Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina, USA.

A number of recent studies suggest that whole exome sequencing of both patients and parents (trio sequencing) can resolve up to 50 % of undiagnosed childhood diseases (Need et al. 2012). We have recently sequenced 15 such trios and evaluated functional consequences of candidate mutations. Here we report functional consequences of three mutations of particular interest. A proband with severe osteopetrosis, autism, and developmental delay, carries a homozygous nonsynonymous mutation, R95H, in sorting nexin 10 (SNX10). SNX10 is essential for osteoclast formation and resorption, which the malfunction will result in osteopetrosis. Our data from in vitro analysis suggests that this mutation leads to loss of function of SNX10. Another hemizygous nonsynonymous mutation (rs148886271; P670S) in BCL6 corepressor (BCOR) was found to down-regulate protein expression in a well-characterized 1721 amino acid (a.a.) isoform, but induced a greater than 50-fold up-regulation in another abundant 1703 a.a. isoform whose function is uncharacterized. The evidence together suggests a complicated influence on normal BCOR function as a transcriptional factor co-regulator and possibly leads to altered downstream transcriptional activation that may cause global developmental delay observed in this patient. Several other genes which are of great interests but still lack functional evidence include ANK3, which has been implicated as a candidate gene in epilepsy and autism. Compound heterozygous mutations in ANK3 were found in one patient with severe intractable seizures in our cohort. We are planning to proceed functional evaluation of ANK3 through mouse hippocampus primary culture. Further studies, including confirming molecular results by patients RNA and protein samples are still required to fully understand the influence of these variants in vivo. Overall, our work illustrates the central role of functional evaluation in the interpretation of personal genomes, since many of the most interesting mutations have not been characterized before. Our work also illustrates that patients with undiagnosed diseases may carry mutations that can inform about human biology more generally, such as the mutations in SNX10 and their potential relevance to bone related disorders.

3102T
Whole exome sequencing identifies novel mutations in three families with GAPO syndrome. *D. Pehlivan[1], E. Karaca[1], T. Gambin[1], S.N. .Jhangiani[2], G. G. Gonzaga-Jauregui[1], Y. Bayram[1], W. Wiszniewski[1], A.H. Cebi[1], D. Muzny[2], M.M. Atik[1], R.A. Gibbs[2], M. Selman Yildirim[3], A. Zamani[3], D. Gul[4], N.H. Elcioglu[5], B. Bozkurt[6], J.R. Lupski[1,7,8].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 3) Necmettin Erbakan University, Meram Medical Faculty, Department of Genetics, Konya, Turkey; 4) Department of Medical Genetics, Gülhane Military Medical Academy, Ankara, Turkey; 5) Department of Pediatric Genetics, Marmara University Medical Faculty, Istanbul, Turkey; 6) Selcuk University Medical Faculty, Department of Ophthalmology, Konya, Turkey; 7) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 8) Texas Children's Hospital, Houston, TX, USA.

GAPO syndrome (MIM# 230740) is the acronym for Growth retardation, Alopecia, Pseudoanodontia, and Optic atrophy. So far about 30 cases have been reported worldwide, making it one of the rarest recessive conditions. Distinctive craniofacial features including alopecia, rarefaction of eyebrows and eyelashes, frontal bossing, high forehead, midfacial hypoplasia, hypertelorism, and thickened eyelids and lips make it a readily recognizable phenotype. However there are additional clinical findings in each reported case. Mutations in ANTXR1 has recently been found to be causative for GAPO syndrome. ANTXR1 plays a role in the regulation of extracellular matrix.In this study we applied whole exome sequencing in three GAPO syndrome families with six affected individuals. Exome sequencing analysis revealed two novel mutations, one frameshift, one splice site, and previously described mutation in the ANTXR1 gene. We here present clinical findings of patients with GAPO syndrome. Our patients contribute to make a better genotype-phenotype correlation in this rare condition and are important to understand the role of ANTXR1 in the function/dysfunction of the extracellular matrix.

3103F
A genome-wide catalogue of genetic variants for nephrotic syndrome via whole genome sequencing. *M.G. Sampson[1,4], A. Tan[2,4], C.A. Gadegbeku[3,5], J. Sedor[3,6], M. Kretzler[3,4], H.M. Kang[2,4], Nephrotic Syndrome Study Network (NEPTUNE), The Michigan O'Brien Renal Center (CPROBE).* 1) Pediatrics and Communicable Disease; 2) Public Health-Biostatistics; 3) Internal Medicine-Nephrology; 4) University of Michigan, Ann Arbor, MI; 5) Temple University, Philadelphia, PA; 6) Case Western Reserve University, Cleveland, OH.

Nephrotic syndrome (NS) is a rare glomerular condition with more than 20 known Mendelian genes and few common risk variants. Yet, these explain a minority of the population prevalence across the lifespan, particularly in adults. To identify known and novel variants associated with NS molecular, histologic, and clinical phenotypes, we are performing low-pass whole genome sequencing (WGS) and exome chip genotyping on 550 affected subjects recruited into observational NS cohorts from North American medical centers. Here we present initial data on the first 256 patients. Subjects were either recruited at time of initial indicated biopsy for suspected primary NS or had an existing diagnosis of focal segmental glomerulosclerosis (FSGS) or membranous nephropathy (MN). All underwent Illumina Exome Chip genotyping and Illumina HiSeq WGS. WGS data underwent GotCloud pipeline with linkage disequilbrium refinement, and comparison to Exome Chip genotypes. The variant frequencies were compared to the latest release of 1000 Genomes and Exome Sequencing Projects, stratified by ancestry. Diagnoses were 52% minimal change disease/FSGS, 25% MN, and 24% other glomerulopathy. Subject ancestry was 49% European (EUR), 29% African (AFR), 15% Admixed American (AMR), and 7% Asian (ASN). Mean sequencing depth was 4.2x. 20.9M SNPs were identified, with transition to transversion ratio (Ts/Tv) 2.16. 330K (16%) SNPs were novel to dbSNP Build 135. Comparing with exome chip data, low-pass WGS was estimated to have 89% power to detect shared variants in 256 cases and 30% power to detect singleton variants. We identified novel loss-of-function and essential splice-site variants shared among our cases in 55 and 64 genes, respectively. In AFR FSGS cases, the minor allele frequency (MAF) of the known FSGS risk haplotype, G1, in APOL1 was 31% (OR=1.6). In EUR MN cases, the allele frequencies of the known MN risk SNPs, rs2187668 and rs4664308, were 33% (OR=4.4) and 83% (OR=3.5), respectively. At the time of abstract submission, we are systematically performing genome-wide association analysis and testing the enrichment of associated variants in regulatory elements. We plan to integrate this genetic data with kidney biopsy-derived RNA-Seq, histologic, and outcome data from the same patients. This should clarify existing, and define new, associations between genetic variation and both endophenotypes and clinical outcomes.

3104W
Personalized functional genomics approach elucidates novel Mendelian disease genes and provides proof of pathogenicity for variants of uncertain significance. *P. Bonnen[1,2], A. Besse[1,2], S. Lalani[2], W. Craigen[2], F. Scaglia[2], R. McFarland[3], C. Bacino[2], R. Taylor[3], K. Scott[2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Wellcome Trust Centre for Mitochondrial Research, The Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

Genome-wide sequencing is now at the forefront of our efforts to identify genetic underpinnings of disease. Bioinformatic analyses of sequence data alone can only predict the impact of variants in known disease genes. For discovery of novel disease genes and to provide proof of pathogenicity of specific alleles, we combine genome-wide sequencing efforts with robust pipelines that allow functional characterization of gene variants in patient cells. This scalable infrastructure is intended to accelerate validation of mutations driving single gene disorders with the ultimate goal of translating annotated variants into clinical diagnostics and therapeutics. Combining the resolution of genome-wide sequencing with functional profiling of variants in patient cells enables the personalized identification of bona fide pathogenic mutations. Our approach combines whole exome sequencing with (1) consistent, systematic and thorough functional profiling on every study patient regardless of variation in clinical presentation, (2) bioinformatic analyses to prioritize pathogenic variants, (3) delivery of the wild-type copy of suspected pathogenic genes to rescue cellular dysfunction and (4) delivery of patient mutations with simultaneous knockdown of the endogenous gene of interest into healthy cells to recapitulate cellular disease phenotypes. Our approach has already revealed several novel disease genes, and importantly, has enabled validation of the pathogenicity of alleles reported as variants of uncertain significance (VUS) in the diagnostic setting. Our current study cohort consists of children with pediatric onset diseases with clinical and diagnostic assessment that present with mitochondrial, neuromuscular, neurological, cardiomyopathy, bone, and rapid aging diseases. Our personalized functional genomics approach has discovered novel disease genes and validated pathogenic mutations in greater than 60% of this cohort. This approach serves as a model for validation of sequence-based genetic findings, and our results provide new insights into the mechanisms underlying the pathology of these single gene disorders and delivers genes and pathways for further study of the pathogenetic mechanisms underlying diseases with more complex etiology.

3105T

Genotype-phenotype correlation in a national mutation study of Danish patients with HHT. *P.M. Torring[1,2], K. Brusgaard[1], L.B. Ousager[1], P.E. Andersen[3], A.D. Kjeldsen[2], HHT centre, Odense University Hospital.* 1) Dept. of Clinical Genetics, Odense University Hospital, Odense, Denmark; 2) Dept. of Otorhinolaryngology, Odense University Hospital, Odense, Denmark; 3) Interventional Radiology, Odense University Hospital, Odense, Denmark.

Purpose: Hereditary Haemorrhagic Telangiectasia (HHT) is an autosomal dominantly inherited vascular disease characterized by the presence of mucocutaneous telangiectasia and visceral arteriovenous malformations (AVM). About 85% of HHT patients carry mutations in the ENG, ACVRL1 or SMAD4 genes. Here, we report on the genetic heterogeneity in the Danish national HHT population and address the prevalence of pulmonary AVM (PAVM). Methods: Probands of 107 apparently unrelated families received genetic testing, including sequencing and multiplex ligation-dependent probe amplification (MLPA) analyses of ENG, ACVRL1 and SMAD4. Results: In 89% of the probands (n=95), a mutation was identified in one of the three genes. We identified 64 unique mutations, primarily in ENG and ACVRL1, of which 27 (41%) were novel. Large deletions were identified in both ENG and ACVRL1. The prevalence of PAVM was 52.3% in patients with an ENG mutation and 12.9% in the ACVRL1 mutation carriers. We diagnosed 80% of the patients clinically, fulfilling the Curaçao criteria, and those remaining were diagnosed by genetic testing. Conclusions: Proper genetic testing requires analysis for both mutations and large rearrangements of all three genes. Inclusion of a pathogenic mutation as a diagnostic criterion is discussed.

3106F

Exome sequencing of subjects with Congenital Insensitivity to Pain and their family members to reveal novel pain genes. *B. Zhang[1], W. He[1], J. Stephens[1], B. Sidders[2], S. Scollen[2], S. Paciga[1], L. Wood[1], M. Sudworth[3], G. Johnson[3], S. John[1], N. Danziger[4], C. Vangjeli[2].* 1) Pfizer Inc, Eastern Point Road, Groton, CT 06340, USA; 2) Pfizer Neusentis, Granta Park, Great Abington, CB21 6GS, UK; 3) Pfizer Ltd, Ramsgate Rd, Sandwich, CT13 9NJ, UK; 4) Department of Clinical Neurophysiology and Pain Center, Groupe Hospitalier Pitié-Salpêtrière, Paris, France.

Congenital insensitivity to pain (CIP) is a rare heterogenous condition comprised of an absence of sensation to noxious stimuli and can be accompanied by a range of other clinical abnormalities, including anosmia, anhydrosis, and mental retardation. Three genes have consistently been shown to harbour mutations in patients with CIP - SCN9A, NTRK1, and NGF - but some studies that have sequenced these genes failed to identify mutations segregating in families with the disorder. It is therefore likely that there are mutations in other genes responsible for CIP, which may represent novel therapeutic targets for pain. We sequenced the exome of 30 subjects - 15 with CIP and 15 unaffected family members - using the Agilent SureSelect v4 all exon 51Mb kit for exome capture and the Illumina HiSeq 2000 for sequencing. Sequences were aligned using BWA and variants called using GATK. Depending on the mode of inheritance and the levels of relatedness within the families, different mutation filtering strategies were employed.Mutations were filtered using frequency cutoffs in public databases including 1000 genomes, dbSNP and the exome variant server, as well as from 400 in-house exome sequenced samples. Autozygosity mapping was used to identify genomic regions of particular interest. Of the 10 families, 2 of them have potentially functional mutations that segregate in known genes - one family has a novel nonsense mutation in SCN9A and another has a 27bp deletion in the NTRK1 gene. For the other 8 families, we have identified lists of segregating mutations, but of these, the causative mutations remains to be elucidated. Future steps include sequencing coding regions that had poor coverage as well as putative regulatory regions of the 3 known CIP genes, as well as GWAS chip genotyping of all samples to more accurately define runs of homozygosity and/or regions identical by descent. It is likely that the CIP in several of the families will not be the result of known CIP genes, and thus these future steps are likely to reveal novel genes responsible for CIP.

3107W

Whole exome sequencing of a dominant retinitis pigmentosa family with female-specific expressivity identifies a novel deletion in PRPF31. *E.E. Davis[1], A. Villanueva[2], J.R. Willer[1], E.T. Dermitzakis[3], N. Katsanis[1].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Virtual Eye Care MD, Merida, Yucatan, Mexico; 3) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland.

Photoreceptor dysfunction is the most common cause of visual impairment and affects ~1 in 3,000 humans. Underscored by an excess of 180 primary disease gene loci, inherited retinal disorders are hallmarked by vast genetic heterogeneity, thereby posing significant molecular diagnostic challenges. However, the extensive clinical variability observed within and among pedigrees with causal mutations at the same locus offers the unique opportunity to dissect cis and trans modulating effects. Here, we report the genetic dissection of a large Mexican family with non-syndromic retinitis pigmentosa (RP) and apparent female-enriched expressivity and age of onset in the first two decades of life. We ascertained all available members of a large, multigenerational pedigree using fundus photography, dark adaptation studies and electroretinography. Next, we performed whole-exome sequencing on two affected first cousins, an obligate carrier, and a married-in, unrelated spouse. Additional Sanger sequencing of candidate pathogenic variants was performed in the entire pedigree, and we identified a novel 14 bp deletion in exon 9 of PRPF31 (p.Arg289Profs*30), a gene implicated previously in adult-onset dominant RP. The mutation segregated with the phenotype of all ten exclusively female affecteds, but was also present in six family members with no fundus changes, two females and four males, a distribution found to represent a significant female bias in expressivity. Subsequent studies in lymphocyte cells showed that the penetrance/expressivity of the PRPF31 deletion allele was concordant with the expression levels of wild-type message. However, neither the known modulators of PRPF31 expression nor cis-eQTLs within 100 kb of the locus could account for the variance in expression of the message or the clinical phenotype. Although non-penetrance and variable expressivity have been reported previously for this locus, primarily for pedigrees of European descent, these data suggest the existence of a novel modulating mechanism of PRPF31 expression in addition to a hitherto unreported gender bias of expressivity.

3108T

A new palmoplantar keratoderma with severe erythermalgia allelic to Olmsted syndrome. *S. Duchatelet[1,2], S. Pruvost[2,3], S. De Veer[1,2], S. Fraitag[4], P. Nitschké[2,5], C. Bole-Feysot[2,3], C. Bodemer[2,6], A. Hovnanian[1,2,6,7].* 1) INSERM, U781, Paris, France; 2) Université Paris Descartes - Sorbonne Paris Cité, Institut Imagine, Paris, France; 3) Genomics Plateform, IMAGINE Foundation, Paris, France; 4) Department of Pathology, Necker-Enfants Malades hospital, APHP, Paris, France; 5) Bioinformatics Plateform, Paris Descartes University, Paris, France; 6) Department of Dermatology, Necker-Enfants Malades hospital, APHP, Paris, France; 7) Department of Genetics, Necker-Enfants Malades hospital, APHP, Paris, France.

We describe a new severe genetic skin disease in a seven-year-old girl from healthy and non consanguineous French parents. The patient presented with progressive and severe palmoplantar keratoderma (PPK) associated with intense erythermalgia manifesting by acute flares of inflammation, itching, burning pain, vasodilatation and redness of the extremities (hands, feet and ears), triggered by heat. Skin biopsy showed hyperplastic epidermis with hyperkeratosis, parakeratosis, hypergranulosis and papillomatosis, together with dilated capillaries in the upper dermis. Finger and toe nails were thin and brittle. Her hair was fine, dry, curly and unmanageable. Microscopic hair examination shows superficial irregularities without specific abnormalities under polarizing microscopy. Whole exome sequencing identified a *de novo* heterozygous p.Leu673Phe mutation within *TRPV3* (transient receptor potential cation channel, subfamily V, member 3) encoding a nonselective cation channel involved in a variety of processes, including temperature sensation and vasoregulation. Recently, distinct missense *TRPV3* mutations were identified in Olmsted syndrome (OS), a rare keratinizing disorder with periorificial hyperkeratosis and mutilating PPK, all features which were absent in our patient. In addition, the association of severe erythermalgia clearly distinguishes our patient from previously described cases of PPK. The p.Leu673Phe missense mutation involves a highly conserved amino acid residue across species and is predicted to be damaging by *in silico* analysis. This mutation introduces a significant change in size at a critical position of the molecule. Using a TRPV3 homology model, we show that Leu673 is located immediately above the predicted activation gate residue Met677 on the preceding S6 transmembrane helical turn. We suggest that this mutation impairs channel activity and causes the unique and severe skin condition seen in our patient. The clinical features in our patient expand the phenotype of palmoplantar keratoderma associated with *TRPV3* mutations. The molecular mechanisms through which *TRPV3* mutations lead to OS or other palmoplantar keratoderma phenotype are yet to be explored.

3109F

A mutation in A-band titin is associated with hereditary myopathy with early respiratory failure in a Japanese family.
R. Izumi[1,2], T. Niihori[1], Y. Aoki[1], N. Suzuki[2], M. Kato[2], H. Warita[2], T. Takahashi[3], M. Tateyama[2], T. Nagashima[4], R. Funayama[4], K. Abe[5], K. Nakayama[4], M. Aoki[4], Y. Matsubara[1]. 1) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 2) Department of Neurology, Tohoku University School of Medicine, Sendai, Japan; 3) Department of Neurology and Division of Clinical Research, National Hospital Organization Nishitaga National Hospital, Sendai, Japan; 4) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 5) Department of Neurology, Okayama University Medical School, Okayama, Japan.

Hereditary myopathy with early respiratory failure (HMERF, MIM #603689) was originally described as an autosomal dominant disease characterized by adult onset proximal or distal myopathy with early respiratory failure and overlapping pathologic findings with myofibrillar myopathy. HMERF had been considered an extremely rare disease, caused by a rare TTN gene mutation (c.97348C>T of the kinase domain) among North European population. In 1993, our group reported a Japanese family with dominantly inherited cytoplasmic body myopathy, which is now included in myofibrillar myopathy. Currently, this family includes 20 patients in five successive generations who show almost homogeneous clinical features characterized by chronic progressive distal muscle weakness and early respiratory failure. However, the underlying genetic etiology in this family was unknown. In this study, we performed linkage analysis and whole exome sequencing to identify the responsible genetic mutation of the family, clinically and pathologically compatible with HMERF and identified a novel c.90263G>T mutation in the TTN gene. Recently, additional seven TTN mutations have been revealed in 29 families with HMERF by massively parallel sequencing in other groups. All the eight mutation, including ours, were clustered in the fibronection type 3 domain (A150 domain) of the A-band TTN. This finding suggested that the A150 domain plays critical roles in the pathogenesis of HMERF as well as the kinase domain, although detailed mechanisms of pathogenesis remain unknown. HMERF is a more frequent disease than previously expected and identified in different ethnic groups. Further studies will be needed to estimate the frequency of TTN mutations in HMERF and to understand the molecular function of titin.

3110W

Exome Sequencing of Familial Hodgkin's Kindreds.
V. Joseph[1,2], T. Thomas[1], M. Artomov[3], K. Schrader[1,2], A. Kiezun[4], C. Manschreck[1], R. Rau Murthy[1], M. Corines[1], X. Wei[5], N. Gupta[6], L. Margolin[6], A. Zelenetz[7], C. Portlock[7], G. Getz[4], R. Klein[1,2], D. Haber[8], M. Daly[3,9], S. Lipkin[5], D. Altshuler[9,10], K. Offit[1,2]. 1) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Cancer Genome Computational Analysis, Broad Institute, Boston, MA; 5) Weill Cornell Medical College, New York, NY; 6) Genomics Platform, Broad Institute, Boston, MA; 7) Lymphoma Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 8) Massachusetts General Hospital Cancer Center, Boston, MA; 9) Medical & Population Genetics, Broad Institute, Boston, MA; 10) Departments of Genetics and Medicine, Harvard Medical School, Boston, MA.

Hodgkin's disease (HD) is a cancer of the immune system characterized by the presence of Reed-Sternberg cells. Concordance for HD seen in identical twins and increased familial aggregation suggests that there are genetic susceptibility factors involved in its predisposition. Genome-wide association studies have shown that there are multiple loci associated with HD at chromosome 6p21 HLA locus and several other loci near genes such as REL, PVT1 and GATA3. A germline mutation in NPAT in a family was described in a Finnish family.

We have performed exome capture and sequencing of 33 individuals with HD (including 10 affected sib-pair) and where available an unaffected sibling and the unaffected parents (n=18) from multiple kindreds with only Hodgkin's lymphoma. The family structure of majority of these HD sib-pairs reflects a plausible recessive mode of inheritance. Targeted exome capture was achieved using Agilent SureSelect 30MB and sequencing was performed using Illumina HiSeq chemistry. Mapping and variant calling were performed using BWA and GATK either at Broad Institute or at MSKCC. QC was performed using standard criteria and variants annotated using SNPEff. Variants were filtered based on public control data such as ESP6500, 1000genomes, prediction algorithms such as SIFT and Polyphen2 and on conservation scores such as GERP and Phylop. Analyses were performed in both recessive and dominant mode where applicable. Recurrence of variant analyses was also performed

We did not observe truncating NPAT mutations in any family. We also did not observe any segregating high quality, high impact mutations in REL, PVT1 or GATA3. No such mutations were observed in the HLA gene cluster. Ongoing efforts are geared to the discovery of segregating rare variants in the germline (including genes reportedly mutated in somatic tissue) in this type of lymphoma. To understand the population risk and the prevalence, the frequency of these genic-variants need to be estimated in a larger cohort of familial lymphomas. Future plans include a targeted capture of prioritized genes in 1000 sporadic HD cases ascertained at MSKCC.

Acknowledgement: Geoffrey Beene cancer research fund, the STARR Cancer Consortium and the Lymphoma Foundation.

3111T

***ITPR2* loss-of-function mutation causes familial generalized anhidrosis and hyperthermia.**
J. Klar[1], C. Hisatsune[2], S.M. Baig[3], M. Tariq[3], A.C.V. Johansson[1], M. Rasool[4], N.A. Naveed[3], K. Sugiura[2], L. Feuk[1], K. Mikoshiba[2], N. Dahl[1]. 1) Department of Immunology, Genetics and Pathology, Science for Life Laboratory. Uppsala University, Sweden; 2) The Laboratory for Developmental Neurobiology, RIKEN Brain Science Institute, Saitama 351-0198 Japan; 3) The Human Molecular Genetics Laboratory, National Institute for Biotechnology and Genetic Engineering (NIBGE), 38000 Faisalabad, Pakistan; 4) The Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

Anhidrosis, defined as the absence of perspiration in the presence of an appropriate stimulus such as heat or exercise, is a rare condition that may be acquired or congenital. The reasons for anhidrosis, or reduced sweating (hypohidrosis), are heterogeneous and may be caused by defects of sweat gland innervation in disorders of the autonomous nervous system or by a reduced number of functional sweat glands in different ectodermal syndromes. Reports on Familial generalized anhidrosis with normal sweat glands (GANSG) (MIM 106190) are very few.

We have identified a consanguineous Pakistani kindred segregating autosomal recessive GANSG in five affected family members. Using gene mapping and targeted re-sequencing we identified a novel homozygous missense mutation (c.7492G>A; p.G2498S) in the *ITPR2* gene associated with anhidrosis. The *ITPR2* gene encodes for the Inositol 1,4,5-trisphosphate receptor type 2 (InsP3R2), a member of the InsP3R protein family of intracellular Ca^{2+} channels. Functional studies showed that the mutation affects the selective filter for Ca^{2+}-ions with a resulting loss of function of InsP3R2. Furthermore, the *Itpr2*$^{-/-}$ mice show a markedly reduced sweat gland response upon pilocarpine stimulation when compared to wild-type littermates. Taken together, our findings indicate that the *ITPR2* mutation p.G2498S underlies anhidrosis in the patients investigated. Furthermore, we have identified InsP3R2 mediated Ca^{2+} release as a critical mechanism for eccrine sweat gland function, perspiration and thermal cooling.

3112F

A comprehensive disease-mutation search of mitochondrial respiratory chain disorder. *M. KOHDA[1], Y. Tokuzawa[2], Y. Moriyama[2], H. Kato[3], Y. Kishita[2], N. Uehara[2], S. Tamaru[4], Y. Yamashita-Sugahara[2], Y. Nakachi[1], N. Matoba[1], T. Yamazaki[5], M. Mori[6], K. Murayama[7], Y. Mizuno[2], A. Ohtake[5], Y. Okazaki[1,2].* 1) Div Translational Res, Research Center for Genomic Medicine, Saitama Med Univ, Hidaka, Saitama, Japan; 2) Div of Functional Genomics & Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Saitama, Japan; 3) Div. of Developmental Biology, Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Saitama, Japan; 4) Dept. of Obstetrics and Gynecology, Saitama Medical University Hospital, Moroyama, Saitama, Japan; 5) Dept. of Pediatrics, Saitama Medical University, Moroyama, Saitama, Japan; 6) Dept. of Pediatrics, Jichi Medical University, Tochigi, Japan; 7) Dept. of Metabolism, Chiba Children's Hospital, Chiba, Japan.

Mitochondria are small organelles and serve as the powerhouse of the living cells because they generate vital energy in their respiratory chain system consisting of protein complexes I to V. Mitochondrial respiratory chain disorder (MRCD) is an intractable disease that develops in childhood. It is a highly frequent inborn error of metabolism that occurs in at least one out of every 7,000 births. Prominent symptoms develop in such organs as the brain, heart, liver and muscles, where a great deal of energy is required. In most cases, effective treatment has not been established yet. It is known that various gene abnormalities cause defects of the protein complexes of the respiratory chain, which results in mitochondrial dysfunction. However, identification of the causative gene and the understanding of pathogenic mechanism of MRCD remain largely unsolved. In this study, We applied SNP array and exome sequencing in combination with stepwise filtering of gene variants. Exome sequencing data were filtered using three different criteria: (i) the presence of mutations in known disease causing genes; (ii) mutations in genes that code for mitochondrial proteins; (iii) unbiased genome-wide approach with strict filtering strategy. For this study, 103 unrelated individuals were chosen who display juvenile-onset mitochondrial disorders. In 18/103 cases, we identified mutations in known disease-causing genes (AARS2, ACAD9, BOLA3, COX10, GFM1, MPV17, NDUFA1, NDUFA10, NDUFAF6, PC, RARS2, SUCLA2, SURF1 and TUFM). Currently we prioritize rare variants in mitochondria-related genes. In addition, unbiased genome-wide analysis allows the identification of new disease genes thereby extending our understanding of the underlying pathomechanisms. The progress of our precisely controlled strategy will be presented at the meeting.

3113W

Mutation in ATP6AP2, an essential accessary subunit of vacuolar ATPase, causes X-linked Parkinson Disease with Spasticity (XPDS). *E. Korvatska[1], T. Strovas[2], D.-H. Chen[3], J.B. Leverenz[2,3,4,5], K. Kiianitsa[6], I.F. Mata[2,3], D.A. Nickerson[7], C.P. Zabetian[2,3,5], B.C. Kraemer[2,9], T.D. Bird[2,3,8], W.H. Raskind[1,4,8].* 1) Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 2) Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle; 3) Department of Neurology, University of Washington, Seattle; 4) VISN-20 Mental Illness Research, Education, and Clinical Center, Department of Veteran Affairs, Seattle; 5) Parkinson's Disease Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle; 6) 6Department of Immunology, University of Washington, Seattle; 7) Department of Genome Sciences, University of Washington, Seattle; 8) Department of Medicine (Medical Genetics), University of Washington, Seattle; 9) Department of Medicine (Gerontology Division), University of Washington, Seattle.

We report a novel gene for a parkinsonian disorder. X-linked Parkinson Disease with Spasticity (XPDS) presents either as typical adult onset Parkinson's disease or earlier onset spasticity followed by parkinsonism. We previously mapped the XPDS gene to a 28 Mb region on Xp11.2-X13.3. Exome sequencing of one affected individual identified five rare variants in this region, of which none was missense, nonsense or frame shift. Using patient-derived cells we tested the effect of these variants on expression/splicing of the relevant genes. A synonymous variant in ATP6P2, c.345C>T (p.S115S), markedly increased exon 4 skipping resulting in overexpression of a minor splice isoform that produces a protein with internal deletion of 32 aa in up to 50% of the total pool, with concomitant reduction of isoforms containing exon 4. ATP6P2 is an essential accessory component of the vacuolar ATPase required for lysosomal degradative functions and autophagy, a pathway frequently affected in Parkinson's disease. Reduction of the full-size ATP6AP2 transcript in XPDS cells and decreased level of ATP6AP2 protein in XPDS brain may compromise V-ATPase function, as seen with siRNA knockdown in HEK293 cells, and may ultimately be responsible for the pathology.

3114T

Identification of genetic defects in cone and cone-rod dystrophy by whole exome sequencing. *C. Lazar[1,2], L. Zelinger[3], M. Mutsuddi[1,4], L. Mizrahi-Meissonnier[3], A. Boleda[1], R.R. Priya[1], E. Banin[3], D. Sharon[3], A. Swaroop[1].* 1) Neurobiology-Neurodegeneration & Repair Laboratory (N-NRL), National Eye Institute, National Institutes of Health, Bethesda, MD, 20892; 2) Institute for Doctoral Studies, Babes-Bolyai University, Mihail Kogalniceanu 1, Cluj-Napoca, Romania, 400084; 3) Department of Ophthalmology, Hadassah - Hebrew University Medical Center, Jerusalem, Israel 81120; 4) Department of Molecular and Human Genetics, Varanasi, India 221005.

Cone dystrophy (CD) and cone-rod dystrophy (CRD) are clinically and genetically heterogeneous retinal disorders displaying autosomal dominant (AD), autosomal recessive (AR) and X-linked inheritance patterns. Several genetic loci have been implicated, and causative mutations have been identified in over twenty genes. The goal of this study is to identify the cause of disease in AD and AR families with cone-dominated retinal phenotypes using whole exome sequencing. Finding causal variants in retinal dystrophies is difficult using traditional methods because of the high degree of genetic heterogeneity. Whole exome sequencing (WES) has been successfully applied for molecular characterization of cohorts affected with a heterogeneous condition as well as identification of novel variants. In this study, WES was performed on 13 individuals from four unrelated Israeli families. Three of the families were clinically diagnosed with CRD and one with CD. To capture the target regions from genomic libraries, we used the Agilent Sure Select Human All Exon V4 capture kit. Sequencing was carried out on an Illumina Genome Analyzer IIx. The sequencing reads that passed the initial quality control achieved an average coverage of 60X. Initial data analysis was performed using Genomatix software. Potential candidate variants are being filtered against an existing Israeli cohort dataset. Additional analysis using other publicly available analysis tools and databases are being used for further filtering and to identify rare variants. We are evaluating the expression of the putative candidates in fetal and adult retina in our global expression databases and will be validating significant causal variants using Sanger sequencing. Initial analysis of three families has led to the identification of rare variants in CDHR1, C8orf37 and GUCY2D genes. CDHR1 and C8orf37 were previously reported to cause AR form of CRD and GUCY2D was previously implicated in the cause of AD CRD. In CDHR1 we have identified a new non-sense mutation predicted to result in a truncated protein with loss of several important domains. Further experiments will be carried out to understand the biological effect of the novel variants in causing the disease pathology. Our studies should help us in better understanding the molecular basis of cone and cone-rod dystrophy, thereby improving clinical management and development of new therapeutic strategies.

3115F

A novel germline *PIGA* mutation in Ferro-Cerebro-Cutaneous Syndrome: A neurodegenerative X-linked encephalopathy with epilepsy and systemic iron-overload.
R.L. Margraf[1], E.M. Coonrod[1], J. Durtschi[1], K. Mallempati[1], A. Kumanovics[1,2], T.M. Newcomb[3], J.M. Opitz[2,4,5,6], J.C. Carey[5,6], H. Zhou[2], B.E. Katz[7], K.V. Voelkerding[1,2], K.J. Swoboda[3,6]. 1) ARUP Institute for Clinical and Experimental Pathology®, ARUP Laboratories, Salt Lake City, UT; 2) Department of Pathology ,University of Utah, School of Medicine, Salt Lake City, Utah; 3) Department of Neurology ,University of Utah, School of Medicine, Salt Lake City, Utah; 4) Department of Human Genetics ,University of Utah, School of Medicine, Salt Lake City, Utah; 5) Department of Obstetrics and Gynecology ,University of Utah, School of Medicine, Salt Lake City, Utah; 6) Department of Pediatrics (Division of Medical Genetics) University of Utah, School of Medicine, Salt Lake City, Utah; 7) Private Practice of Pediatrics and Pediatric Neurology (retired), Twin Falls, Idaho, 83303, USA.

This report presents a family with three males affected by Ferro-Cerebro-Cutaneous Syndrome, a novel X-linked syndrome of neurodegeneration, cutaneous abnormalities, and systemic iron overload. Linkage studies demonstrated a shared haplotype at Xp21.3-Xp22.2, a region spanning 18 megabases and containing more than 100 candidate genes. When candidate gene sequencing proved unsuccessful, exome sequencing was used to determine the causal variant. A heuristic filtering approach identified several rare mutations in the linkage region, but only the PIGA mutation segregated with disease in the family. The *PIGA* gene contains a germline three base pair in frame deletion (deleted positions chrX:15342943-15342945, p.110delLeu). The leucine residue that is deleted in the affected males is highly conserved from humans to zebrafish. The unaffected great-grandfather had the same X allele as all the affected males, but without the *PIGA* mutation. This indicates that the *PIGA* mutation arose *de novo* on the X allele from the great-grandfather to the grandmother who then passed the *PIGA* mutation on to several children, including an affected male. *PIGA* encodes an enzyme in the GPI anchor biosynthesis pathway. Clonal expansion of cells with somatic *PIGA* mutations causes paroxysmal nocturnal hemoglobinuria (PNH). In PNH, blood cells including erythrocytes are deficient in GPI anchor proteins, reducing expression of critical cell surface proteins and predisposing cells to complement-mediated lysis. Recently, a family with a germline *PIGA* mutation was reported where affected males had multiple congenital anomalies and severe neurologic impairment resulting in infantile lethality. In contrast, affected boys in this report were born without anomalies and were apparently neurologically normal prior to onset of seizures after 6 months of age, with 2 surviving to the second decade. An affected individual was tested for the presence of GPI anchor proteins in granulocytes, monocytes, and erythrocytes; and only granulocytes were found to be deficient in GPI anchored proteins. This could explain the milder phenotype seen in our patients compared to *PIGA* mutations that lead to erythrocyte lysis or early lethality. The novel *PIGA* mutation in this family likely caused a reduction in GPI-anchor protein cell surface expression in various cell types, resulting in the observed novel and complex phenotype involving the central nervous system, skin and iron metabolism.

3116W

Deciphering the genetic background of PEHO-like syndrome.
M. Muona[1,2,3], A. Laari[3,4,5], A.-K. Anttonen[3,4,5,6], M. Somer[7], A. Palotie[1,8,9], A.-E. Lehesjoki[3,4,5]. 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland; 3) Folkhälsan Institute of Genetics, University of Helsinki, Helsinki, Finland; 4) Haartman Institute, Department of Medical Genetics and Research Program's Unit, Molecular Medicine, University of Helsinki, Helsinki, Finland; 5) Neuroscience Center, University of Helsinki, Helsinki, Finland; 6) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 7) Norio Centre, Rinnekoti Foundation, Helsinki, Finland; 8) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK; 9) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

PEHO syndrome (progressive encephalopathy with edema, hypsarrhythmia, and optic atrophy) is an infantile onset neurodegenerative disorder. The clinical presentation in Finnish patients, homozygous for a founder mutation (unpublished), is relatively uniform and includes hypotonia, infantile spasms with hypsarrhythmia, profound psychomotor retardation, optic atrophy, and progressive brain atrophy originating in cerebellum. A significant number of patients manifest many of these features in the absence of the typical neuroradiological findings, or with no sign of progression. These patients remain without proper diagnosis but are often classified as PEHO-like. We aim to decipher the genetic basis of PEHO-like syndrome by using exome sequencing. We have selected 29 Finnish, mostly sporadic, PEHO-like patients, who have been excluded for the PEHO founder mutation. Exome sequencing was performed using Illumina HiSeq 2000 platform. We carried out variant filtering using strategies assuming two different patterns of inheritance: recessive and de novo. In the former we selected potentially deleterious variants - either homozygous or compound heterozygous - of essential splice site, nonsense, frameshift, and missense types with a 1000 genomes minor allele frequency below 1%. In the '*de novo*' analysis we included only heterozygous, potentially damaging variants absent from the control databases. Analysis of the first 21 exomes revealed likely pathogenic hemizygous mutations in *CDKL5*, a known infantile epileptic encephalopathy gene, in two male patients. Capillary sequencing of the patients' parents showed that the mutations occurred *de novo*. In an affected sibpair we identified likely pathogenic compound heterozygous mutations in *ABAT*, previously linked to epileptic encephalopathies in a few cases. One patient had a three-amino-acid duplication previously reported as pathogenic in *SPTAN1*. Finally, we have also identified potentially pathogenic mutations in genes without a previous connection to encephalopathies. We are currently analysing the exomes of the remaining eight patients. To facilitate identification of heterozygous de novo variants we have also sequenced the exomes of parents of seven patients. Our findings imply that PEHO-like syndrome is genetically highly heterogeneous. A subset of patients had mutations in previously established disease genes, indicating the utility of exome sequencing as a diagnostic tool.

3117T

Aicardi Goutieres Syndrome (AGS) - phenotypic variability and diagnosis in a series of cases using whole-exome sequencing. *V. Narayanan[1,2], S. Szelinger[1,3,4], J.J Corneveaux[1,3], I. Schrauwen[1,3,5], A.L. Siniard[1,3], A.A. Kurdoglu[1,3], I. Malenica[1,3], K.M. Ramsey[1,3], D.W. Craig[1,3], M.J. Huentelman[1,3].* 1) Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix, AZ; 2) Pediatric Neurogenetics Center, Barrow Neurological Institute, Phoenix AZ; 3) Neurogenomics Division, Translational Genomics Research Institute, Phoenix AZ; 4) Molecular and Cellular Biology Interdisciplinary Graduate Program, School of Life Sciences, Arizona State University, Tempe AZ; 5) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium.

AGS is an autosomal recessive heterogeneous disorder, with causal mutations reported in 6 genes (TREX1, RNAseH2A, RNAseH2B, RNAseH2C, SAMHD1, ADAR) to date. Markers of AGS include elevated CSF interferon-alpha, elevated CSF neopterin and tetrahydrobiopterin, and basal ganglia or cerebellar calcification, are not universal in all cases. Due to a wide range of clinical symptoms that do not completely follow the known genetic risk models many suspected AGS cases remain undiagnosed and possible new causal genes remain unidentified. Here we describe 3 families with affected children in whom AGS was suspected and causal variations identified by family based whole-exome sequencing. Each study participant was pooled into multiplexed libraries of 6 total samples by TruSeq Exome Enrichment v2 chemistry and sequenced on two lanes of a HiSeq2000 flowcell to a mean target depth of 97±70X. In Family 1, with a single affected male and parental consanguinity, initial clinical diagnosis of leukodystrophy was made but biochemical and genetic testing was negative. Exome sequencing uncovered a single homozygous variant in exon 3 of RNASEH2B that is highly conserved, deleterious and never before seen in any public SNP databases. Both parents were identified as carriers, fitting autosomal recessive inheritance, which was validated by Sanger sequencing. In Family 2, a single affected female with normal development until 13 months of age but progressive encephalopathy after immunization was observed. Exome sequencing uncovered compound heterozygous variants in Exons 2 and 9 of (ADAR1). Both variants were highly conserved and deleterious as defined by in-silico prediction algorithms and one of the two causal variants is a previously described risk allele for AGS cases. In Family 3, with two affected boys, the initial clinical diagnosis was acute disseminated encephalomyelitis (ADEM), but questioned after the younger sibling developed a subacute encephalopathy. Exome sequencing uncovered deleterious compound heterozygous variants in Exon 2 and 8 of RNASEH2A. One of the variants was never before seen, while the other had no reported frequency in target population. In conclusion AGS may be more prevalent than previously thought and should be considered in cases of acute encephalopathy with features of white matter disease, following infection or immunization. Whole exome sequencing may be an efficient approach to diagnosis, and discovery of new genes linked to AGS.

3118F

Exome sequencing identifies mutations in a novel gene in patients with Noonan syndrome. *T. Niihori[1], Y. Aoki[1], T. Banjo[2], N. Okamoto[3], S. Mizuno[4], K. Kurosawa[5], T. Ogata[6], F. Takada[7], M. Yano[8], T. Ando[9], T. Hoshika[10], C. Barnett[11,12], H. Ohashi[13], H. Kawame[14], T. Hasegawa[15], T. Okutani[16], T. Nagashima[17], S. Hasegawa[18], R. Funayama[19], T. Nagashima[19], K. Nakayama[19], S. Inoue[1], Y. Watanabe[2], T. Ogura[2], Y. Matsubara[1].* 1) Department of Medical Genetics, Tohoku University School of Med, Sendai, Miyagi, Japan; 2) Department of Developmental Neurobiology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan; 3) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Japan; 4) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Kasugai, Japan; 5) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 6) Department of Pediatrics, Hamamatsu University School of Medicine, Hamamatsu, Japan; 7) Department of Medical Genetics, Kitasato University Graduate School of Medical Sciences, Sagamihara, Japan; 8) Department of Pediatrics, Akita University School of Medicine, Akita, Japan; 9) Department of Pediatrics, Municipal Tsuruga Hospital, Tsuruga, Japan; 10) Department of Pediatrics, Tottori Prefectural Central Hospital, Tottori, Japan; 11) South Australian Clinical Genetics Service, SA Pathology, Women's and Children's Hospital, North Adelaide, Australia; 12) School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, South Australia, Australia; 13) Division of Medical Genetics, Saitama Children's Medical Center, Saitama, Japan; 14) Department of Genetic Counseling, Ochanomizu University, Tokyo, Japan; 15) Department of Pediatrics Keio University School of Medicine, Tokyo Japan; 16) Division of NICU, General Perinatal Medical Center, Wakayama Medical University, Wakayama, Japan; 17) Department of Pediatrics, Jikei University School of Medicine, Tokyo, Japan; 18) Department of Pediatrics, Niigata Graduate School of Medical and Dental Sciences, Niigata, Japan; 19) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan.

RAS GTPases mediate a wide variety of cellular functions, including proliferation, cell survival and differentiation. Recent studies have revealed that germline mutations and genetic mosaicism of classical RAS, including HRAS, KRAS and NRAS, show a wide spectrum of genetic disorders: Noonan syndrome and related disorders (RAS/mitogen-activated protein kinase (MAPK) pathway syndromes or RASopathies), nevus sebaceous and Schimmelpenning syndrome. In the present study, we identified a total of nine missense, nonsynonymous mutations in a gene encoding a member of the RAS family, in 17 of 180 individuals (9%) with Noonan syndrome and related conditions without mutations in known genes. Clinical manifestations in the mutation-positive individuals are consistent with those of Noonan syndrome, which are characterized by distinctive facial appearance, short stature and congenital heart defects. Seventy percent of mutation-positive individuals had hypertrophic cardiomyopathy, a high frequency compared with the 20% incidence in individuals with Noonan syndrome overall. Luciferase assays in NIH3T3 cells showed that five mutants identified in children with Noonan syndrome enhanced ELK1 transactivation. The introduction of mutant mRNAs of this gene into one cell-stage zebrafish embryos was found to result in a significant increase of embryos with craniofacial abnormalities, incomplete looping and a hypoplastic chamber in the heart and an elongated yolk sac. These results demonstrated that gain-of-function mutations in this gene cause Noonan syndrome, showing a similar biological effect to mutations in other RASopathy genes. Note: The name of the gene will be disclosed at the meeting.

3119W

Identification of pathogenic variants in Idiopathic Scoliosis. *S. Patten[1,2], E. Alix[3,4], A. Labalme[3,4], S. Girard[5], B. Biot[6], C. Poizat[3,4], D. Sanlaville[3,4], J. Berard[7], G. Rouleau[5], F. Clerget-Darpoux[8], P. Drapeau[1], F. Moldovan[2], P. Edery[3,4].* 1) Pathology and Cell Biology, Université de Montreal, Montreal, Quebec, Canada; 2) CHU Sainte Justine and Faculté de Médecine Dentaire, Université de Montréal, Québec, Canada; 3) Hospices Civils de Lyon, Department of Genetics, Lyon, France; 4) Inserm U1028, CNRS UMR5292, University Lyon 1, Neuroscience Research Centre, Lyon, France; 5) Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada; 6) Croix-Rouge française, CMCR des Massues, Lyon, France; 7) Hospices Civils de Lyon, Service d'Orthopédie Pédiatrique, Lyon, France; 8) Département de Génétique, Unité INSERM U-781, IHU Imagine et Université Paris Descartes Hôpital Necker-Enfants Malades, Paris, France.

Idiopathic scoliosis (IS) is a spine deformity affecting up to 3% of adolescents. Despite strong evidences of genetic contributions to the etiology of IS, the causative genes remain unidentified. We previously identified a new disease gene location (3q12.1 or 5q13.3) in a large extended family with apparent monogenic inheritance (Edery P et al, Eur J Hum Genet 2011). In the present study, we performed SNP genotyping to further refine these loci. Whole exome sequencing was performed in 7 affected individuals from the large family and 4,564 single nucleotide variants (SNVs) were identified in our targeted chromosomal regions. These SNVs were subsequently filtered and the pathogenic nature of candidate SNVs was assessed by functional studies in zebrafish. We identified one rare missense SNV (MAF<1%) in a gene, within the 5q13.3 IS interval, that segregated with IS in the family and resulted in spinal deformities similar to that observed in IS patients when overexpressed in zebrafish. Sanger sequencing of this gene was subsequently performed in 40 additional multiplex families, and in 150 unrelated IS cases. Co-transmission of that same SNV and the disease was also observed in 3/40 additional families and it was identified in 3/150 unrelated IS cases. We also identified a missense mutation in this gene in 1/40 family, and another rare missense variant in the gene in 5/150 IS patients. None of these three nucleotide variants were found in 206 control chromosomes of individuals in the same geographical area. Functional analyses of these mutations in zebrafish also resulted in spine deformities similar to that observed in IS patients. These results indicate that these variants cause IS and, to our knowledge, we report on the first ever causative gene in idiopathic scoliosis.

3120T

Exome sequencing is the preferred approach for identifying the genetic cause in consanguinous and non-consanguinous recessive disease. *H. Smeets[1], J. Vanoevelen[1], M. Gerards[1], R. Kamps[1], T. Theunissen[1], B. De Koning[1], I. Boesten[1], M. van Geel[1], P. Lindsey[1], C. Stumpel[1], M. Nguyen[1], M. Gerrits[1], S. Ghesquiere[1], S. Stevens[1], C. de Die[1], B. van den Bosch[1], I. De Coo[2].* 1) Dept Clin Genet, Maastricht UMC, Maastricht, Netherlands; 2) Dept Neurology, Erasmus UMC, Rotterdam, Netherlands.

Exome sequencing is becoming the primary approach in genetic testing, particularly suited to solve *de novo* cases and patients with recessive disease. We performed exome sequencing in 27 families with predominantly recessive neurological syndromes and, if available, a metabolic cellular phenotype. In most families more than one child was affected. Counseling information and SNP-array data was used to classify the parents as consanguinous or non-consanguinous. Exome sequencing was done on a single patient per family, using an Illumina HiSeq2000 system. Quality criteria were >90 million reads per sample, >90% coverage per base at 10X, more than 65% coverage per base at 20X, <50% duplicate reads, >80X mean corrected coverage and a specificity of >65%. Data analysis was performed using an in-house optimized pipeline. Filtering of the variants was based on frequency (<1%), functional impact, function of the gene, the genetic model (homozygous or compound heterozygous), genetic localization and segregation in the family. If the number of remaining variants was too high or the the data did not match the quality criteria (<10%) an additional patient was sequenced. A total of seven patients were homozygous or compound heterozygous for mutations in the genes LPIN1, SCL19A3 (2 times), AARS2, DHODH, SERAC1, MMP14, five of which could be directly linked to the phenotype and two were confirmed after *in vitro* validation. In 16 patients possibly pathogenic variants were identified in genes, which could be connected to the phenotype. If possible, these variants are functionally validated or confirmed in additional patients, but providing the definite evidence can be cumbersome. Still, we expect to solve the majority of the families studied. In the consanguinous families, the phenotypes are often not caused by a single gene defect and more genes are mutated, each explaining different parts of the clinical spectrum. As the parents involved have a risks of offspring with different pathologies, it is advisable in those families to screen both parents completely for shared recessive mutations to determine recurrence risks and prenatal options.

3121F

Next-Generation Sequencing Identities PXDN Mutations in Patients with Complex Microphthalmia and Anterior Segment Eye Disease. *A. Slavotinek[1], A. Choi[1], R. Lao[2], P. Ling-Fung Tang[2], W. Mayer[1], E. Wan[2], T. Bardakjian[3], P.Y. Kwok[2], A. Schneider[3].* 1) Dept Pediatrics, Division of Genetics, Univ California, San Francisco, San Francisco, CA; 2) Cardiovascular Research Institute, UCSF, San Francisco, CA; 3) Division of Medical Genetics, Einstein Medical Center, Philadelphia, PA.

Anophthalmia and microphthalmia (A/M) are common and significant because of visual loss. Mutations in SOX2 and other genes can cause A/M, but more than half of affected individuals do not receive a molecular diagnosis. A/M can be isolated ('simplex A/M') or accompanied by additional ocular defects ('complex A/M'). We used exome sequencing to study children with simplex A/M (n=14), complex A/M (n= 5) and other eye defects including anterior segment dysgenesis (ASD) and retinal dysplasia (n=3). Samples were initially analyzed for known genes associated with A/M and if negative, further analysis was undertaken. We present two families who had PXDN mutations as a cause of their eye defects. The first family had an affected sib pair with complex A/M and ASD, sclerocornea, microphthalmia, developmental delays and hypotonia. The older male sibling had glaucoma. Both sibs had two PXDN mutations: c.1021C>T, predicting p.Arg341* and a frameshift mutation, c.2375_2397del, predicting p.Tyr791Serfs*65 and premature protein truncation. The second family was a single affected male with bilateral ASD who had healthy parents. There was no known consanguinity. This child had a paternally inherited, missense mutation, p.Gln316Pro, and a maternally inherited frameshift mutation, p.Tyr398Thrfs*40. Mutations in PXDN have been described in 3 families with congenital cataracts, microcornea, sclerocornea and glaucoma, but extraocular anomalies were not noted (Khan et al., 2011). Peroxidasin is usually located in the endoplasmic reticulum and can be secreted into the extracellular space. The protein contains a peroxidase domain and motifs that are typical for extracellular proteins, including an amino-terminal secretory signal sequence and ligand binding domains, with leucine-rich repeat regions, C2-type immunoglobulin-like motifs and a von Willebrand factor type C domain (Khan et al., 2011). The peroxidasins are involved in hydrogen peroxidase metabolism, although the substrates for this reaction in the eye have not been characterized. It is unclear if PXDN provides adhesive support for the corneal and lens epithelium, or if the reduced antioxidant capacity that results from lack of this enzyme leads to increased reactive oxygen intermediates and insoluble aggregates that result in cataracts and corneal clouding. We conclude that PXDN mutations should be considered in complex microphthalmia and anterior segment dysgenesis.

3122W

Diagnostic exome sequencing to elucidate the genetic basis of likely recessive disorders in consanguineous families. *P. Makrythanasis[1], M. Nelis[1,2], F.A. Santoni[1], M. Guipponi[3], A. Vannier[1], F. Béna[3], S. Gimelli[3], E. Stathaki[3], S. Temtamy[4], A. Megarbané[5,6], A. Masri[7], M.S. Aglan[4], M.S. Zaki[4], A. Bottani[3], S. Fokstuen[3], S. Kitsiou-Tzeli[8], H. Fryssira[8], E. Kanavakis[8], N. Al-Allawi[9], A. Sefiani[10,11], S. Al-Hait[12], S.C. Elalaoui[10], N. Jalkh[5], L. Al-Gazali[13,14], F. Al-Jasmi[13,14], H. Chaabouni Bouhamed[15], E. Abdalla[16], D.N. Cooper[17], H. Hamamy[1], S.E. Antonarakis[1,3].* 1) Dept Genetic Medicine & Dev, University of Geneva, Geneva, Switzerland; 2) Estonian Genome Centre, University of Tartu, Tartu, Estonia; 3) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 4) Department of Clinical Genetics, National Research Centre, Cairo, Egypt; 5) Medical Genetics Unit, Saint Joseph University, Beirut, Lebanon; 6) Institut Jerôme Lejeune, Paris, France; 7) Pediatric Department, The University of Jordan, Amman, Jordan; 8) Department of Medical Genetics, University of Athens, Athens, Greece; 9) Department of Pathology, College of Medicine, University of Dohuk, Dohuk, Iraq; 10) Département de Génétique Médicale, Institut National d'Hygiène, Rabat, Morocco; 11) Centre de Génomique Humaine, Faculté de Médecine et de Pharmacie, Université Mohamed V Souissi, Rabat, Morocco; 12) Genetics & IVF Department, The Farah Hospital, Amman, Jordan; 13) Department of Paediatrics, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates; 14) Department of Pediatrics, Tawam Hospital, United Arab Emirates University, Al-Ain, United Arab Emirates; 15) Department of Human Genetics, University Tunis El Manar, Faculty of Medicine, Tunis, Tunisia; 16) Medical Research Institute, Alexandria University, Alexandria, Egypt; 17) Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK.

Consanguinity is a risk factor for autosomal recessive (AR) disorders. Although many of the clinical phenotypes presenting in the offspring of consanguineous couples are of unknown etiology, advances in sequencing the protein-coding portion of the human genome (exome) provide an opportunity to arrive at a molecular diagnosis in at least a proportion of unresolved phenotypes, thereby identifying novel candidate genes responsible for AR phenotypes. Samples were collected from 41 consanguineous families characterized by a wide spectrum of clinical phenotypes suggestive of AR inheritance. DNA was taken from the patient(s), all unaffected siblings and the parents. All samples were genotyped with a 720K SNP array to define the identical-by-descent chromosomal regions likely to contain the responsible pathogenic variants. Exome sequencing was performed on one affected individual per family. Variants within the identified target areas were called and filtered using bioinformatic tools. The putative pathogenic variant was found in known disease-causing genes (VLDLR, DMP1, FKTN, SEPSECS, GUCY2D, BBS4, SYNE1 and POMGNT1, TACO1) in 9 families. In 4 families evidence for a simultaneous presence of two monogenic disorders was identified. In 21 families, variants of likely pathogenicity were found. In 7 families, no plausible candidates were identified. Consanguineous families provide a unique opportunity to identify pathogenic variants in both known and candidate genes responsible for recessive. We identified a putative causative variant in 9/41 families tested and in further 4/41 evidence for affection from two monogenic disorders. High-throughput sequencing represents a substantial improvement in our ability to diagnose recessively inherited disorders and identify novel candidate genes.

3123T

Exome sequencing identifies loss of function mutations in *UBE3B* in a family with intellectual and developmental delay, hypotonia, distal digital hypoplasia, and hearing loss. *Y.R. Li[1,2], S.A. Shalev[5], J. Liang[3], Y. Guo[2], J. Zhang[3], B.J. Keating[2,4,10], Y. Chen[3], L. Tian[2], S.S. Vergano[6], X. Xu[3], H. Hakonarson[2,4,10], J. Wang[3,7,8,9], M.A. Deardorff[4,11].* 1) Medical Scientist Training Program; Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104; 3) BGI-Shenzhen, Shenzhen 518083, China; 4) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 5) Genetic Institute, Emek Medical Center, Afula, Israel; 6) Division of Medical Genetics and Metabolism, Children's Hospital of the King's Daughters, Norfolk, Virginia; 7) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 8) King Abdulaziz University, Jeddah, Saudi Arabia; 9) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; 10) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 11) Department of Pediatrics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104.

Recessive, loss of function, mutations in the *UBE3B* gene have recently been reported in the Blepharomphimosis-Ptosis-Intellectual Disability (BPIDS) and Kaufman oculocerebrofacial syndromes (KOS). Here we present an additional family with clinical features caused by recessive mutations in *UBE3B*, identified using of autozygosity mapping and exome sequencing. The family presented with known consanguinity, including multiple affected siblings as well as an affected paternal cousin with severe growth and global intellectual developmental delays, hypotonia, severe sensorineural hearing loss, genitourinary (GU) anomalies, bilateral aplasia of the distal phalanges, micrognathia, microcephaly, and hypoplasia of the distal phalanges of one or more toes. The proband and brother both died of unknown causes at ages 9 and 2, respectively; metabolic testing of all three affected individuals revealed no abnormalities. Given the complex presentation and lack of a clear definitive diagnosis, we performed exome sequencing of the affected proband, two unaffected siblings, and unaffected parents to identify a Mendelian mutation that would be causative for this disorder. Given the pattern of inheritance, we expected to identify a homozygous mutation in the proband that was present in a heterozygous state in the parents, and either absent or heterozygous in the unaffected siblings. We identified a deleterious homozygous (A>G) splice acceptor site mutation in intron 13 of *UBE3B* (c.1451-2A>G) predicted to result in skipping of exon 14. The *UBE3B* gene resides in a 591kb region of shared autozygosity on chromosome 12q24.11, identified by SNP array in the affected siblings and affected paternal cousin. We confirmed the mutation to be homozygous in the three affected children and heterozygous in both the parents and two unaffected siblings. The features in this family, while distinct, demonstrate significant clinical overlap with those identified in BPIDS and KOS. Our findings suggest that there is a broader clinical spectrum of features caused by mutations in *UBE3B*, underscoring the evolving genetic landscape of phenotypically overlapping disorders in the neurocognitive syndromic developmental axis.

3124F

Affected Sib Analysis of Human X-Exome Data to Identify Novel X-linked Intellectual Disability Genes. *T. Niranjan[1,2], C. Skinner[3], T. Turner[1,2], M. May[3], L. Holloway[3], A. Chakravarti[1], C. Schwartz[3], T. Wang[1,4].* 1) The McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins University, Baltimore, MD; 3) Greenwood Genetics Center, Greenwood, SC; 4) Department of Pediatrics, Johns Hopkins University, Baltimore, MD.

X-linked Intellectual Disability (XLID) is a group of genetically highly heterogeneous disorders caused by mutations in genes on the X chromosome. More than 90 XLID genes have been found, which account for ~50% of XLID disorders. To systematically identify responsible genes for the remaining XLID, we sequenced the X-exome of 55 XLID families (a single affected male from 27 families and two affected males from 28 families) using an Agilent SureSelect system and the HiSeq2000 platform. An average of 1,844 ± 246 variants were identified in each X-exome with a mean coverage of 49x. To enrich disease-causing mutations, we initially tried to use variant datasets from dbSNP and the male portion of 1000 Genomes as filters. However, these databases present several challenges for an efficient enrichment of disease-causing mutations. dbSNP contains potentially unannotated pathological mutations and available male data from 1000 Genomes contain an excess of ambiguous variants for the X chromosome. To solve these problems, we tested and optimized a novel strategy by sequencing affected sibs in proband families to enrich for shared variants and by eliminating neutral variants using a small cohort of unrelated males. This approach achieved a substantial (98.1%) reduction in the number of variants for follow-up studies and significantly enriched for known XLID genes, including *ATRX*, *HUWE1*, *MECP2*, *MED12*, *MAOA*, and *PHF8* ($p < 0.005$, hypergeometric test). Novel candidate XLID genes including *ZC4H2* are also identified. We conclude that the affected sib-based approach can be used to effectively enrich for disease-causing mutations in genetically heterogeneous X-linked disorders and that publicly available human reference databases including dbSNP and 1000 Genomes should be used with caution as automatic filters to enrich for disease-causing mutations in X-linked disorders.

3125W

Novel C10orf2 mutations cause Perrault syndrome. *H. Kawakami, H. Morino, R. Miyamoto, H. Maruyama.* Dept Epidemiology, Hiroshima Univ RIRBM, Hiroshima, Japan.

Purpose: Perrault syndrome, first reported in 1951, is characterized by sensorineural hearing loss in both genders and gonadal dysfunction in females, and genetically and clinically heterogeneous autosomal recessive disorder. **HSD17B4**, **HARS2**, **CLPP** and **LARS2** were reported as responsible genes for Perrault syndrome so far. By exome sequencing, we searched further mutations of the syndrome.

Methods: Patients are 38-year-old and 32-year-old sisters. There is no genetic relationship between their parents. The parents and brother have no similar symptoms. We performed exome-sequencing. We used BWA for alignment/mapping, Samtools and Picard for SAM/BAM handling, GATK and Samtools for variant call, and Annovar for annotation. Functional predictions due to amino acid changes were estimated by PolyPhen2 , SIFT and Mutation Taster.

Results: Total reads of two affected patients are 9,003,317,959 and 10,599,079,582 bp, and mean coverages are 84.87 and 99.92. Total variant numbers are 116,627 and 120,840 and 100 common variants which were not found in open variant database, dbSNP135, 1000 genomes and ESP, nor in in-house variant database were filtered. Among these variants, we detected only one gene, C10orf2 which contained multiple variants, c.1172G>A (p.R391H) and c.1754A>G (p.N585S). These variants are predicted to be damaging by all the three algorithms. C10orf2:c.1172G>A was paternal and c.1754A>G was maternal allele. Affected two women had both variant alleles, but unaffected brother had only paternal one.

Conclution: We found novel compound heterozygous mutations in C10orf2, encoding Twinkle which cause Perrault syndrome. C10orf2 was also known as the causative gene of PEOA4 and IOSCA. However, the phenotype of this family is different from these diseases. Twinkle is a DNA helicase which unwinds short stretches of double-stranded DNA and, along with mitochondrial single-stranded DNA biding protein and mtDNA polymerase gamma, is thought to play a key role in mtDNA replication. Since **HARS2**, **CLPP** and **LARS2** which were known as responsible genes for Perrault syndrome are associated with mitochondrial function, the mutations of C10orf2 are reasonable as the cause of the syndrome.

3126T

Novel clinical presentations associated with mutations in nuclear genes of the mitochondrial translation apparatus. *M. RIO[1,2], Z. ASSOULINE[1], A.S. LEBRE[1], N. BODDAERT[2], I. DESGUERRE[2], A. MUNNICH[1,2,3], A. ROTIG[3].* 1) SERVICE DE GENETIQUE, HOPITAL NECKER-ENFANTS MALADES, PARIS, France; 2) DEPARTEMENT DE PEDIATRIE, HOPITAL NECKER-ENFANTS MALADES, PARIS, France; 3) INSTITUT IMAGINE, INSERM U781, HOPITAL NECKER-ENFANTS MALADES, PARIS, France.

The mitochondrial protein synthesis apparatus allows the synthesis of the 13 respiratory chain (RC) subunits encoded by mitochondrial DNA. The deficiency of mitochondrial protein synthesis can be caused by mutations in any component of the translation apparatus including tRNA, rRNA and proteins. We report here 4 patients from 3 independent families with mutations in 3 nuclear genes encoding proteins of the mitochondrial translation apparatus. All patients presented a multiple RC deficiency in muscle, liver or fibroblasts and an abnormal pattern assembly of RC complexes in fibroblasts. Interestingly the clinical presentations of the patients differ for the previously reported patients with mutations in these genes. Patients 1-2 are two brothers born to non consanguineous parents. They presented hypoglycemia, trunk hypotonia and lactic acidosis at 2 days of age. They died at 6 and 5 days of age. Exome sequencing detected two heterozygous mutations of MTO1 encoding the mitochondrial translation optimization 1. MTO1 mutations have been identified in two families with hypertrophic cardiomyopathy and lactic acidosis. Patient 3 is a girl born to consanguineous parents of Turkish origin. She has sideroblastic anemia, psychomotor retardation, microcephaly, exocrine pancreatic dysfunction, short stature, and mild mental retardation. She has a homozygous PUS1 mutation. PUS1 encodes the pseudouridylate synthase 1 that converts uridine to pseudouridine once it has been incorporated into an RNA molecule. PUS1 mutations were previously reported in patients with multiple RC deficiency and myopathy, lactic acidosis, and sideroblastic anemia (MLASA). Patient 4 is a girl born to non consanguineous parents. She presented trunk hypotonia in the first months of life. She developed tonic spasms without hypsarrhythmia on EEG between 6 and 18 months of age. At 2.5 years, she had delayed development, growth retardation, hypotrophia, microcephaly and spastic diplegia. She had permanent hyperlactatemia. Brain MRI detected bilateral anomalies of thalamus. Exome sequencing identified two heterozygous mutations in GFM1 encoding the mitochondrial elongation factor G1. Until now GFM1 mutations were associated with either severe encephalopathy or liver insufficiency. These results give support to the clinical and genetic heterogeneity of mitochondrial disorders and highlight how difficult genotype-phenotype correlations are in OXPHOS deficiency.

3127F

Targeted sequencing of mitochondrial exome in pediatric patients with mitochondrial diseases. *M. Tesarova[1], T. Honzik[1], H. Kratochvilova[1], A. Vondrackova[1], V. Stranecky[2], M. Rodinova[1], H. Hansikova[1], J. Zeman[1].* 1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague 2, Czech Republic; 2) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague 2, Czech Republic.

Mitochondrial diseases (MD) represent both clinically and genetically heterogeneous group of disorders. More than 130 nuclear genes have been described so far whose mutations lead to MD. Accurate targeting of the genetic analysis based on clinical symptoms or laboratory tests is possible for only a few types of MD (e.g. MNGIE syndrome, mitochondrial encephalocardiomyopathy - *TMEM70*, Leigh syndrome due to *SURF1* mutations). For most of MDs, even specialized enzymatic and protein analyses do not allow unambiguously to narrow a group of candidate genes. Next generation sequencing is a significant milestone in discovery of genetic bases of inherited diseases and it already transforms genetic diagnostics. In last two years, more than 25 MD-genes were identified by exome sequencing. In a group of 26 patients with mitochondrial disease we performed targeted sequencing of mitochondrial exome (>1100 nuclear genes mostly based on MitoCarta Inventory; http://www.broadinstitute.org/pubs/MitoCarta/index.html). In 6 patients, mutations in known MD-genes *TSFM*, *COX10*, *AIFM1*, *TK2*, and *MGME1* were found. In 3 patients no gene was prioritized. In the other patients candidate variants in genes with previously unknown association with MD were selected and are further evaluated to support their pathogenicity. *Supported by research projects UNCE 204011, RVO-VFN64165/2012 and grants IGA NT13114/4, IGA NT14156/3..*

3128W

Whole Genome Sequencing for Rapid Identification of Sequence Variants Associated with Recessive Canine Disease Models. *G.S. Johnson[1], D.P. O'Brien[2], R. Zeng[1], D. Gilliam[1], T. Mhlanga-Mutangadura[1], J.R. Coates[2], M.L. Katz[3], J.F. Taylor[4], R.D. Schnabel[4].* 1) Department of Veterinary Pathobiology, University of Missouri, Columbia, MO, USA; 2) Department of Veterinary Medicine and Surgery, University of Missouri, Columbia, MO, USA; 3) Mason Eye Institute, University of Missouri, Columbia, MO, USA; 4) Division of Animal Sciences, University of Missouri, Columbia, MO, USA.

Over the past 25 years we have accumulated over 100,000 canine DNA samples, many from dogs with a variety of inherited diseases. One area of focus is the identification of the causal mutations underlying these diseases with two objectives: (1) to facilitate the breeding of healthy dogs, and (2) to identify canine models for heritable human diseases. In 2010, we began sequencing the genomes of dogs believed to bear distinct recessive diseases. To date, we have generated genome sequence data from 28 dogs. Collaborators at UCLA and T-GEN have provided an additional 34 genome sequences from domestic dogs and closely related wild canids. Ongoing sequencing efforts and collaborations are expected to produce in excess of 100 genome sequences by the end of 2013. Our model for the identification of causal variants uses 1 case and N-1 controls as a starting point, where the case is assumed to be homozygous for a variant that is not present in any of the N-1 control animals. Additional models under development will accommodate more complex genetic etiologies such as compound heterozygosity and incomplete penetrance. Currently, our efforts have resulted in the identification of five variants that are apparently causal for their respective diseases. We shall describe an *SAG* mutation associated with progressive retinal atrophy as an example. Furthermore, these sequenced genomes have provided us with an opportunity to explore reverse genetics which involves predicting recessive phenotypes from heterozygous sequence variants, identifying homozygous individuals by screening our archived DNA samples, and finally, verifying the predicted phenotype.

3129T

Whole Exome Sequencing on two fetuses with Centronuclear Myopathy of consanguineous parents from Sudan shows two possible candidates. *N. Dohrn[1], V.G. Le[2], A. Petersen[3], P. Skovbo[4], I.S. Pedersen[2], A. Ernst[2], H. Krarup[2], M.B. Petersen[1].* 1) Department of Clinical Genetics, Aalborg University Hospital, Denmark; 2) Section of Molecular Diagnostics at Department of Clinical Biochemistry, Aalborg University Hospital, Denmark; 3) Institute of Pathology, Aalborg University Hospital, Denmark; 4) Department of Obstetrics and Gynaecology, Aalborg University Hospital, Denmark.

Arthrogryposis multiplexa congenita (**AMC**) is a rare congenital disorder characterized by multiple joint contractures. There are several possible underlying conditions but the common pathway is the lack of fetal movement. Usually termination of pregnancy is performed due to the known severity of the disorder. Exact diagnosis is identified at autopsy subsequently. We present a consanguineous healthy couple from Wadi Halfa, Sudan, with two separate pregnancies with female fetuses showing AMC. The mother's great grandfather was the grandfather of the father. Ultra sound (**US**) in the 13th week showed AMC, leading to termination of the first pregnancy. The second pregnancy showed the same malformation on US in the 13th week, also leading to termination. Array-CGH, karyotype (46,XX) and investigation for Congenital Myotonic Dystrophy were all normal. A full autopsy including CNS and muscle histopathological examination of the two foetuses were performed. The muscle examination included immunohistochemistry and electron microscopy. A diagnosis of Centronuclear Myopathy with autosomal recessive inheritance (**AR-CNM**) was proposed as there was a greater amount of centrally located nuclei and a greater variation in the muscle fibre diameter than expected for the gestational age and as no other specific myopathies were found.

CNM is seen with either X-linked, autosomal dominant or recessive inheritance. The autosomal recessive form of CNM is currently only reported to be caused by mutations in either the *BIN1* or *RYR1* genes, and many patients are without any genetic clarification. Some cases were detected during fetal life, but none of them were severe enough to result in termination of pregnancy, in contrast to our case. Blood samples and cell cultures from the parents and foetuses, respectively, were analyzed with Whole Exome Sequencing (WES). The WES result showed no mutations in the two previously known genes, *RYR1* and *BIN1*. This is in agreement with results from sanger sequencing previously carried out on *BIN1*. The WES suggested two possible candidates for a novel gene responsible for AR-CNM. They were identified by using multiple filtration parameters to filter from the starting collections of twelve thousands variants for each of the four exomes. Both genes are without any previous connection to a neuromuscular disease. We are in the process of carrying out confirmation on our WES results concerning the two candidates.

3130F

Value of whole exome sequencing for novel variant identification in Leber congenital amaurosis. *Y. Guo[1,14], I. Prokudin[2,3,14], C. Yu[4,5,14], J. Liang[5], Y. Xie[5], M. Flaherty[6], L. Tian[1], S. Crofts[6], F. Wang[1], J. Snyder[1], C. Donaldson[6], N. Abdel-Magid[1], L. Vazquez[1], B. Keating[1,7,8], H. Hakonarson[1,7,8,15], J. Wang[5,9,10,11,15], R. Jamieson[2,3,12,13,15].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Eye and Developmental Genetics Research Group, Western Sydney Genetics Program, The Children's Hospital at Westmead, Sydney, NSW, Australia; 3) Children's Medical Research Institute, Westmead, Sydney, NSW, Australia; 4) College of Life Sciences, Sichuan University, Key laboratory for Bio-resources and Eco-environment of Ministry of Education ,Sichuan Key Laboratory of Molecular Biology and Biotechnology, Chengdu 610064, PR China; 5) BGI-Shenzhen, Shenzhen 518083, China; 6) Department of Ophthalmology, The Children's Hospital at Westmead, Sydney, NSW, Australia; 7) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 8) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 9) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 10) King Abdulaziz University, Jeddah, Saudi Arabia; 11) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; 12) Discipline of Ophthalmology & Save Sight Institute, University of Sydney, Sydney, Australia; 13) Disciplines of Paediatrics and Child Health & Genetic Medicine, University of Sydney, Sydney, NSW, Australia; 14) Equal contribution; 15) Corresponding authors.

Leber congenital amaurosis (LCA) is a severe form of retinal dystrophy and children present with nystagmus and poor vision in the first year of life. There are at least 21 known disease genes and few distinguishing clinical features to focus candidate gene sequencing. Allele-specific assays testing for known variants and Sanger sequencing of specific gene regions provide answers for some patients. We applied whole exome sequencing in a family with LCA, where previous targeted assay and sequencing approaches had been unsuccessful in disease gene identification. We implemented two independent bioinformatic pipelines, BWA-GATK and SOAP, to interrogate the sequence data and maximize our capacity for valid variant identification. Our first pipeline resulted in 16 prioritized variants, while the second led to 8 prioritized variants. Four variants overlapped between the two groups of prioritized variants, one of which was the top candidate predicted from both pipelines, a novel homozygous nonsense mutation in *TULP1*. **TULP1** is a member of the tubby gene family and has roles in rhodopsin movement, synapse formation and survival of the photoreceptors, and causes approximately 1% to 2% of cases with LCA. The disease-causing variant was confirmed on Sanger sequencing and segregated appropriately in the family. This study highlights the value of a whole exome sequencing approach as a rapid means for known and novel variant identification in disease genes in the genetically heterogeneous retinal disease of LCA.

3131W

C5orf42 is the major gene responsible for OFD syndrome type VI. *E. Lopez[1], C. Thauvin-Robinet[1,2,3], N. El Khartoufi[4], L. Devisme[5], M. Holder[6], H. Ansart-Franquet[5], D. Lacombe[7], P. Kleinfinger[8], I. Kaori[9], J.I. Takanashi[10], M. Le Merrer[4,11], J. Martinovic[12,13], C. Noël[14], M. Shboul[15], L. Ho[15], Y. Güven[16], F. Razavi[4,11,17], L. Burglen[18,19], V. Darmency-Stamboul[21], M. Avila[1], J. Thevenon[1,2,3], B. Aral[1,20], H. Kayserili[22], S. Lyonnet[4,11,17], C. Le Caignec[23], B. Franco[24], B. Reversade[15], L. Faivre[1,2,3], J.B. Rivière[1,20], T. Attié-Bitach[4,11,17]*. 1) Equipe d'accueil EA 4271 GAD "Génétique des Anomalies du Développement", IFR Santé STIC, Université de Bourgogne, Dijon, France; 2) Centre de Référence Anomalies de Développement et Syndromes Malformatifs de l'interrégion Grand-Est, Hôpital d'Enfants, CHU, Dijon, France; 3) Centre de Génétique, Hôpital d'Enfants, CHU, Dijon, France; 4) Département de Génétique, Hôpital Necker-Enfants Malades, APHP, Paris, France; 5) Département d'Anatomo-Pathologie, Centre de Biologie et de Pathologie, CHRU, Lille, France; 6) Service de Génétique Clinique, Hôpital Jeanne de Flandre et Service de Neurologie Pédiatrique, Roger Salengro, CHRU, Lille, France; 7) Service de Génétique Médicale et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, CHU de Bordeaux, Laboratoire MRGM, EA4576, Université de Bordeaux, Bordeaux, France; 8) Département de génétique, Laboratoire Cerba - Saint-Ouen l'Aumone et Service de Gynécologie-Obstétrique, Hôpital R. Dubos, Pontoise France; 9) Department of Pediatrics, Shimada Center for Rehabilitation and Neurodevelopmental Intervention, Tama, Tokyo, Japan; 10) Department of Pediatrics, Kameda Medical Center, Kamogawa, Japan; 11) Unité INSERM U781 et Fondation IMAGINE, Hôpital Necker Enfants-Malades, Paris, France; 12) AP-HP, Unité de Foetopathologie, Hôpital Antoine Béclère, Clamart, France; 13) Département de Foetopathologie et Anatomopathologie, Laboratoire Cerba - Saint-Ouen l'Aumone, France; 14) Service de Gynécologie-Obstétrique, Hopital R. Dubos, Pontoise, France; 15) Laboratory of Human Embryology, Institute of Medical Biology, Agency for Science, Technology and Research (A*STAR), Singapore; 16) Istanbul University, Faculty of Dentistry, Department of Pedodontics, Istanbul, Turkey; 17) Service de génétique Hôpital Armand Trousseau, AP-HP, Paris, France; 18) Centre de Référence des malformations et maladies congénitales du cervelet, Hôpital Armand Trousseau, Paris, France; 19) Université Paris Descartes, Sorbonne Paris Cité, Paris, France Hôpital Necker-Enfants Malades; 20) Laboratoire de Génétique Moléculaire, Plateau Technique de Biologie, CHU Dijon, France; 21) Service de Pédiatrie 1, Hôpital d'Enfants, CHU, Dijon, France; 22) Medical Genetics Department, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey; 23) Service de Génétique Médicale, CHU, Nantes, France; 24) Telethon Institute of Genetics and Medicine, Napoli, Italy, Medical Genetics, Department of Medical Translational Sciences, University of Napoli Federico II, Napoli, Italy.

The oral-facial-digital syndrome type VI (OFD VI) or Varadi-Papp syndrome has recently been caracterised by the following diagnostic criteria: molar tooth sign (MTS) and one or more of the following: 1) tongue hamartoma and/or additional frenula and/or upper lip notch; 2) mesoaxial polydactyly of one or more hands or feet and 3) hypothalamic hamartoma. Because of the MTS, OFD VI belongs to the 'Joubert Syndrome Related Disorders' (JSRD). Its genetic etiology remains largely unknown although mutations in the TMEM216 gene, responsible for Joubert (JBS2) and Meckel-Gruber (MKS2) syndromes, have been reported in only two patients. Few mutations in the X-linked OFD1 gene have also been reported in male JBS patients with oral defects or female patients with OFD syndrome and mild MTS. Combined exome and Sanger sequencing identified compound heterozygous or homozygous causal mutations in the C5orf42 gene in 8/9 families including a severe fetus with microphthalmia, cerebellar vermis hypoplasia, corpus callosum agenesis, bilateral hand and feet preaxial and postaxial polydactyly, unilateral hand mesoaxial polydactyly with Y-shaped metacarpal phalange, and severe skeletal dysplasia but absent oral manifestations. Altogether, we identified 12 novel C5orf42 mutations in 11 cases from 8 families. All patients presented distal anomalies with constant feet preaxial polydactyly (11/11 cases) and frequent mesoaxial polydactyly (7/11 cases) in at least one distal extremity with Y-shaped metacarpal abnormality (6/8 cases). Mutations of this gene have already been reported in Joubert but classical additional features usually reported in JSRD appeared absent, such as polycystic disease and retinal disease. In contrast to highly heterogeneous JSRD with 20 causative genes identified to date, OFD VI seemed genetically highly homogeneous. In conclusion, we report that C5orf42 is the major gene responsible for OFD VI syndrome and confirm that OFD VI and JBS syndromes are allelic disorders, enhancing the complex, highly heterogeneous group of ciliopathies.

3132T

Independent origin of multifocal skin tumors and high frequency of mosaicism in patients with Tuberous Sclerosis Complex (TSC). *M.E. Tyburczy[1], J. Wang[2], S. Li[2], R. Thangapazham[2], Y. Chekaluk[1], J. Moss[3], D. Kwiatkowski[1], T. Darling[2]*. 1) Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Uniformed Services University of the Health Sciences, Bethesda, MD; 3) Cardiovascular and Pulmonary Branch, NHLBI, National Institutes of Health, Bethesda, MD.

Tuberous sclerosis complex (TSC) is an autosomal dominant disorder caused by mutations in TSC1 or TSC2. It is characterized by formation of tumors that are thought to arise through a two-hit gene inactivation model. Two-thirds of TSC cases are sporadic, and mosaicism is known to occur at low frequency in TSC. We sought to examine several questions on the pathogenesis of TSC skin lesions through analysis of a set of TSC patients enriched for those with adult presentation and diagnosis (PMID 21690595). Fibroblast-like cells were grown from 33 skin tumor samples obtained from 23 TSC patients and 23 parallel samples of normal-appearing skin. Long range PCR was used to amplify all coding regions and most nearby intronic sequence from each of TSC1 and TSC2, followed by library preparation for next generation sequencing. Over 3000x read depth was achieved, and sequencing data was analyzed using BWA, Picard, GATK, Python, and Matlab to identify all sequence variants present at an allele frequency ≥1%. We identified TSC2 mutations in 27 of 33 cultures, with allele frequencies ranging from 1% to 56%, and a single TSC1 mutation in one culture. Five cultures, 4 from patients with adult presentation of TSC, had no mutations in either TSC1 or TSC2. Sixteen cultures had two different mutations in TSC2, including one case of loss of heterozygosity (LOH). Twelve had only a single mutation identified. Seven (39%) of 18 patients showed mosaicism. Blood DNA analysis of one patient showed no evidence of the mosaic mutation seen in the cultured cells. In four patients, multiple tumors showed distinct second hit mutations in TSC2. Strikingly, 7 of the 14 second hit point mutations were CC>TT dinucleotide mutations, likely due to UV-induced damage. Such CC>TT mutations have never been identified as germline mutations in TSC1 or TSC2. In summary, there are multiple novel observations: 1) Skin tumors in TSC develop through two-hit inactivation of TSC2. 2) Different TSC skin lesions have distinct second hit events, indicating that they must arise independently. 3) Genomic loss or classic LOH is rare in TSC skin lesions. 4) UV-induced mutation is likely a contributing genetic event to development of skin lesions in TSC. 5) Mosaicism is common in TSC patients with adult presentation. 6) Lack of mutation identification in some TSC patients with adult presentation is consistent with extreme mosaicism as a potential cause.

3133F

REPS1 is a novel gene of Neurodegeneration with Brain Iron Accumulation. *A. DRECOURT[1], N. BODDAERT[2], I. DESGUERRE[2], D. CHRETIEN[1], A. MUNNICH[1,2], A. ROTIG[1]*. 1) INSTITUT IMAGINE, INSERM U781, PARIS, France; 2) DEPARTEMENT DE PEDIATRIE, HOPITAL NECKER-ENFANTS MALADES, PARIS, FRANCE.

Neurodegeneration with brain iron accumulation (NBIA) encompasses a group of rare neurogenerative disorders transmitted with an autosomal recessive inheritance. This heterogeneous group of disorders can be differentiated by clinical, brain MRI and molecular features. The hallmark features of NBIA include progressive extrapyramidal dysfunction (dystonia, rigidity, choreoathetosis) and iron accumulation in the brain, primarily in the basal ganglia. To date, mutations in PANK2, PLA2G6, FA2H, ATP13A2, C2orf37, CP, FTL and WDR45 have been associated with NBIA. The largest subgroup (50%) of NBIA observed so far is PKAN (pantothenate kinase associated neurodegeneration), caused by PANK2 mutations. The remaining individuals are said to have 'idiopathic NBIA', meaning that the underlying cause is not yet known. Consequently, the disease causing genes are still to be identified in several patients with idiopathic NBIA. We have performed exome sequencing in two sisters with NBIA and identified two heterozygous mutations in REPS1 gene encoding RALBP1 associated Eps domain containing 1. Reps1 is involved in endocytosis and the two mutations (p.Ala113Glu and p.Val78Leu) are located in its EH1 domain that interacts with Rab11-FIP2. The function of Reps1 in iron metabolism is currently unknown but it has been well demonstrated that its protein interacting partner, RabB11-FIP2, functions in transferrin recycling. Western Blot analysis detected a low Reps1 protein level in patient fibroblasts indicating a probable misfolding and degradation of the protein due to the missense mutations, as REPS1 transcript was found in normal amount. We investigated the iron metabolism and oxidative stress in patient's fibroblasts. Colorimetric ferrozine-based assay allowed us to detect a dramatic increase of iron levels in patient fibroblasts. We also observed an increase of ferritin and iron responsive protein (IRP1) amounts suggesting iron accumulation in those cells. In keeping with that, SOD2 and IRP1 protein amounts were highly increased whereas aconitase activity was decreased. All these result indicate that Reps1 mutation induce deregulation of iron metabolism and oxidative stress. In conclusion we report REPS as a novel gene for NBIA. Improvement in our understanding of the biochemistry and pathophysiology of this form of NBIA will help for novel therapeutics of this neurological condition.

3134W

Challenging diagnostic cases resolved by whole exome sequencing.
S.F. Suchy[1], J.E. Abdenur[2], T. Ben-Omran[3], W.K. Chung[4], K. Leydiker[2], S.J. Bale[1], A. Daly[1], E.V. Haverfield[1]. 1) GeneDx, Gaithersburg, MD; 2) Division of Metabolic Disorders, CHOC Children's, Orange CA; 3) Section of Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Weill Cornell Medical College, Doha, Qatar; 4) Columbia University Medical Center, New York, NY.

Accurate diagnosis is essential to the appropriate management of genetic disorders, particularly inborn errors of metabolism (IEM). We present three challenging cases in which the molecular cause was identified by whole exome sequencing (WES). Case 1: A one year-old child presented with hypotonia, developmental delay, encephalopathy, seizures, intermittent strabismus and dystonia. There were two similarly affected siblings. Extensive work-up of blood and urine, and imaging studies were non-revealing. CSF studies were declined by the parents. WES was performed (Agilent SureSelect XT2 All Exon V4 kit and an Illumina HiSeq2000), on the proband and parents. A homozygous p.Phe375Leu mutation in the tyrosine hydroxylase gene (TH) was identified. Both parents were carriers of the mutation. Case 2: A two year-old male presented with congenital ichthyosis, developmental delay, white matter changes, mild hip dysplasia and constipation. Studies included: a normal metabolic work-up (organic acids, very long chain fatty acids, amino acids, acylcarnitine profile) and a normal chromosome microarray. WES revealed two mutations in trans in the ALDH3A2 gene. One was a frameshift (p.Asn255IlefsX11) and the second was a deletion of exons 1-3. Case 3: A 3 1/2 year-old male was referred from neurology to the metabolic genetics clinic due to distal muscle weakness, motor delay, an unbalanced gait and speech delay. He had no calf hypertrophy or scoliosis. The metabolic work-up and MRI of the head and spine were normal. WES revealed a novel p.Arg844Cys mutation in the ATP7A gene. There are three main phenotypes associated with ATP7A mutations: Menkes disease, occipital horn syndrome and X-linked distal motor neuropathy. Patients with identical mutations who had different phenotypes have been described (Kennerson et al., AJHG, 86, 2010). This proband's phenotype and the location and the type of mutation are consistent with reported cases of X-linked distal motor neuropathy. A different mutation, p.Arg844His, affecting this same codon has been reported previously in association with Menkes disease, and this missense change was observed to result in a mis-localized protein (Moller et al., Hum Mut 26, 2005). The diagnosis of many genetic disorders remains a challenge. Disorders characterized by a broad phenotypic spectrum or occurring in a patient with an atypical presentation can further confound the diagnosis. WES offers another option for these diagnostic dilemmas.

3135T

Exome Sequencing Utility for Implication of Genes in Mendelian Disorders.
S. Jhangiani[1], M. Bainbridge[1], C. Gonzaga-Jauregui[2], W. Wiszniewski[2], T. Gambin[2], J. Lu[1], H. Doddapaneni[1], M. Wang[1], C. Buhay[1], H. Dinh[1], Y. Han[1], C. Kovar[1], J. Santibanez[1], J. Reid[1], D. Valle[3], E. Boerwinkle[1], D. Muzny[1], J. Lupski[2], R. Gibbs[1], Centers for Mendelian Genomics. 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland,.

Whole exome sequencing has proven to be the most successful way to elucidate the molecular basis of Mendelian disorders. The Human Genome Sequencing Center at Baylor College of Medicine has generated over 25 Tb of data for more than 250 phenotypes. Our high-throughput multiplexed exome pipeline, producing 2K capture libraries per month, is automated from library construction through annotated VCF generation in part using the Mercury pipeline. Since 2012, we joined the Centers for Mendelian Genomics with the Baylor-Johns Hopkins Center for Mendelian Genomics and have collected samples from 20 countries and initiated collaborations with over 86 institutions. The collection encompasses over 200 phenotypes and has generated 20 Tb of whole exome Illumina data.

Coverage of the genome's most understood regions is far greater with exome sequencing and still the most cost efficient method for gene association to disease. Whole genome shotgun sequencing generated from a patient with Charcot Marie Tooth syndrome was compared against whole exome data from the same individual. Not only were the majority of disease associated mutations found in the exome data, including the causative CMT locus in the patient, but an additional mutation was found in *SH3TC2* because of increased target coverage using this application.

Gene-disease association is still highly debated; however, discovery of a candidate mutation in several unrelated individuals with similar phenotypes gives power to disease implication. A candidate mutation was discovered in one family with a novel phenotype, sharing characteristics to Bohring-Opitz syndrome. The proband showed a *de novo* truncating mutation in *ASXL3*. Additional patients were found searching for other patients with the same *de novo* nonsense mutation in the identical gene - fostering the idea that molecular diagnostics is the future of the field. Whole exome data in conjunction with the molecular diagnostics will play an integral role in disease associations to genes.

3136F

Autosomal Dominant FSGS without extra-renal symptoms due to mutations in LMX1B gene.
K.J Johnson[1], O. Boyer[2, 3, 5], S. Woerner[2], F. Yang[1], E.O Oakeley[6], B. Linghu[1], O. Gribouval[2], M-J. Tete[2], JS. Duca[1], L. Klickstein[1], A.J. Damask[1], J.D. Szustakowski[1], F. Heibel[7], M. Matignon[8], V. Baudouin[9], F. Chantrel[10], J. Champigneuelle[11], L. Martin[12], P. Nitshke[13], M-C. Gubler[2], S-D. Chibout[6], C. Antignac[2,5,4]. 1) Novartis Institutes for Biomedical Research, Cambridge, Massachusetts; 2) Inserm U983, Hôpital Necker-Enfants Malades, Paris, France; 3) Service de Néphrologie Pédiatrique, Centre de Référence MARHEA, Hôpital Necker-Enfants Malades; 4) Département de Génétique, Centre de Référence MARHEA,Assistance Publique-Hôpitaux de Paris, Paris, France Hôpital Necker-Enfants Malades; 5) Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris; 6) Novartis Institutes for Biomedical Research, Basel, Switzerland; 7) AURAL et Service de Néphrologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France; 8) Service de Néphrologie, Hôpital Henri Mondor, Faculté de Médecine Paris XII, Assistance Publique-Hôpitaux de Paris, Paris, France; 9) Service de Néphrologie Pédiatrique, Hôpital Robert Debré, MARHEA, Université Denis Diderot, Assistance Publique- Hôpitaux de Paris, Paris, France; 10) Service de Néphrologie et Médecine Interne, Centre Hospitalier de Mulhouse, Mulhouse, France; 11) Service de Pathologie, CHU Nancy Brabois, Vandoeuvre, France; 12) Service de Pathologie, et Inserm UMR1098, CHU Dijon, Dijon, France; 13) Plateforme de Bioinformatique, Hôpital Necker, Enfants Malades, Paris, France.

Inherited forms of Focal segmental glomerular sclerosis (FSGS) are genetically heterogeneous and mutations in six genes (WT1, LMX1B, ACTN4, TRPC6, INF2 and ARHGAP24) have been reported in about 20% of families with autosomal dominant (AD) forms of the disease. In some AD cases various types of extra-renal involvement have been observed. For example LMX1B mutations are found in nail-patella syndrome (NPS, OMIM #161200) in which some cases exhibit FSGS with specific type III collagen lesions in the glomerular basement membrane (GBM). In one family out of a worldwide cohort of 74 with AD inheritance and no identified mutation, whole exome sequencing identified a mutation in LMX1B that segregated with FSGS in 5 affected individuals, all of whom lack the extra-renal features of NPS, including the GBM ultrastructural changes. Screening of the 73 other families in this cohort identified mutations in the same codon (R246) in two further families. Protein structure modeling suggests that R246 is a critical residue for interaction of LMX1B with the homeodomain of target genes. The results indicate that isolated FSGS can arise due to mutations in LMX1B, a gene normally associated with a broader syndromic phenotype. Other genes that underlie syndromes of which FSGS is a part of the phenotype should be considered in screening families in which no underlying cause of their familial FSGS has been identified.

3137W

Niemann-Pick Disease Type C Caused by Maternal Uniparental Isodisomy of Chromosome 18.
M. Fietz[1], S. Stark[1], A. Leo[1], G. Jenkins[2], B. Bennetts[2,3,4], J. Christodoulou[3,4,5]. 1) Dept Biochemical Genetics, SA Pathology (at WCH), North Adelaide, SA, Australia; 2) Dept Molecular Genetics, Western Sydney Genetics Program, Children's Hospital at Westmead, Sydney, NSW, Australia; 3) Discipline of Paediatrics and Child Health, Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 4) Discipline of Genetic Medicine, Sydney Medical School, University of Sydney, Sydney, NSW, Australia; 5) Genetic Metabolic Disorders Service, Western Sydney Genetics Program, Children's Hospital at Westmead, Sydney, NSW, Australia.

Background: Niemann-Pick disease type C (NP-C) is an autosomal recessive lysosomal storage disorder with an estimated Australian incidence of 1:140,000. It has a broad clinical spectrum, ranging from neonatal lethal presentation through to adult-onset neurodegenerative disease. NP-C is caused by mutations in either the *NPC1* (18q11.2, ~95% of cases) or *NPC2* (14q24.3) genes. Mutations in these genes lead to abnormal intracellular cholesterol trafficking, detectable by analysis of cholesterol accumulation and esterification in cultured cells. Case Study: Patient ES was a term baby born to non-consanguineous parents, presenting at 2-months of age with conjugated hyperbilirubinaemia and hepatosplenomegaly. Cultured fibroblast studies revealed that he was affected by NP-C. He was commenced on oral miglustat therapy at 5-months of age on a compassionate basis. When reviewed at 15-months of age he had motor and growth delay, but cognitive development was progressing well. Purpose of Study: The aim of the study was to determine the causative mutation(s) for NP-C in patient ES, to enable future prenatal testing for this family. Results: Sequence analysis of *NPC1* revealed homozygosity for a previously unreported single nucleotide deletion in exon 15 (c.2336delT). However, analysis of parental DNA revealed that the mutation was carried by his mother but not his father. Subsequent microsatellite analysis of chromosome 18 demonstrated maternal uniparental isodisomy (UPD) over the length of chromosome 18, revealing the cause of the homozygous mutation in ES. Further, these results strongly suggest a post-zygotic error as the cause of the UPD. Conclusion: We believe this is the first report of NP-C being caused by UPD for chromosome 18, and this case highlights the importance of parental carrier testing for confirmation of the genetic basis of Mendelian disorders.

3138T

Leber congenital amaurosis with early-onset severe macular atrophy and optic atrophy is likely pathognomonic of NMNAT1 mutations. *I. Perrault[1], M. Nicouleau[1], S. Hanein[1], N. Delphin[1], B. Gilbert-Dussardier[2], C. Vincent-Delorme[3], C. Edelson[4], C. Hamel[5], E. Silva[6], S. Defoort-Delhemmes[7], L. Fares-Taie[1], S. Gerber[1], X. Gerard[1], A. Goldenberg[8], A. Duncombe[8], G. Le Meur[9], P. Calvas[10], A. Munnich[1], O. Roche[11], H. Dollfus[12], J. Kaplan[1], J. Rozet[1].* 1) Genetics of retinal degeneration, INSERM U781-IMAGINE, Paris, France; 2) Centre Hospitalier Universitaire de Poitiers-Poitiers, FRANCE; 3) Centre Hospitalier d'Arras-Arras, FRANCE; 4) Department of Ophthalmology, Fondation Ophtalmologique Adolphe de Rothschild, Paris, France; 5) Neurosciences Institute, Hôpital Saint Eloi, Montpellier, France; 6) Department of Ophthalmology, Coimbra University Hospital, Coimbra, Portugal; 7) Department of Exploration of Vision and Neuro-Ophthalmology, Hôpital Roger Salengro, Centre Hospitalier Universitaire Régional, Lille, France; 8) Department of Genetics, Centre Hospitalier Universitaire, Rouen, France; 9) Department of Ophthalmology, Centre Hospitalier Universitaire, Nantes, France; 10) Department of Medical Genetics, Purpan Hospital, Centre Hospitalier Universitaire, Toulouse, France; 11) Department of Ophthalmology, Université Paris Descartes-Sorbonne Paris Cité, Paris, France; 12) Department of Clinical Genetics, Strasbourg University Hospital, Strasbourg, France.

Introduction: The nuclear nicotinamide mononucleotide adenyltransferase 1 (NMNAT1) encodes a homohexameric NAD- synthesizing enzyme as well as a chaperone that protects against neuronal activity-induced degeneration. Recently, NMNAT1 mutations have been reported to cause a highly specific Leber congenital amaurosis phenotype characterized by severe neonatal neurodegeneration of the central retina with early-onset optic atrophy. The purpose of the present study was to search for second NMNAT1 disease alleles in single heterozygote patients harboring the NMNAT1 phenotype. Methods: Nine sporadic cases and two sibs harboring single heterozygote NMNAT1 mutations were screened for copy number variations (CNV) using array comparative genomic hydridization (CGH, CytoScanHD) and mutations affecting regulatory elements or the splicing using Sanger sequencing of genomic DNA and/or lymphoblastoid cDNA. NMNAT1 haplotypes were constructed using SNP and microsatellite segregation analysis to search for shared ancestral alleles. Results: cDNA sequencing allowed identifying a one-exon duplication in the two sibs and 1/9 sporadic case sharing a same NMNAT1 haplotype. Array CGH analysis is ongoing to characterize the mutation at the genomic level. In 4/9 apparently unrelated sporadic cases, genomic DNA sequencing identified a same deep intronic change that affects a predicted regulatory element. All four cases were born to a parent originating from a French Indian Ocean island: La Réunion, suggesting that the change occurred on an ancestral allele. Haplotype analysis was consistent with this hypothesis. cDNA sequencing in 1/4 individual evidenced lack of amplification of the mRNA encoded by the allele harboring the deep intronic change. This suggests that the change, or a variant in linkage disequilibrium, is responsible for down regulation or instability of the mRNA. Conclusion: Here, we report that at least 6/10 LCA patients with single heterozygote NMNAT1 mutations carry a second disease allele undetected by exome sequencing. This result suggests that severe neonatal neurodegeneration of the central retina with early-onset optic atrophy is pathognomonic of NMNAT1 mutations. In-depth molecular analysis of the gene and surrounding regulatory elements should be considered in all patients harboring this highly specific LCA phenotype.

3139F

Lujan-Fryns Syndrome - conglomeration of different genetic entities? *K. Hackmann[1], S. Maas[2], N. Di Donato[1], E. Schrock[1], V.M. Kalscheuer[2], A. Rump[1].* 1) Institut fuer Klinische Genetik, TU Dresden, Dresden, Germany; 2) Max Planck Institute for Molecular Genetics, Berlin, Germany.

Lujan-Fryns Syndrome (LFS) has been described as X-linked intellectual disability (XLID) with a marfanoid habitus. Previous analyses identified mutations in three different genes on the X chromosome, namely MED12, UPF3B, and ZDHCC9. We investigate these genes for mutations in 30 patients clinically diagnosed with LFS. No mutation was detected in any of the three genes. Molecular karyotyping revealed one case of Phelan-McDermid syndrome and one case of microduplication 16p11.2. In a third case we found a previously undescribed de novo duplication on chromosome 12p13, which we consider as potentially pathogenic. In several other cases we identified familial variants of unknown pathogenicity. Five individuals without noticeable deletions or duplications were analyzed by X-exome sequencing. After filtering against entries in publicly available databases two missense mutations in two genes currently not connected with disease remained. These might account for the phenotypic appearance but additional studies have to be carried out. The mutations were found in NKAP and LAS1L, respectively. It seems that there are cases that resemble LFS at least on a clinical level and do not necessarily have to have an X-linked cause. Exome sequencing should be the next step in order to identify the genetic causes in these patients. The clinical diagnosis LFS implies an XLID and might therefore lead to a misguided diagnostic procedure. We assume that one essential criterion for the clinical diagnosis of LFS is an X-linked segregation of the disorder. If this is not the case there is probably no higher incident of aberrations on the X chromosome to be expected in LFS cases than in any other form of intellectual disability of unknown cause. It remains to be corroborated by studying additional cases, if LFS is a single clinical entity or if there is a phenotypic overlap between several disease entities.

3140W

Expression of mouse acidic mammalian chitinase in periplasmic space of Escherichia coli. *K. Okawa, A. Kashimura, K. Ishikawa, Y. Kida, K. Iwabuchi, Y. Matsushima, M. Sakaguchi, Y. Sugahara, F. Oyama.* Biotechnology Laboratory, Department of Applied Chemistry, Faculty of Engineering, Kogakuin University, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan.

Acidic mammalian chitinase (AMCase) have been shown to be closely associated with asthma in mouse model, allergic inflammation and food processing. Biochemical characterization of AMCase requires large quantities of purified protein. Currently, efforts to characterize these proteins structurally and biochemically rely on mammalian, insect cell expression systems and E. coli. We describe herein an E. coli-expression system allowing periplasmic production of active AMCase in E. coli. The mouse AMCase cDNA was cloned into the E. coli expression vector for extracellular expression in E. coli. Most of the expressed AMCase was present in the periplasmic space with chitinolytic activity measured using the chromogenic substrate, p-nitrophenyl β-D-N, N'-diacetylchitobiose (PNP-diacetylchitobioside). This recombinant protein can be used to elucidate detailed biomedical functions of the mouse AMCase.

3141T

Familial generalized seizures due to LGI1 mutation: importance of family history for genetic testing. *D.R.M. Amrom[1,3,4,7], F. Andermann[2,3,5,7], E. Andermann[1,3,4,6,7].* 1) Neurogenetics Unit; 2) Epilepsy Service and Seizure Clinic; 3) Montreal Neurological Hospital & Institute, Montreal, Quebec, Canada; 4) Department of Neurology & Neurosurgery; 5) Department of Pediatrics; 6) Department Human Genetics; 7) McGill University, Montreal, Quebec, Canada.

It is well known that patients with temporal lobe epilepsy may present with generalized seizures, and the temporal localization depends on further investigation. We present a family with 5 individuals in 3 generations where the clinical pattern consisted largely of generalized seizures, but who were then shown to have epilepsy due to an LGI1 mutation. The proband is a 46-year-old female who had four nocturnal seizures, all occurred around 5-6 am, the first one at age 19 years. She has been seizure free since adequate compliance with treatment. Her first EEG performed at 19 years showed an excess of slow waves at 2-4 Hz over both posterior regions without epileptic activity. Her second EEG at 22 years showed spikes and slow spike waves alternating over both temporal regions, mostly during drowsiness, and increased during hyperventilation. During intermittent photic stimulation, a photomyoclonic response appeared. Her 40-year-old sister had her first tonic-clonic seizure at 12 years which was generalized from the onset. All but one of her subsequent attacks occurred during sleep. Before the only seizure that occurred while awake, she felt diffuse numbness and heard a whooshing sound suggestive of neocortical temporal lobe involvement. The third sister is 52 years old; she had her first generalized seizure at 19 years. A year prior to this, she had transient symptoms of a tingling sensation associated with a wooshing noise. She later had other generalized attacks preceded by this aura. A diagnosis of neocortical or lateral temporal, possibly autosomal dominant partial epilepsy with auditory features (ADPEAF), was suggested. Although the proband only had nocturnal generalized seizures, LGI1 sequencing was performed on the basis of the family history. A c.611delC mutation leading to a frameshift and premature termination of the protein was identified. Generalized nocturnal and diurnal seizures associated with interictal generalized spike-wave activity occurring in a family with ADPEAF is unusual. They may represent secondarily generalized seizures or primary generalized seizures or both. In addition, photosensitivity in the proband is unusual as well. This family illustrates that patients with ADPEAF may present with generalized seizures and generalized spike and wave epileptic discharges. Intensive monitoring and attention to aura with auditory features should lead to accurate diagnosis of this genetically determined epileptic syndrome.

3142F

Type V Osteogenesis Imperfecta: don't miss the subtleties. *E.M. Carter[1], C.L. Raggio[1,], J.G. Davis[1,2].* 1) Center for Skeletal Dysplasias, Hospital for Special Surgery, New York, NY; 2) NY-Presbyterian Hospital, Weill Cornell Medical College, New York, NY.

Osteogenesis Imperfecta (OI) is a group of heritable connective tissue disorders characterized by predisposition to fracture. Diagnosis is made by clinical and radiographic criteria, family history, and increasingly through molecular genetic testing. Type V OI (OI5) accounts for ~4-5% of all dominant OI. It is a clinically heterogeneous group caused by dominant mutation of *IFITM5* (c.-14C>T). Distinguishing features can include hyperplastic callus formation after fracture/surgery, calcification of the radioulnar interosseous membrane, and/or radial head dislocation. Here we present two cases. Case 1: A now 12 yr old girl referred by her pediatrician at 5 yrs because of multiple fractures since the age of 9 mos. Family history: no consanguinity, no one else with fractures, skeletal dysplasia, or metabolic bone disease. X-rays: calcification of the radioulnar interosseous membrane, hyperplastic callus formation at the distal fibulas, mild scoliosis, mild bell-shaped thorax, gracile ribs, and osteopenia. Physical exam: ligamentous laxity, limited forearm supination/pronation, white sclerae, no dentinogenesis imperfecta (DI), and no contractures. No mutation detected in *col1α1/col1α2*. Findings were consistent with OI5. Molecular confirmation through *IFITM5* sequencing is pending. Case 2: A now 2 yr old boy of mixed Hispanic ancestry referred by his orthopaedic surgeon at 1 yr because of multiple fractures since the age of 3 mos. Family history: no consanguinity, no one else with fractures, skeletal dysplasia, or metabolic bone disease. After a radius/ulna fracture at 10 mos a skeletal survey showed multiple healed fractures. No mutation was detected in *col1α1/col1α2*. X-rays: gracile ribs, hypertrophic chondrocostal junctions, bilateral bowing of the radius and ulna, vertebral compression fractures, hypertrophic tibia callus, osteopenia, and moderate metaphyseal sclerosis. Physical exam: ligamentous laxity, white sclerae, and no DI. Findings were consistent with OI5, confirmed through mutation analysis of *IFITM5* (c.-14C>T). With accurate clinical diagnosis, >90% of individuals with OI have a *col1α1/col1α2* mutation. For individuals with no col1α1/col1α2 mutation, a skeletal survey should be performed to look for distinguishing features of OI5. Although clinical variability is wide, genetic testing is sensitive: 100% have the c.-14C>T *IFITM5* mutation. Accurate diagnosis is important in evaluation of non-accidental injury and for genetic counseling.

3143W

Mesenteric Cyst A Case Report and a Review of the Genetic Implications of this Rare Disease. *J. Davalos.* Area of Health # 1, Quito, Pichincha, Ecuador.

Case Report A 44 -year- old woman, using by 13 years a IUD, asymptomatic, the IUD was removed and we finded a 18 centimeters pelvic mass and a red spotty skin pigmentation in the left arm.The ultrasound scan reported a pelvic cyst containing 1,7 liters and a 3,1 cm right ovarian cyst. A computerized tomographic scan revealed a mesenteric cyst, the mass measured 14.1 × 10.8 cm and contained approximately one liter of serous fluid and also reported a small hepatic cyst. Discussion Mesenteric cysts are one of the most rare intra-abdominal tumors.1 The reported incidence ranges from 1/20,000 to 1/250,000 admissions to hospital.2 As proposed by Gross, mesenteric and omental cysts are thought to represent benign proliferations of ectopic lymphatics that lack communication with the normal lymphatic system.3 Other etiologic theories include (1) failure of the embryonic lymph channels to join the venous system, (2) failure of the leaves of the mesentery to fuse, (3) trauma, (4) neoplasia, and (5) degeneration of lymph nodes.4 Genetic Considerations Carney complex (CNC) is a familial multiple neoplasia syndrome transmitted as an autosomal dominant trait. CNC was initially described as the association of myxomas, spotty skin pigmentation, and endocrine overactivity. A variety of endocrine and nonendocrine tumors occur in patients with CNC.Primary pigmented nodular adrenocortical disease (PPNAD), a rare cause of ACTH-independent Cushing syndrome. PPNAD is observed in one-fourth of patients with CNC. Approximately half of the cases of CNC are familial. Putative genetic loci have been identified by linkage analysis at chromosome 2p16 and 17q22-24. Recently, the responsible gene on 17q22-24, PRKAR1A, was identified. Mutations in the PRKAR1A tumour suppressor gene were recently found in CNC1 mapping kindreds, while the CNC2 and perhaps other genes remain unidentified. Additional mutations in this gene were described later in other kindreds).Overall, inactivating mutations of this gene have been observed in ~41% of CNC kindreds.5,6Polycystic liver disease (PCLD) is characterized by the growth of fluid-filled cysts of biliary epithelial origin in the liver. PCLD is most often associated with autosomal dominant polycystic kidney disease (ADPKD).A causative gene, PCLD, was mapped to chromosome 19p13.2-13.1, with a maximum LOD score of 10.3. Haplotype analysis refined the PCLD interval to 12.5 cM flanked by D19S586/D19S583 and D19S593/D19S579.7.

3144T

The Role of Molecular Genetic Analysis in the Diagnosis of Primary Ciliary Dyskinesia. *R.H. Kim[1], D.A. Hall[2], E. Cutz[3], M.R. Knowles[4], K. Nelligan[5], K. Nykamp[6], M.A. Zariwala[7], S.D. Dell[8].* 1) Clinical Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Respirology, St Michael's Hospital, University of Toronto, Toronto, ON, Canada; 3) Division of Pathology, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 4) Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina USA; 5) Division of Respiratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 6) Prevention Genetics, Marshfield, Wisconsin, USA; 7) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina USA; 8) Division of Respiratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; Department of Pediatrics and Institute of Health Policy, Managment and Education, University of Toronto, Toronto, ON, Canada.

Rationale: Primary Ciliary Dyskinesia (PCD) is an autosomal recessive genetic disorder of motile cilia. The diagnosis of PCD has previously relied on ciliary analysis with transmission electron microscopy or video microscopy. However, patients with PCD may have normal ultrastructural appearance and ciliary analysis has limited accessibility. Alternatively, PCD can be diagnosed by demonstrating biallelic mutations in known PCD genes. Genetic testing is emerging as a diagnostic tool to complement ciliary analysis where interpretation and access may delay diagnosis. **Objectives:** To determine the clinical utility of genetic testing of patients with a confirmed or suspected diagnosis of PCD in a multi-ethnic urban centre. **Methods:** 28 individuals with confirmed PCD on transmission electron microscopy of ciliary ultrastructure and 24 individuals with a probable diagnosis of PCD based on a classical PCD phenotype and low nasal nitric oxide had molecular analysis of 12 genes associated with PCD. **Results:** 28/49 (57%) who underwent ciliary biopsy were diagnosed with PCD through an ultrastructural defect. Of the 52 individuals who underwent molecular genetic analysis, 22 (42%) individuals had two mutations in known PCD genes. 24 previously unreported mutations in known PCD genes were observed. Molecular findings also aided in the interpretation of inconclusive biopsy results. Combining both diagnostic modalities of biopsy and molecular genetics, the diagnostic yield increased to 69% compared to 57% based on biopsy alone. **Conclusions:** The diagnosis of PCD is challenging and has traditionally relied on ciliary biopsy which is unreliable as the sole criteria for a definitive diagnosis. Molecular genetic analysis can be used as a complementary test to increase the diagnostic yield and facilitate carrier testing and genetic counseling.

3145F

Genotype-Phenotype Correlation in Bardet- Biedl syndrome. *O. MHAMDI[1], H. CHAABOUNI[1,2].* 1) Department of Human Genetics, Faculty of Medicine, Tunis, Tunisia; 2) Department of Hereditary disorders, Charles Nicolle Hospital. Boulevard 9 Avril 1006 Tunis.

Bardet-Biedl syndrome (BBS) is an autosomal recessive disease characterized by retinal dystrophy, obesity, postaxial polydactyly, learning disabilities, renal involvment, and male hypogenitalism. BBS is a genetically heterogeneous with 17 genes (BBS1-BBS17), mutations in BBS genes accounting for approximately 70-80% of cases. Triallelic inheritance has been suggested in about 5% of cases. The genotype-phenotype correlation in BBS was unclear and can be explained by genetic locus heterogeneity, triallelism and modifier genes such as CCDC28B.In order to analyze more clearly, we evaluated a world wide cohort of 371 patients from 330 families with Bardet-Biedl syndrome in whom disease-causing genes were identified. The phenotype spectrum of studied BBS patients was ranged between mild to severe. Thus, our studies have important implication for molecular diagnosis and genetic counselling in BBS patients.

3146W

Fragile X triplet repeat instability is influenced by both the presence and position of AGG interruptions in intermediate and premutation alleles. *S. Nolin[1], R. Cao[2], J. Taylor[2], A. Glicksman[1], N. Ersalesi[1], W.T. Brown[1], J. Coppinger[2], G. Latham[2], A. Hadd[2].* 1) New York State Institute for Basic Research in Developmental Disabilities, Staten Island, NY; 2) Asuragen, Inc., Austin, TX.

Fragile X syndrome is a common cause of inherited intellectual disability that results from the expansion of a CGG repeat in the 5' untranslated region of the FMR1 gene to >200 copies. We examined FMR1 repeat instability in 762 transmissions of fragile X intermediate (45-54 repeats) and small premutation (55-90 repeats) alleles. Novel PCR assays were used to identify the number and position of the AGGs that occur within the CGG repeat region. Twenty-eight percent of the transmissions contained no AGGs, 44% had one AGG and 27% had two AGGs. Consistent with prior studies, the absence of AGG interruptions within the repeat was associated with an increased risk for instability on transmission as well as an increased risk for expansion to a full mutation in larger alleles. Alleles with two AGGs were least likely to exhibit instability or expansion to a full mutation on transmission. Alleles with one AGG had moderate risks for instability or expansion to full mutation. Surprisingly, the position of the single AGG interruption in these alleles was correlated with instability. Most of the one-AGG alleles had an interruption at position 10 while a minority had one at position 11. Analysis of one-AGG alleles with 45-59 repeats revealed an increased risk of instability for alleles with an AGG at position 10 compared to those with an AGG at position 11. Similarly, analysis of one-AGG alleles with 70-90 repeats revealed an increased risk for full mutation expansion for alleles with an AGG at position 10 compared to those with the AGG at position 11. These results and the fact that AGGs in expanded alleles occur more frequently at position 10 than position 11, suggest that AGGs at position 11 have a greater influence in stabilizing the FMR1 repeat region on transmission.

3147T

Identification of a novel nonsense mutation and a missense substitution in the AGPAT2 gene causing congenital generalized lipodystrophy type 1. *A. Haghighi[1], M. Razzaghy-Azar[2], A. Talea[2], M. Sadeghian[3], S. Ellard[4], A. Haghighi[5].* 1) University of Toronto, Toronto, Ontario, Canada; 2) Endocrinology and Metabolism Research Center, Tehran University of Medical Sciences, Tehran, Iran; 3) Pediatric Gastroenterology Department, H. Aliasghar Hospital, Tehran University of Medical Sciences, Tehran, Iran; 4) Institute of Biomedical and Clinical Science, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7PS, UK.

Congenital generalized lipodystrophy (CGL) is an autosomal recessive disease characterized by the generalized scant of adipose tissue. CGL type 1 is caused by mutations in gene encoding 1-acylglycerol-3-phosphate O-acyltransferase-2 (AGPAT2). A clinical and molecular genetic investigation was performed in affected and unaffected members of two families with CGL type 1. The AGPAT2 coding region was sequenced in index cases of the two families. The presence of the identified mutations in relevant parents was tested. We identified a novel nonsense mutation (c.685G>T, p.Glu229*) and a missense substitution (c.514G>A, p.Glu172Lys). The unaffected parents in both families were heterozygous carrier of the relevant mutation. The results expand genotype-phenotype spectrum in CGL1 and will have applications in prenatal and early diagnosis of the disease. This is the first report of Persian families identified with AGPAT2 mutations.

3148F

Phenotype - genotype correlation in a Colombian Oculocutaneous Albinism (OCA) cohort. *C. Lattig[1], D. Sanabria[1], A. Fernandez[2], O. Urtatiz[1], Fundacion Contraste - Albinos por Colombia.* 1) Laboratorio de Genética Humana, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogota, Colombia; 2) Hospital Militar, Bogotá, Colombia.

Oculocutaneous albinism (OCA) is a genetic condition of melanin synthesis characterized by a complete lack or generalized reduction in pigmentation of hair, skin and eyes. OCA1, one of the four types of OCA is the most frequent worldwide and is caused by mutations in TYR gene. The TYR gene consists of 5 exons spanning about 65 kb of genomic DNA and encoding a protein of 529 amino acids. In the present study, we have 20 unrelated cases of albinism with complete ophthalmological evaluation. Sequencing of the TYR gene in these individuals revealed three novel mutations, one nonsense mutation, c.551C>G (S184X) and two missense mutation, c.739T>C (C246R) and c.163 T>G (p.C55G), all of them in exon 1. We also found the p.R402Q temperature-sensitive mutation in three individuals. In three cases we were not able to detect any mutations in any of the five exons of the TYR gene. Individuals 07 and 08 are siblings and both present the novel mutation p.C246R in heterozygous state; in addition individual 08 has the variant R402Q in heterozygous state and has a more severe albino phenotype than individual 07. We present the first report of a phenotype - genotype correlation in a Colombian OCA cohort.

3149W

Support to the International Rare Diseases Research Consortium: a new service to the research community. *P. Lasko[1,2,3], B. Cagniard[4], S. Höhn[4], R. Favresse[4], D. Désir-Parseille[4], S. Peixoto[4], M. Bellanger[4], N. Lévy[4], S. Aymé[4].* 1) International Rare Diseases Research Consortium, Montreal (IRDiRC), Quebec, Canada; 2) CIHR Institute of Genetics, Montreal, Quebec, Canada; 3) Department of Biology, McGill University, Montreal, Quebec, Canada; 4) SUPPORT IRDiRC, Paris, France.

The International Rare Diseases Research Consortium (IRDiRC) brings together members that share common goals and principles and have agreed to work in a coordinated and collaborative manner within a multinational consortium to advance research in this critical area. The IRDiRC scientific secretariat has been established at the Rare Diseases Platform in Paris. IRDiRC's two main objectives are to deliver 200 new therapies for rare diseases and means to diagnose most rare diseases by the year 2020. Its members are over 30 public and private research funding organizations that have each dedicated over 10 million US$ to research into rare diseases. A number of challenges will be addressed through collaborative actions: establishing and providing access to harmonised data and samples, performing the molecular and clinical characterisation of rare diseases, boosting translational, preclinical and clinical research, and streamlining ethical and regulatory procedures. The consortium has established three Scientific Committees. The Diagnostics committee advises on research related to the diagnoses of rare disease, including sequencing and characterization of these diseases. The Interdisciplinary committee provides expertise on cross-cutting aspects of rare diseases research including issues related to ontologies, natural history, biobanking, and registries. The therapies committee gives guidance for the pre-clinical and clinical research aiming to deliver new therapies for rare diseases. The guiding principles of IRDiRC, its plan for action and its achievements so far will be presented, as well as ways for the genetic research community to get engaged in this global effort.

3150T

Role of CFTR in regulating spermatogenesis and Implications in male infertility. *H. Sharma[1], R. Prasad[1], S.K. Singh[2], R. Mohan[2].* 1) Department of Biochemistry, PGIMER, Chandigarh, India; 2) Department of Urology, PGIMER, Chandigarh, India.

The cystic fibrosis transmembrane conductance regulator (CFTR) is a cAMP - activated Chloride and HCO3 conducting channel, mutations of which are known to be associated with male infertility. CFTR mutations have been identified in patients presented with spectrum of genital phenotypes ranging from impaired spermatogenesis to CAVD and obstruction in reproductive tract other then vas-deferens or epididymis, although some contradictory reports are also available. Moreover spectrum of mutation identified in Indian patients with classical CF and infertile CBAVD male is heterogeneous and is completely different from that of Caucasians population, therefore the Present study was undertaken to establish the spectrum and frequency of CFTR mutation in Indian infertile male population and to understand the possible involvement of CFTR gene in the etiology of male infertility other than CAVD. Blood samples from infertile males with, obstructive azospermia (n=25) with palpable vas deferens and oligospermia (n=100) were used for genomic DNA isolation and screening mutations in seven exons (2,3,4,7,8,10) hot spots of CFTR gene. Delta F-508, N1303K, R553X, G551D, G542X, 621+1G>T and W1282X were the other known most common mutation screened in Indian infertile males through allele specific ARMS PCR analysis. IVS8-Poly T allele polymorphism was also determined in all infertile male patients. Out of 25 cases with obstructive azospermia 17(68%) were found to have mutation in single CFTR allele where as among 100 oligospermic males only 17% males were also found to have mutation in CFTR gene. Delta F-508 (9.5%) and IVS8 (5T) allele (22%) were the most common mutations identified in Indian infertile male population. Two novel mutations in exon 10 of CFTR were also indentified in patients with obstructive azospermia. The increased frequency of CF mutations in males with reduce sperm quality and quantity and in males with azospermia without CBAVD suggest CFTR protein may be involved in the process of spermatogenesis or sperm maturation apart from playing critical role in the development of epididymal gland and vas-deferens. Moreover exon 10 in NBD1 region of CFTR playing important role in etiology male infertility,therefore Indian infertile male presented with obstructive azospermia or spermatogenetic defect and opting for assisted reproductive technology (ART) should be advised for screening of exon 10 of CFTR.

3151F

Milder course in Duchenne patients with nonsense mutations and no muscle dystrophin. *M. Zatz, R.C.M. Pavanello, M. Lazar, N.C.V. Lourenço, A. Cerqueira, L. Nogueira, M. Vainzof.* Human Genome and stem cell research center, Biosciences Institute, Universidade de Sao Paulo, Sao Paulo, SP, São Paulo, Brazil.

Duchenne muscular dystrophy (DMD) is a lethal X-linked condition caused by mutations in the dystrophin gene which results in the absence of muscle dystrophin. The course is usually severe and very similar in affected patients. Without careful management, loss of ambulation occurs between 10-12 years of age. On the other hand, in Becker muscular dystrophy (BMD) there is a wide variability in the severity of the phenotype which has been mainly associated with the site of the deletion and to the amount of muscle dystrophin. Therefore, the quantity/quality of muscle dystrophin has been strongly associated with appropriate muscle function. Animal models with null mutations and no muscle dystrophin are of great interest. The mdx mice are almost asymptomatic, while the GRMD (golden retriever muscular dystrophy) dogs are usually severely affected representing the best model for DMD. Such clinical difference might be related to the size of the dystrophic muscle. We have previously identified two exceptional GRMD dogs with a very mild phenotype and no muscle dystrophin. A DMD patient with a milder course despite the absence of muscle dystrophin was also reported (Dubowitz, 2006). Here we report two additional very rare DMD patients with nonsense mutations in the DMD gene and no muscle dystrophin. Case 1, currently age 14, is only mildly affected, with discrete calves' hypertrophy, some difficulties for running and climbing stairs but with normal ability to walk. His youngest brother shows a severe course, being confined to a wheelchair at age 9. Both carry an out of frame duplication in exon 2. Case 2, is an isolated DMD patient who, at age 15, is able to walk without difficulties and climb stairs with the aid of the bannister. He carries an out-of-frame deletion encompassing exons 51 to 54. Dystrophin IF and WB with three antibodies against the N-terminal, rod domain and C-terminal, showed a typical DMD pattern, in both of them. Merosin was positive, and sarcoglycans were faint, as observed in most DMD dystrophin deficient patients. Therapeutic trials aiming the amelioration of muscle function have been targeting the production of muscle dystrophin in affected DMD patients. Importantly, our observations indicate that it is possible to have a functional large muscle even without dystrophin. Finding the underlying protective mechanisms is of utmost importance and may lead to new avenues for treatment.

3152W

MEFV Gene Mutation Detection In Arabic Patients. *R. Taha[1], S. Ayesh[2], M. Kambouris[1,3], H. El-Shanti[1,4].* 1) Molecular Genetics, Shafallah Medical Genetics Center, Doha, Qatar; 2) Gene Medical Labs, Gaza, Palestinian Territory; 3) Yale University School of Medicine, Genetics, New Haven CT, USA; 4) University of Iowa, Pediatrics, Iowa City, IA, USA.

Autoinflammatory diseases are characterized by inflammation in the absence of high-titer autoantibodies or antigen-specific T cells. Familial Mediterranean fever (FMF) is the archetypal hereditary autosomal recessive periodic fever syndrome & auto-inflammatory disease characterized by recurrent self-limiting episodes of fever & painful polyserositis. It is found in families of Mediterranean ancestry, especially non-Ashkenazi Jews, Armenians, Turks, and Arabs. The offending MEFV gene localizes at Hsa 16p, encodes the pyrin (marenostrin) protein, it is highly polymorphic with multiple disease causing mutations and normal polymorphisms. In Arabic FMF patients the spectrum and distribution of MEFV mutations are distinctive and the portion of unidentified mutations [50%] is the highest amongst the groups affected by FMF. The MEFV genomic region in 100 Palestinian patients with clear FMF symptomatology consistent with the clinical diagnostic criteria and with only one identified pathogenic mutation was screened to identify the second pathogenic mutation as well as coding and non-coding variations, large duplications or deletions and intronic variations. Mutation analyses involved sequencing of exons and splice sites, sequencing putative regulatory regions by using Long range PCR, Multiplex Ligation-dependent Probe Amplification (MLPA) to detect large deletions or duplications & sequencing of the entire genomic region of MEFV. No second pathogenic mutation was identified in any of the samples by sequencing MEFV exons, splice sites, as well as putative regulatory regions. MLPA did not detect any large MEFV genomic deletions or duplications. Twenty rare intronic variants were identified (each in 1-3 patients) and were not present in ≈700 ethnically matched control chromosomes. The biological significance of these variations could not be determined. There are strong evidences of preferential amplification of one allele over the other due to extensive polymorphism within the genomic sequence that would account for the lack of detection of the second pathogenic mutation. Alternatively, the effects of modifier genes or other loci that influence the clinical picture of FMF in Arabic populations can not be excluded. The comprehensive identification of MEFV mutant alleles among FMF patients is essential for the efficient examination of specific genotype-phenotype correlation patterns and for development of molecular tools to support the clinical diagnosis.

3153T

PhenomeCentral: An Integrated Portal for Sharing and Searching Patient Phenotype Data for Rare Genetic Disorders. *M. Brudno[1], M. Girdea[1], S. Dumitriu[1], S. Köhler[2], P.N. Robinson[2], A.J. Brookes[3], K. Boycott[4], C.F. Boerkoel[5], W.A. Gahl[5], Canadian CARE for RARE Consortium (FORGE) and NIH Undiagnosed Diseases Program.* 1) Centre for Computational Medicine, Hospital for Sick Children & University of Toronto, Toronto, ON, Canada; 2) Charite Hospital, Berlin, Germany; 3) Department of Genetics, University of Leicester, Leicester, United Kingdom; 4) Children's Hospital of Eastern Ontario, Ottawa, ON Canada; 5) National Institutes of Health Undiagnosed Diseases Program.

The availability of low-cost genome sequencing has allowed for the identification of the molecular cause of hundreds of rare genetic disorders. Solved disorders, however, only represent the 'tip of the iceberg'. Because the discovery of disease-causing variants typically requires confirmation of the mutation or gene in multiple unrelated individuals, an even larger number of genetic disorders remain unsolved due to difficulty identifying second families. With many groups now tackling these remaining undiagnosed disorders, which may be present in only a handful of individuals seen at different hospitals and sequenced by different centers, it is critical to establish effective and secure data-sharing techniques that allow clinicians and scientists to identify additional families via phenotype and genotype searches.

To address this need, we have developed PhenomeCentral (http://phenomecentral.org), a repository for secure data sharing targeted to the rare disorder community. Each patient record within PhenomeCentral consists of a thorough phenotypic description capturing observed abnormalities as well as relevant absent manifestations, expressed using Human Phenotype Ontology terms. Furthermore, each record can be labeled by the creator as: *private* hidden from everyone except the contributor; *public* viewable and searchable by all registered users; or *matchable* the record cannot be directly viewed or searched, but is reachable via an automated phenotype matching system (following Cafe Variome principles) which informs contributors of the existence of profiles similar to their cases. The phenotypic features shared among these records are presented without revealing additional patient information or the contributors, enabling direct communication for any subsequent data sharing.

PhenomeCentral currently incorporates phenotype data for >400 patients with rare genetic disorders without a molecular diagnosis, including: 200 from the Canadian CARE for RARE project and 150 from the NIH Undiagnosed Diseases Program (UDP). Clinical geneticists and scientists studying rare disorders can request accounts, and new patients can be added either using the PhenoTips User Interface, built into PhenomeCentral, or uploaded in bulk. An interface allowing for the deposition and analysis of genomic data (whole-exome VCF files and CNVs) is under development.

3154F

Ocular phenotypes in aneurysm syndromes collected from GenTAC (National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions) Registry. *G. Oswald[1], KW. Holmes[1,4], S. LeMaire[2], W. Ravekes[1], NB. McDonnell[3], C. Maslen[4], RV. Shohet[5], RE. Pyeritz[6], R. Devereux[7], DM. Milewicz[8], EM. Reynolds[9], JP. Habashi[1], GenTAC Registry Consortium.* 1) Johns Hopkins University, Baltimore, MD; 2) Baylor College of Medicine, Houston, TX; 3) NIA at Harbor Hospital, Baltimore, MD; 4) Oregon Health & Science University, Portland OR; 5) Queen's Medical Center, Honolulu, HI; 6) The University of Pennsylvania, Philadelphia, PA; 7) Weill Cornell Medical College of Cornell University, New York, NY; 8) University of Texas Medical School at Houston, Houston, TX; 9) University of Maryland, Baltimore, MD.

The National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC) has enrolled over 3400 patients with aortic aneurysm syndromes, including Marfan (MFS), Loeys-Dietz (LDS), vascular Ehlers-Danlos (vEDS), Bicuspid Aortic Valve with enlargement (BAVe) and Familial Thoracic Aortic Aneurysm Disease (FTAAD). The GenTAC Registry was queried for prevalence of ocular features, showing its utility to investigators studying non-cardiac features of these disorders. Ectopia lentis (ECL) was reported in 53% of patients with MFS, with equal distribution between age and gender, consistent with previously published data. ECL was not observed in LDS or vEDS, and rarely reported in other aneurysm conditions (<1%). Retinal detachment (RD) was observed in 8.7% of MFS and rarely in other aneurysm syndromes (<1%). ECL and RD were concomitant in 51 of 61 MFS patients. Glaucoma and cataracts were reported in 4% and 8.3% of MFS patients, respectively. Myopia was most commonly reported in MFS, but not significantly different than that observed in the general population. LDS showed the highest prevalence of strabismus/amblyopia of 12.3% as compared to MFS (8.3%) and other syndromes. Patients with MFS are commonly counseled to avoid LASIK correction of refractive error due to speculation that surgery's mechanical forces may increase RD risk. Analysis of the operative data available in a subset of patients found that 3.1% of enrolled patients with a confirmed diagnosis of MFS reported having undergone LASIK. Of the 12 MFS patients who had LASIK, 7 had ECL (58.3%) and 2 had RD (16.7%). 2/12 patients who had lens extraction (16.7%) and 5/22 with lens implants (23%) had RD. Direct causation cannot be inferred because the timing of LASIK in relation to the RD was not available. This review of ocular data within the GenTAC registry identifies a >50% prevalence of lens dislocation in MFS and a notable absence of lens dislocation in LDS as previously reported; therefore, presence or absence of lens dislocation is a useful diagnostic tool in differentiating patients. Furthermore, this data supports the theoretical risk that LASIK correction may increase the risk of retinal detachment in patients with connective tissue disorders however more detailed studies investigating the temporal relationship need to be done.

3155W

Budd-Chiari syndrome as the presenting symptom of familial thrombotic thrombocytopenic purpura caused by a novel frameshift mutation of ADAMTS13. *R. Parvari[1,3], M. Arafat[1], K. Akel[2], B. Yerushalmi[2], M. Aviram[2], D. Landau[2].* 1) Shraga Segal Dept of microbiology, immunology and Genetics, Ben Gurion Univ of the Negev, Beer Sheva, Israel; 2) Department of Pediatrics, Soroka Medical Center, Beer Sheva, Israel; 3) National Institute of Biotechnology, Ben Gurion Univ of the Negev, Beer Sheva, Israel.

Both Budd Chiari syndrome (BCS) and thrombotic thrombocytopenic purpura (TTP) are rare pediatric diagnoses. Thrombotic microangiopathy, the pathological hallmark of TTP, may be due to acquired or congenital defects in the protease ADAMTS13, which cleaves the vWF multimers and prevents uncontrolled thrombosis. We have been treating an extended Bedouin family of 5 patients with different TTP manifestations. The index case, a 10 years old boy previously healthy was admitted because of headache, jaundice, epistaxis and macro hematuria of 3 days duration. He was further diagnosed with BCS caused by occlusion of the hepatic veins. The child has been since treated with fresh frozen plasma infusions and enoxaparin. He experienced several TTP relapses. Liver biopsy shows evidence of BCS, without cirrhosis. The child has experienced an early (hematologic) good response to plasma treatment but no venous recanalization could be obtained, leading to signs of severe portal hypertension Four other affected members in the extended family experienced TTP-like episodes during childhood or pregnancy. Three patients were found to have very low (< 5%) ADAMTS13 levels. Sequencing ADAMTS13 has revealed a novel frameshift mutation that will result in the absence of 82 amino acids the carboxy terminus deleting the CUB2 domain. This is the first description of familial TTP presenting with BCS, a major vessel thrombosis.

3156T

In silico and molecular analyses of mutations that alter mRNA splicing of COL1A1. *J. Schleit, S. Bailey, T. Tran, D. Chen, P.H. Byers.* Pathology, University of Washington, Seattle, WA.

Approximately 10-20% of inherited pathogenic mutations alter mRNA splicing. Phenotypes of affected individuals are often dependent on the stability of the mRNA produced by the mutant allele. Currently, the downstream effects of mutations on mRNA splicing are often difficult to predict, complicating diagnoses based on DNA sequence alone. As the use of molecular diagnostics moves increasingly to DNA sequencing, the need to understand the effects of splice site mutations will increase further. We sought to identify factors that determine the mRNA splice products in the heritable disorder osteogenesis imperfecta (OI). To better understand the relationship between genotype and mRNA splicing outcomes, we examined clinical and molecular data from 219 OI patients with 122 unique mutations within the introns COL1A1, which encodes the proa1(1) chain of type I collagen. We examined mutations representing OI types I-IV with approximately 73% of patients diagnosed with OI Type I, 11% with OI Type IV, 6% with OI Type III, and 10% with OI Type II. 106 of the 219 (48%) of patients had mutations in the canonical GT-AG splice donor or acceptor site. All patients with OI Type I, for which cDNA was examined, produced unstable mRNAs with premature termination codons due to translational frameshifts. 34 individuals with G to A transitions in the final nucleotide of COL1A1 introns (IVS(X)-1G>A) had OI type I phenotypes, due to shift of the splice acceptor site one nucleotide downstream when the first nucleotide of the exon was guanine. In contrast, 72 individuals with IVS(X)+1G>A mutations had highly variable OI phenotypes (Types I-II) and the splice outcomes yielded diverse and often multiple mRNA products. In patients with OI Types II-IV we detected abnormal, stable mRNAs resulting from in-frame splice outcomes. For splice donor site mutations there was no correlation between outcome and intron size or the predicted strength of native and alternative splice sites in these individuals. To examine additional factors which may impact splice outcome, we treated wild-type human fibroblasts with the transcriptional inhibitor Actinomycin D. Using intron specific primers, we measured the speed and order of intron splicing of COL1A1 using quantitative PCR and capillary electrophoresis. mRNA splice outcomes were independent of the speed at which COL1A1 introns were spliced, however the order in which introns are normally removed appears to play a role in determining splice outcome .

3157F

Characterization of Escherichia coli-expressed acidic mammalian chitinase. *A. Kashimura, K. Okawa, K. Ishikawa, Y. Kida, K. Iwabuchi, Y. Matsushima, M. Sakaguchi, Y. Sugahara, F. Oyama.* Biotechnology Laboratory, Department of Applied Chemistry, Faculty of Engineering, Kogakuin University, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan.

Chitinase hydrolyze chitin, a polymer of N-acetyl-D-glucosamine, which is present in a wide range of organisms, including fungi, insects, and parasites. Although mammals do not have chitin synthase, chitinase genes were found in both human and mouse genomes. It has been known that chitinase activity is significantly increased in plasma from patients with Gaucher disease, an autosomal recessive lysosomal storage disorder. Recent researches have shown that elevated expression of chitinases is closely associated with progression of allergy and asthma. The pathophysiological role for the mammalian chitinases remains unproven. To understand how these chitinases interact with chitin and degrades it, we expressed and purified acidic mammalian chitinase in Escherichia coli as a fusion protein. The recombinant protein showed a robust peak of activity with maximum was observed at pH 2.0, where the optimal temperature was 54 °C. This protein also shows a heat-stable until 56 °C at both pH 2.0 or 56 °C at pH 7.0 for 30 min. The chitinolytic activity of the recombinant AMCase against PNP-diacetylchitobioside was comparable to CHO-expressed one. Furthermore, the recombinant AMCase bound to chitin beads and cleaved colloidal chitin and released mainly chitobioside fragments. Thus, E. coli-expressed AMCase possesses chitinase functions comparable to CHO-expressed one.

3158W

Total uniparental isodisomy of all 23 paternal chromosomes with whole genome homozygozity in blood but not in skin in a Czech girl with deafness. *P. Seeman[1], M. Malíková[2], J. Neupauerová[1], M. Trková[3], M. Putzová[3].* 1) Child Neurology, DNA Laboratory, 2nd Medical Faculty, Charles University and University Hospital Motol Prague, Praha 5, Czech Republic; 2) Institute of Biology and Medical Genetics, 2nd Medical Faculty, Charles University and University Hospital Motol Prague, Praha 5, Czech Republic; 3) Gennet Prague, Genetics and Reproduction Medicine Center, Praha 7, Czech Republic.

We report an extremely unusual and probably not yet reported case of uniparental isodisomy of all 23 chromosomes . A Czech patient, 12 year of age was originally tested for non-syndromic deafness. A novel mutation p.V234I in the ESRRB gene was detected in homozygous state. Subsequent testing in parents and brother revealed this mutation only in the father in heterozygous state, but not in the mother. Therefore parentity was tested by a set of microsatellite markers form chromosome 17 and X, but only paternal alleles were detected in homozygous state and maternal alleles were absent. Initially we thought on a deletion deletion, but MLPA from the CMT1A regions showed no copy number variation. Extensive testing using 36 microsatellite markers spread over the entire genome showed homozygozity in all markers from all chromosomes 23 with always only paternal allele present . SNP array confirmed the whole genome homozygozity without copy number change. DNA isolated from blood was used for all this initial testing. Since we could not believe such status could be possible in all tissues without major problems we repeated all the above testing with DNA isolated from cultivated skin fibroblasts and buccal cells. In skin and buccal DNA alleles from both parents were present - no isodisomy and no loss of heterozygozity was present. Finally we tested DNA from blood sampled from the patient 3 years after the first blood sampling and the results were identical. So we concluded this patient has a mosacism for blood cells with parental isodisomy of all chromosomes. Except the prelingual sensorineural deafness the phenotype of this girl was without abnormalities and with normal mental and motor development and status. Consequences for the health status of this girl in the future is hardly predictable, but there may be increased risk for blood malignancy. Supported by IGA MH NT 14348 and NT 11521-4.

3159T

Genotype and Family Analysis of 68 Thai families with Duchenne Muscular Dystrophy. *L. Choubtum[1], K. Taweechue[2], W. Khunin[3], S. Nujarean[3], C. Limwongse[2,4], D. Wattanasirichaigoon[3].* 1) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 3) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 4) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

Background: Duchenne muscular dystrophy (DMD; OMIM#300377) is a severe X-linked devastating disorder, resulted from mutation in the dystrophin gene which encodes the dystrophin protein. The DMD gene is located on Xp21.2-p21.1. It contains 79 coding exons and encodes a 3,685 amino acid residues-containing peptide. The aim of this study is to describe genotypic defects and family structure including positive family history, carrier status of the mother, and number of affected individuals of the families. Methods: Patients with the diagnosis of DMD based on clinical manifestation, family history and elevated levels of creatine kinase were enrolled in the study through genetics/neuromuscular clinics and network for prevention of reoccurrence for DMD as part of the Development Potentials of the Thai People Project, Faculty of Medicine Ramathibodi Hospital. Multiplex PCR and/or multiple ligation-dependent probe amplification (MLPA) were employed as the first tier genetic analysis of the DMD gene. If the result of multiplex PCR/MLPA came out negative, then it was followed by PCR-direct sequencing for all 79 coding exons and brain/muscle promoter region of the gene. Results and Conclusion: A total of 68 unrelated DMD families were enrolled in the study. Multiplex PCR/MLPA revealed 72% of the probands harboring exon(s) deletion (41) or duplication (8). Nineteen probands had sequencing analysis, 10 of whom the result are available which show family-specific including nonsense and frameshift mutations. There were positive family history in 17 (28%) families, and family data was incomplete in 8 families. There were a total of 115 affected (deceased and living) cases with the maximum affected case of 8 in one family. Thirty-one mothers of the 43 probands with negative family history were tested for carrier status, and 39% (12 out of 31) were shown to be carrier. The results of this study support the previous observation that large deletion/duplication of the DMD gene are accounting for two-thirds ad the genetic pathology of DMD. In this study, there was a lower rate of positive family history and maternal carrier for sporadic cases compared to those previously described in the literature.

3160F

Novel mutations in *EDA* gene in hypodontia and curly hair. *J.-W. Kim*[1,2], *K.-E. Lee*[1], *J. Ko*[1]. 1) Pediatric Dentistry, Seoul National University School of Dentistry, Seoul, South Korea; 2) Molecular Genetics, Seoul National University School of Dentistry, Seoul, South Korea.

Hypodontia (tooth agenesis) is the developmental absence of at least one tooth except third molar. Familial hypodontia can occur as an isolated form or as part of a genetic syndrome. Mutations in *MSX1*, *PAX9* and *AXIN2* genes has been identified in familial non-syndromic hypodontia. Ectodermal dysplasia is a group of syndromes affecting ectodermal origin tissues and comprises more than 150 different forms. Mutations in the ectodysplasin-A (*EDA*) gene have been associated with X-linked hypohidrotic ectodermal dysplasia and partial disruption of the EDA signaling pathway has been shown to cause isolated form of hypodontia. We have identified two X-linked hypodontia families, and performed mutational analysis of *EDA* gene. Mutational analysis revealed two novel *EDA* mutations: c.866G>T, p.Arg289Leu; c.1135T>G, p.Phe379Val (reference sequence NM_001399.4). These mutations were perfectly segregated with hypodontia and curly hair within each family and were not found in 150 control X-chromosomes with same ethnic background and 1000 genome project. This study broadens the mutational spectrum of *EDA* gene and understanding of X-linked recessive hypodontia with curly hair. This work was supported by grant (02-2013-0002) from the SNUDH Research Fund and by grants from the Bio & Medical Technology Development Program (2011-0027790), the Science Research Center grant to Bone Metabolism Research Center (2012-0000487) by the Korea Research Foundation Grant.

3161W

Defects in *Nek8* result in abnormal specification of developmental patterning, polycystic kidney disease, and impaired response to replication stress. *D.R. Beier[1, 2], D.K. Manning[2], M. Sergeev[3,4], S. Houghtaling[1], P. Czarnecki[3,4], M. Garnass[2], H.J.C. Choi[5], W. Goessling[2], K.A. Cimprich[5], J.V. Shah[3,4]*. 1) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 2) Brigham and Women's Hospital, Division of Genetics, Boston MA; 3) Harvard Medical School, Systems Biology, Boston MA; 4) Brigham and Women's Hospital, Renal Division, Boston MA; 5) Stanford University School of Medicine, Department of Chemical and Systems Biology, Stanford, CA.

The serine threonine kinase *Nek8* is emerging as a multifunctional protein with roles in disparate cellular functions. We discovered *Nek8* in an analysis of the juvenile cystic kidneys *(jck)* mouse model of PKD; affected mice carry a missense mutation in its presumptive regulatory domain. Missense mutations in this domain of *Nek8* have also been identified in patients with the renal cystic disease nephronophthisis (NPHP9) and in the Lewis PKD rat. We generated a *Nek8*-null allele; homozygous mutant mice die at birth and exhibit randomization of left-right asymmetry, cardiac anomalies, and glomerular kidney cysts. Ciliogenesis is intact in *Nek8*-deficient embryos and cells, but nodal ciliary signaling is perturbed as left-sided marker genes are misexpressed . We have also made *jck/Nek8-* compound heterozygotes; these mutants develop less severe cystic disease than *jck* homozygotes and suggest that the *jck* missense allele may encode a gain-of-function protein. Notably, *Nek8-/-* and *Pkd2-/-* embryonic phenotypes are strikingly similar. PC2 is expressed properly in *Nek8*-deficient embryos and cells; however, similar to cells lacking PC2, *Nek8*-depleted IMCD cells exhibit diminished calcium influx in response to fluid shear stress, which suggests *Nek8* may play a role in mediating PC2-dependent signaling. Mass spectrometry revealed *Nek8* binds *Anks6*, the gene mutated in the *cy/+* PKD rat. Morpholino mediated disruption of *Anks6* in zebrafish results in the formation of pronephric cysts. However, unlike loss of function mutations in *Nek8, Invs* and *Nphp3*, defects in L/R asymmetry were not seen. More recently, *Nek8* has been identified as a key effector of the ATR (ATM and Rad3-related) mediated replication stress response. Cells lacking *Nek8* form spontaneous DNA double-strand breaks, and *Nek8* protein physically interacts with ATR and travels with the replication fork. Of note, elevated DNA damage signaling is evident in *jck* homozygous kidneys at 3 weeks of age, well prior to when they exhibit extensive cystic disease. Thus, in addition to the well-established specific localization of *Nek8* to the Inversin compartment of the ciliary axoneme, it is also present in the nucleus and binds to chromatin. Functionally, it plays a role in mediating embryonic nodal signaling, maintenance of renal tubular integrity, and DNA damage response. It will be of interest to assess whether these activities reflect multiple independent *Nek8* activities, or are mechanistically related.

3162W

Mutation in the mouse homolog of *C5ORF42* disrupts ciliogenesis and causes cerebellar defects and other Joubert Syndrome phenotypes associated with the disruption of SHH signaling. *R. Damerla[1], C. Cui[1], G. Gabriel[1], X. Liu[1], B. Gibbs[1], R. Francis[1], Y. Li[1], B. Chatterjee[1], M. Srour[2], JL. Michaud[2], GJ. Pazour[3], CW. Lo[1].* 1) Developmental Biology, University of Pittsburgh, Pittsburgh, PA; 2) Centre of Excellence in Neurosciences of Université de Montréal and Sainte-Justine Hospital Research Center, Montreal, Quebec, Canada H3T 1C5; 3) Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

Mutations in cilia related genes are observed to cause a host of rapidly expanding genetic disorders known as ciliopathies. Two recent studies of patients in Canada and Saudi Arabia linked mutations in an uncharacterized gene *C5ORF42* with Joubert syndrome (JBTS). We report for the first time a mouse model for Joubert syndrome with a missense mutation in *2410089E03Rik*, the mouse homolog of *C5ORF42*. This mutant, named Heart Under Glass (Hug) was recovered from a forward genetic screen with ENU mutagenesis using fetal ultrasound imaging to identify congenital heart defects. Genome scanning to map the mutation followed by exome sequencing analysis identified the mutation in Hug as *2410089E03Rik* (c.T757C:pS235P). Hug mutants exhibited a host of defects characteristic of JBTS patients, including polydactyly, craniofacial anomalies, cystic kidneys, and congenital heart defects comprising of outflow tract malalignment associated with pulmonary atresia. These developmental anomalies are also reminiscent of other ciliopathies, suggesting they may arise from a ciliogenesis defect and the disruption of cilia dependent planar cell polarity pathways (PCP) and Shh signaling. Immunocytochemistry in fibroblasts derived from both the Hug mutant mice and from a JBTS patient with a *C5ORF42* mutation showed defects in ciliogenesis. This was associated with perturbation of the cilia transition zone. Shh signaling was observed to be disrupted in Hug mutant fibroblasts, but stereocilia patterning was not perturbed, suggesting cochlear PCP regulation is not affected. Detailed analysis of the brain showed abnormal cerebellar development associated with a significant reduction in vermis foliation, defects that are clinically relevant for assessing JBTS prognosis. Together, these studies showed the novel Hug mutant mouse model will be invaluable for further interrogating the role of cilia biology in the pathophysiology of JBTS.

3163W

NIH Study `Clinical and Molecular Investigations into Ciliopathies': Findings on Joubert Syndrome and Related Disorders (JSRD). *J.K. de Dios[1], D. Doherty[2], I.G. Phelps[2], T. Vilboux[1], A. Cullinane[1], D. Yildirimli[1], J. Shendure[3], B. O'Roak[3], J. Bryant[1], W. Zein[4], B. Brooks[4], R. Fischer[1], K. Daryanani[5], B. Turkbey[6], P. Choyke[6], J. Snow[7], T. Heller[8], M. Parisi[9], M. Huizing[1], W.A. Gahl[1], M. Gunay-Aygun[1]*. 1) NHGRI, NIH, Bethesda, MD; 2) Division of Developmental Medicine and Genetic Medicine, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA; 4) NEI, NIH; 5) NIH Clinical Center; 6) NCI, NIH; 7) NIMH, NIH; 8) NIDDK, NIH; 9) NICHD, NIH.

Human diseases caused by defects of the primary cilium/basal body/centrosome are a group of distinct disorders with overlapping features including hepatorenal fibrocystic disease, obesity, retinal degeneration and other functional and structural defects of the eyes and the brain, skeletal dysplasia, polydactyly, and situs inversus. JSRD is a clinically and genetically heterogenous group of ciliopathies defined based on the distinctive constellation of midbrain and hindbrain malformations that result in the `molar tooth sign' on brain imaging. Most JSRD patients display hypotonia, developmental delay, abnormal eye movements, and an abnormal respiratory pattern in infancy. Variable features include fibrocystic kidney disease, congenital hepatic fibrosis, retinal degeneration, retinal colobomas, and polydactyly. 21 genes identified to date account for approximately 50% of JSRD. The exact frequency and nature of kidney, liver, retina, brain and other organ system involvement in JSRD is not well defined. Under our ciliopathy study (www.clinicaltrials.gov, trial NCT00068224), ongoing since 2003, we have evaluated a total of 275 patients including 101 JSRD patients at the NIH Clinical Center. Enrollment criterion for JSRD was `molar tooth sign' on brain MRI. Evaluation included abdominal magnetic resonance imaging and high resolution ultrasonography, echocardiogram, comprehensive kidney and liver function tests, hormone levels, cognitive evaluations and DNA sequencing. Our cohort of 101 JSRD patients included 87 families; 10 families had 2 and 2 had 3 affected siblings. 46 families (53 %) had mutations in known JSRD genes including mutations in TMEM67 (15 patients, 12 families), C5orf42 (13, 10), CC2D2A (7, 6), AHI1 (5), CEP290 (4), NPHP1 (3, 2), TMEM216 (2), OFD1 (2), MKS1 (2) and KIF7 (1). Ages ranged from 0.6 to 36 years (9.2 + 7.6); there were 42 females and 14 adults. Kidney disease was diagnosed in 34% (34/101) and liver disease was present in 30% (30/101). Retinal degeneration was present in 34% (34/101) and 30% (30/101) had uveal colobomas. Neurocognitive functioning in JSRD ranged from average to severely impaired; the distribution suggested that genotyping may explain the variability. Identification of the underlying genetic causes of the remaining JSRD patients and further description of the full clinical spectrum of the related phenotypes will provide the groundwork for more focused studies and future therapeutic interventions.

3164W

Nephronophthisis type 3 with bone marrow fibrosis caused by NPHP3 gene mutations; expanding the clinical phenotype of a ciliopathy. *E. Leon[1], T. Biagi[1], K. Mistry[2].* 1) Genetics, Children's National Medical Center, Washington, DC; 2) Nephrology, Children's National Medical Center, Washington, DC.

Nephronophthisis (NPHP) is a recessive cystic kidney disease phenotypically related to Joubert syndrome, Meckel syndrome, and Senior-Loken syndrome. These associated conditions are thought to be united by the dysfunction of the primary cilium/basal body complex. NPHP is the most frequent genetic cause of end-stage kidney disease in children and young adults and can be caused by mutations in nine different genes encoding nephrocystins (NPHP1-8 and ALMS1). Disease recurrence has never been reported in kidneys transplanted into patients with NPHP. Mutations in the NPHP3 gene can cause two phenotypes: NPHP type 3 characterized by kidney and hepatic fibrosis and Renal-Hepatic-Pancreatic dysplasia that presents with oligohydramnios, hepatic and pancreatic cysts and fibrosis along with cardiac and brain anomalies resembling Meckel syndrome. We present a 4-year-old boy born at term to an unrelated Caucasian young couple via induced vaginal delivery due to oligohydramnios and enlarged left kidney. He was initially diagnosed with polycystic kidney disease, developed renal failure at 4 months and had a left nephrectomy that showed kidney fibrosis. He was initially treated with peritoneal dialysis followed by chronic hemodialysis. He developed pancytopenia and a bone marrow biopsy showed fibrosis. He was found to have delayed gastric emptying and severe oropharyngeal dysphagia. He developed failure to thrive and a G-tube was placed. He was found to have hepatosplenomegaly at age 2. Secondary hypertension and hypogammaglobulinemia were also noted. He had kidney and liver transplants at 3.5 years of age. His kidney transplant failed without signs of rejection or acute tubular injury. An echocardiogram showed a bicuspid aortic valve with associated dilation of the ascending aorta. A brain MRI showed normal structure. NPHP3 gene sequencing showed two previously reported pathogenic mutations c.2570+1G>T and c.406delA. He is now anuric, hemodialysis-dependent with generalized decreased muscle bulk and unable to walk. His cognition and speech are intact. This case presentation expands the clinical phenotype produced by NPHP3 gene mutations. The new unreported clinical findings are bone marrow fibrosis, persistent splenomegaly, and delayed gastric emptying. The reason of the failed renal transplant in this case continues to be unknown. The multiorgan compromise seen in this ciliopathy needs to be considered in any patient with the clinical diagnosis of NPHP.

3165W

Investigating the clinical features and genetics of idiopathic generalized epilepsy starting in mothers of babies with epilepsy. *S. Ghavimi[1], H. Azimi[2,3].* 1) Mofid Hospital, Shahid Beheshti Medical University, Tehran, Iran; 2) All Saints University School of Medicine, Dominica; 3) PsychoGenome, Ottawa, Ontario, Canada.

Objective: To investigate the clinical features and genetics of idiopathic generalized epilepsy starting in mothers of babies with epilepsy. Methods: Patients with general spike, defined as generalized seizures with spike or polyspike and wave on EEG, were studied in the setting of a first seizure clinic where an early postictal EEG record is part of the protocol. This outpatient study was conducted on 200 mothers who came to the clinic of Mofid's Hospital. Results: Of 200 mothers with an electro-clinical diagnosis of IGE, 56 (28%) were diagnosed as adult onset IGE. The seizure patterns in these 30 cases were tonic-clonic seizures + absences (8), tonic-clonic seizures + myoclonus (16), and tonic-clonic seizures alone (45). Tonic-clonic seizures were often precipitated by alcohol or sleep deprivation. The proportion of affected first and second degree relatives did not differ between the classical and adult onset IGE groups. Twenty adult onset cases were treated with sodium valproate, four with other antiepileptic drugs, and 50 were untreated. Conclusions: Babies born to mothers who are epileptic can gave an increase in chance of being epileptic themselves. Although it is still not clear how or when the epilepsy will start, it is clear that these children, just like their mothers carry the genes, or mutation required to have idiopathic generalized epilepsy . Adult onset IGE is a relatively frequent and benign disorder. Seizures are usually provoked and are easy to control. Patients in this age group may often be misdiagnosed as having non-lesional partial epilepsy. Early postictal EEG and sleep deprivation studies may improve the detection of these patients. Pedigree analysis suggests that adult onset IGE, like classical IGE, has a genetic etiology.

3166W

The role of fibrillin-1 in human mesenchyme stem cell adipogenesis. *M.R. Davis[1], C. Duffy[2], P. DeSousa[2], V. MacRae[1], K.M Summers[1].* 1) Genetics and Genomics, The Roslin Institute, The University of Edinburgh, Roslin, Midlothian , United Kingdom ,EH25 9RG; 2) MRC Centre for Regenerative Medicine, SCRM Building, The University of Edinburgh, Edinburgh bioQuarter, 5 Little France Drive, Edinburgh, EH16 4UU.

The extracellular matrix is important in maintaining the structure of connective tissues, including bone, skin, blood vessels and adipose. Fibrillin-1 is a leading component of 10 nm microfibrils, which provide strength and elasticity to connective tissues. Mutations in the fibrillin-1 gene lead to the connective tissue disorder Marfan syndrome (MFS), which affects the cardiac, ocular and musculoskeletal systems. Many patients suffering from MFS present with depletion in adipose tissue throughout their bodies, which has also been demonstrated in mouse models. Therefore it is important to further investigate the significance of fibrillin-associated microfibrils in adipose formation. This study used mesenchymal stem cells (MSCs), which are capable of differentiating into multiple connective tissue lineages, to investigate the role of fibrillin-1 in formation of adipocytes. Fibrillin expression at the mRNA and protein level was examined using fluorescent immunocytochemistry, quantitative PCR (qPCR) and bioinformatics validation utilizing various online databases. The study demonstrated the presence of fibrillin-1 microfibrils and RNA expression early in primary human MSC differentiation to the adipose lineage. Fibrillin-1 was degraded and mRNA levels decreased as differentiation proceeded. Since many MFS patients lack appropriate formation of adipose tissue, we suggest that a fibrillin matrix is necessary for the early stage of differentiation into this connective tissue lineage but not required to develop the specific differentiated state.

3167W

Reappraising the child and adult findings from Eastern Quebec kindreds: A protective-compensatory model may reconcile the genetic and the developmental findings in schizophrenia. *M. Maziade[1,2], T. Paccalet[1], M. Battaglia[1,2].* 1) Centre de recherche Institut universitaire en santé mentale de Québec, Québec, Canada; 2) Laval university, Faculty of Medicine, Québec, Canada.

Background: The neurodevelopmental, the multifactorial-oligogenic and the gene-environment diathesis models have provoked advances in schizophrenia research, yet the exact pathophysiology remains indefinable. We broadened our analysis of 20 years of findings in adults and children descending from densely affected families in the Québec population with a founder effect. The goal was to inspect the link between these family-genetic and developmental findings.

Method: 48 multigenerational families affected by schizophrenia or bipolar disorder represented a quasi-total sample of affected kindreds in the Eastern-Quebec catchment area. Among the 1274 adult family members with lifetime best-estimate diagnoses, 341 had DSM-IV schizophrenia or bipolar disorder. Young offspring of an affected parent were studied with the same clinical, physiological and cognitive measures as the adults.

Results: First, in clinical, neuropsychological and molecular genetic terms, the reported findings in these strongly familial schizophrenia patients and relatives bore a striking resemblance to those reported in sporadic samples. Second, we observed a considerable degree of heterogeneity despite the origin from a founder-effect population in the pedigrees. Third, cognitive deficits in some non-affected adult relatives were as severe as those in patients. Finally, children/adolescents descending from the kindreds displayed neurodevelopmental endophenotypic anomalies comparable to those of adult patients.

Discussion: These four observations could be reconciled under the hypothesis that highly familial and sporadic cases share mechanisms based on defective protective genes, a model to an extent similar to cancer and cardiovascular and metabolic findings. These defective protective genes running in families would longitudinally disturb the compensatory mechanisms in children inheriting them and might be at the core of the schizophrenia process.

3168W

Heritability of obesogenic growth trajectories during development in a model system. *C. Schmitt[1], S. Service[1], R. Cantor[2], A. Jasinska[1], M. Jorgensen[3], J. Kaplan[3], N. Freimer[1].* 1) Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA; 2) Department of Human Genetics, UCLA, Los Angeles, CA; 3) Department of Pathology, Section on Comparative Medicine, Wake Forest University School of Medicine, Winston-Salem, NC.

Obesity is increasingly prevalent worldwide, and has severe negative impacts on public health. Obesity arises from a complex interaction of genetic predisposition and environment that can accumulate throughout life. Although increasing evidence points to the importance of early development in the manifestations of adult disease, few studies have been undertaken of developmental measures that might be associated with adult obesity risk. The search for obesogenic markers during development in humans is complicated by the ubiquity of diets high in fat and simple carbohydrates, and the difficulty in assessing the actual diets of study subjects. This research investigates the genetic underpinnings of obesogenic growth trajectories from birth to adulthood in a genetically well-characterized model system under a controlled diet and environment: the African green monkey (*Chlorocebus aethiops sabaeus*) in the Vervet Research Colony at Wake Forest School of Medicine.

We used growth curve analysis on measures taken thrice yearly on body size and composition - body weight (BW), BMI and waist circumference (WC) - in a population of 641 monkeys measured from 2000 through 2012. 33 individuals, 6 M and 27 F, presented with signs of chronic abdominal obesity - defined as having an adult WC above 40.5 cm for at least three successive measurements. Individuals measured < 6 times were excluded from analysis (max = 30, mean = 12.6 ± 6.12). Growth was modeled using three-parameter logistic growth curves in nonlinear mixed effects models, with parameters modeled as fixed effects and subject within sex and obesity status modeled as random effects. As expected, we found a strong effect of sex and obesity status on all parameters of growth. We assessed heritability of individual growth parameters using maximum likelihood variance components analysis in SOLAR. Growth parameters were highly and significantly heritable, with sex as a significant covariate (e.g., BW: ϑ_1, asymptote of growth, $h^2 = 0.54$, $p < 0.0001$; ϑ_2, midpoint of growth, $h^2 = 0.27$, $p < 0.0001$; ϑ_3, rate of growth, $h^2 = 0.34$, $p < 0.0001$). This study suggests that adult obesity is a developmental process driven in part by heritable obesogenic trajectories resulting in faster and longer growth to larger adult size. A more detailed examination of growth trajectories may be used to assess early obesity risks and promote the discovery of novel biomedical interventions.

3169W

Functional assessment of type 2 diabetes associated loci in pancreatic β-cell specification. *E.A. Robb, N.A. Zaghloul.* University of Maryland School of Medicine, Baltimore, MD.

An estimated 24 million people in the United States are affected by the multifactorial disease, Type 2 Diabetes (T2D). Efforts to pinpoint the causative genes successfully identified 66 loci and at least as many genes contributing to T2D susceptibility using genetic association studies. However, nearly two decades later, little is still known about the precise role each identified variant plays in the disease. A key characteristic of T2D patients is the inability to maintain a functional population of pancreatic β-cells accompanied by a failure to produce and secrete sufficient insulin. In order to assess the extent to which these T2D-associated genes contribute to the generation and maintenance of β-cells, we have designed a semi-high throughput assay using the zebrafish model. Antisense oligonucleotide morpholinos, designed to individually target each T2D gene, were injected into double-transgenic embryos at the 1-2 cell stage. Expression of fluorescent mCherry (under the control of the *insulin* promoter) was utilized to specifically assess developing β-cells while overall pancreatic development was evaluated using green fluorescent protein (under the control of the *ptf1a* promoter). For each injected morpholino, we quantified the area (mass) and intensity of mCherry expression while also assessing potential effects of the morpholino on general pancreatic development by examining GFP-expressing exocrine progenitor cells. Examination of 61 T2D-associated genes (note that 7 T2D-associated genes did not have orthologs in the zebrafish), 22 genes are necessary for β-cell generation as they produce a reduction in mass and intensity; of these, 9 were previously known to affect β-cells. Four additional genes result in decreased insulin expression in β-cells. Thus, novel roles in β-cell development have been identified for several of these genes. Glucose regulation (insulin secretion and glucose level) assays were carried out for each of the 22 T2D-associated genes found to contribute to the development of the pancreatic β-cell. These genes were also investigated using whole-mount *in situ* hybridization (ISH) to identify the developmental stage of pancreatic disruption and mechanism of beta cell specification. Taken together, these data not only suggest the importance of β-cell production in genetic susceptibility to T2D, but also corroborate previous findings, supporting the use of the zebrafish model to investigate the role of T2D genes of unknown function.

3170W

Recapitulation of Metatropic Dysplasia Phenotypes in Mice Expressing Mutant TRPV4. *M.M. Weinstein[1, 2], S.W. Tompson[1], Y. Chen[3], B. Lee[3], D.H. Cohn[1, 2].* 1) Molecular, Cell, and Developmental Biology, University of California, Los Angeles; 2) Department of Orthopaedic Surgery and Orthopaedic Hospital Research Center, University of California, Los Angeles; 3) Howard Hughes Medical Institute and Department of Molecular and Human Genetics, Baylor College of Medicine, Huston, TX.

Mutations that activate the calcium channel TRPV4 have been shown to cause a spectrum of skeletal dysplasias including autosomal dominant brachyolmia, spondylometaphyseal dysplasia - Kozlowski type, and both lethal and non-lethal metatropic dysplasia. The mechanism by which these activating mutations cause growth plate dysfunction is not currently understood. In order to better understand the pathology caused by *TRPV4* activating mutations in the skeleton, we created transgenic mice overexpressing either wild-type or mutant TRPV4 in cartilage.

Mice transgenic for wild-type TRPV4 showed no morphological changes during the embryonic stage, but did have a slight delay in skeletal mineralization. Mice transgenic for the R594H TRPV4 mutant allele recapitulated defects seen in the skeletons of humans with lethal metatropic dysplasia including abnormalities of the ribcage, autopod, and vertebrae as well as displaying dumbbell-shaped long bones and perinatal death. The differences between expression of normal versus mutant TRPV4 indicate that increasing the quantity of normal TRPV4 at the cell surface is not sufficient to cause a severe skeletal dysplasia, and that a defect in the regulation of channel function, as seen in the mutant mice, is necessary for producing the abnormalities that characterize the TRPV4 skeletal dysplasias.

In addition to modeling the *TRPV4* disorders in mice, we were also able to better understand some key features of the skeletal pathology caused by TRPV4 defects. In both moderately and severely affected mutant TRPV4 transgenic mice, growth plates had enlarged reserve zones with reduced numbers of poorly organized proliferating and hypertrophic chondrocytes. In the most severely affected mutants, there were only a few hypertrophic cells with no distinct hypertrophic zone and there was a complete lack of endochondral ossification. Despite these abnormalities, the overall pattern of expression of the cartilage collagens within the growth plate was appropriate, indicating that the primary mechanism by which *TRPV4* mutations exert their effect is by inhibiting growth plate chondrocyte differentiation.

3171W

Mandibular *Prdm16* expression is critical to normal secondary palate formation in the mouse. *B.C. Bjork, L.N. Furlan, B.T. Nelson, L. Pitstick.* Dept Biochemistry, Midwestern Univ, Downers Grove, IL.

Non-syndromic orofacial clefting is one of the most common human birth defects with both multigenic and environmental influences that contribute to its etiology. Pierre Robin sequence (PRS)-type clefting is a `palate-extrinsic' mechanism of cleft secondary palate that results as an indirect consequence of an abnormally small mandible and a posteriorly-displaced tongue. Autosomal recessive ENU-induced and gene trap null loss-of-function mutant alleles of *Prdm16* model PRS, in that the mutant phenotype includes cleft palate, mandibular hypoplasia and a posteriorly-displaced tongue.

PRDM16 is a zinc finger transcription factor and known regulator of TGFβ superfamily signaling. We previously showed that TGFβ signaling is perturbed in *Prdm16 loss-of-function* mutants. *Prdm16* is critical in determining brown adipose cell fate and in the maintenance of hematopoeitic and neuronal stem cell populations. Our research aims to identify genes whose expression is altered as a consequence of *Prdm16* loss to gain a more complete understanding of the genetic, molecular and developmental mechanisms involving *Prdm16* during mouse and human palate and mandible development. To this end, we performed a qualitative assessment of the molecular consequences of *Prdm16 loss-of-function* by whole-mount *in situ* hybridization using a subset of probes specific to genes critical during normal mandible and palate development. In particular, *Goosecoid (Gsc)* exhibited reduced expression levels in the mandible and nasal prominences, and as a target of TGFβ/SMAD signaling, *Gsc* presents a strong candidate for examination as a direct transcriptional target of PRDM16.

Finally, we used the $Prdm16^{cGT}$ conditional gene trap null mutant mouse model to investigate the mechanism by which *Prdm16* loss results in clefting. We employed the *Hand2::cre* and *Osr2::cre* mouse strains to ablate *Prdm16* expression specifically in the mandible and palate shelf mesenchyme, respectively. We determined that *Prdm16* expression in the mandible, but not the palate, is required for normal palate development. These findings further demonstrate the critical role that *Prdm16* plays during mandible patterning and palate formation and validate the exploration of *PRDM16* as a candidate gene with contributions to the etiology of at least some PRS-type clefting cases.

3172W

Mouse models reveal an essential role for RERE in eye development. *B. Kim[1], O. Shchelochkov[2], M. Justice[1], B. Lee[1], D. Scott[1].* 1) Molecular & Human Genetics, Baylor College Med, Houston, TX; 2) Department of Pediatrics, University of Iowa, Iowa IA.

Microphthalmia occurs in approximately 1 out of 10,000 individuals and can be caused by alterations in genes involved in the early development of the eye. In an effort to identify novel genes involved in the development of microphthalmia, we carried out an autosomal recessive ENU screen. We found a novel mouse stain (*eyes3*) with microphthalmia. The mutation responsible for the eyes3 phenotype was mapped by linkage analysis to a region of mouse chromosome 4 that is synthetic to human chromosome 1p36.31-p36. *Rere* (arginine-glutamic acid dipeptide repeats) maps to this region and was selected as a positional candidate based on its role as a nuclear receptor co-regulator. Sequencing revealed a homozygous c.578T>C change in *Rere*, which produces a single amino acid change in a highly conserved BAH domain of RERE (p.Val193Ala). To confirm that the microphthalmia seen in the *eyes3* strain was due to a defect in *Rere*, these mice were crossed with mice carrying an RERE null allele (*om*). Homozygous *om* embryos (*Rere*$^{om/om}$) die around E9.5, but a portion *Rere*$^{om/eyes3}$ mice lived into adulthood but have microphthalmia and optic nerve atrophy. We examined the expression pattern of RERE in the developing mouse eye. At E13.5, RERE expressing cells were primarily located in the lens epithelial cells and the optic cup margin. RERE was detected in the ganglion cell layer and the lens at birth. At postnatal day 14 (P14), RERE was expressed in the outer nuclear layer (ONL), inner nuclear layer (INL), and ganglion cell layer (GCL). Histological examination revealed that the lens was not formed in *Rere*$^{om/eyes3}$ embryos at E15.5. In wild type embryos, invagination of the lens placode results in formation of the optic cup and the lens vesicle around E10.5. In *Rere*$^{om/eyes3}$ embryos, the lens vesicle is not developed due to failure of invagination of the lens placode which, consequently, leads to shallow lens pit and abnormal development of the optic cup. This is consistent with RERE's role in retinoic acid signaling which is required for reciprocal interactions between the optic vesicle and invaginating lens placode. We conclude that RERE is required for lens induction in mice. It is likely that RERE performs a similar role in humans and may contribute to the development of ocular phenotypes through its effects on retinoic acid signaling in the developing eye.

3173W

Functional and mutational analysis of long-range enhancers of ZIC3 in patients with congenital heart defects and laterality in zebrafish models. *J. Marino[1,2], S. Hook[2], P. Hu[2], R. Hart[2], E. Rossler[2], J.A. Towbin[3], J.W. Belmont[3], L. Ribeiro-Bicudo[1], M. Muenke[2].* 1) Genetics, Hospital for Rehabilitation of Craniofacial Anomaly Rua Sílvio Marchione, 3-20 Vila Universitária 17012-900 - Bauru, SP - Brasil - Caixa-postal: 1501; 2) National Institute of Health, Bethesda, MD, USA National Institutes of Health 35 Convent Drive, Bldg 35, Room 1B202 Bethesda, MD 20892-3717; 3) Baylor College of Medicine, Houston, TX, USA.

Mutations in ZIC3 frequently result in X-linked heterotaxy in humans, a syndrome consisting of left-right patterning defects, midline abnormalities, and cardiac malformations. Studies in mouse models of Zic3 dysfunction also result in heterotaxy, indicating conserved mammalian function of this developmental transcription factor. Deletion of the entire ZIC3 locus in humans, or gene dysfunction in the classical mouse mutant bent tail model, results in heterotaxy, indicating that loss-of-function is the most common pathogenic mechanism. Here we use comparative genomic alignments of non-coding elements near ZIC3 to identify and test for conserved function of potential human regulatory elements in the zebrafish model. We set out to examine the ZIC3 genomic locus by testing all of the potential enhancers and promoter elements identified by the ECR browser tool (http://ecrbase.dcode.org/). Functional elements were then subjected to Sanger sequencing in 380 control samples compared with 366 patients with congenital heart defects and/or laterality. We now describe two mutations in the ECR1 within a 521bp fragment, a C-G transition and G-C transition (both in affected males). Similarly, ECR2 showed a male specific C-A transition mutation in a 433bp functional enhancer fragment. These findings are in addition to several other mutation positive regulatory elements at the ZIC3 locus. Here we describe both the functional and mutational landscape of the laterality gene ZIC3 and provide evidence that non-coding elements will likely prove to be important in our understanding of disease pathogenesis and in future genetic testing.

3174W

Functional characterization of Gli2 in normal breast development and in breast cancer. *C. Zhao, PA. Beachy.* Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, CA, 94305.

The highly conserved Hedgehog (Hh) signaling pathway plays critical roles in embryogenesis and adult tissue homeostasis. Gli2, an essential transcriptional factor of the Hh pathway, is required for embryonic development and postnatal growth of various tissues. Germline mutations in Gli2 often lead to developmental disorders, such as holoprosencephaly or hypopituitarism; whereas somatic mutations or gene amplifications of Gli2 have been found in a growing list of malignancies, including breast cancer. Despite progress in identifying germline or somatic Gli2 mutations in ectoderm-derived tissues, such as in the skin, little is known about function of Gli2 in mammary gland development and in mammary tumors. We find that Gli2 is expressed in mammary stromal cells and its expression is differentially regulated via secreted factors from mammary epithelial cells during a mammary regenerative cycle. Importantly, genetic ablation of Gli2 function in mammary stromal cells results in a distended mammary gland with disrupted terminal end buds (TEB). Intriguingly, a subset of Gli2 mutations in human breast cancer when expressed in FACS-isolated mammary stromal cells causes elevated Hh pathway activity. We propose a paracrine model that involves epithelial-secreted factors to regulate stromal-specific Gli2 expression in normal mammary development and deregulation of this epithelial-stromal interaction contributes to mammary malignancies.

3175W

Modeling *Foxf1* deficiency and overexpression in mice. *A.V. Dharmadhikari[1, 2], B. Carofino[1,2], M.G. Hill[1], X. Ren[3], T.V. Kalin[3], J. Zabielska[4], W-Y. Wan[5], T. Majewski[6], H.B. Brown[1], A. Gambin[4], P. Szafranski[1], V.V. Kalinichenko[3], M.J. Justice[1,2], P. Stankiewicz[1,2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Program in Translational Biology & Molecular Medicine, Baylor College of Medicine, Houston, TX; 3) Division of Pulmonary Biology, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH; 4) Mossakowski Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland; 5) Dept of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX; 6) Dept of Pathology, University of Texas MD Anderson Cancer Center, Houston, TX.

Alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV) is a rare neonatally-lethal diffuse developmental disorder of the lungs caused by haploinsufficiency of *FOXF1*. All affected newborns die in the first month of life due to severe respiratory distress and pulmonary hypertension. *Foxf1* null mice die by midgestation as a result of defects in mesodermal differentiation and cell adhesion. *Foxf1* heterozygous mice exhibit up to 90% neonatal mortality, depending on genetic background. For the current study, *Foxf1* heterozygous mice with a deletion of the forkhead binding domain were generated and are congenic on the C57BL/6J background. In contrast to recent reports that *FOXF1* is incompletely paternally imprinted in the human lungs, early postnatal mortality was observed regardless of parental transmission of the deleted allele; no differences were seen in *Foxf1* expression in embryonic and postnatal lung tissues from reciprocal crosses. Analysis of RNA from postnatal day 0.5 *Foxf1*$^{+/-}$ and wildtype lungs using Illumina mouse WG-6 v2.0 expression bead chip microarray revealed statistically significant deregulation ($p<0.05$, $fdr<0.05$) of several genes, including those involved in pulmonary vascular development (*Sema3C*, *Dll4*, and *Ednrb*), lung branching morphogenesis (*Fgf10* and *Lama1*), and the blood pressure regulating renin-angiotensin system (*Ren1*, *Cma1*, and *Cpa3*). To study the effects of *Foxf1* overexpression, we knocked a Cre-inducible *Foxf1* allele into the *ROSA26 locus*. These mice have been mated to *CMV*-cre and Tie2-cre mice to obtain whole body and vascular endothelial cell specific overexpression of *Foxf1*, respectively. We are currently mating *ROSA26Foxf1*; *Tie2*-cre mice to *Foxf1*$^{+/-}$ mice, which we hypothesize will rescue early postnatal mortality. This could inform future gene therapy studies in patients with ACDMPV.

3176W

Transgenic zebrafish expressing mutant skeletal muscle actin *acta1a* genes model human nemaline myopathy. *O. Ceyhan, A.H. Beggs.* Division of Genetics and Program in Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital and Harvard Medical School, Boston MA, USA.

The nemaline myopathies (NMs) are a group of rare genetic neuromuscular disorders that tend to present at birth or infancy with moderate to severe muscle weakness and defined by an accumulation of rod-like structures (nemaline bodies) in myofibers. Heterozygous (dominant) mutations in the skeletal muscle α-actin gene (*ACTA1*) account for ~25% of all NM cases and ~50% of the severe presentations. Despite our current understanding of normal actin function, the mechanisms that lead to defects in muscle development and function in patients with *ACTA1* mutations remain unclear and there are no curative therapies available for treatment. In order to elucidate the molecular defects underlying the muscle pathology in *ACTA1*-related NM, we are generating a panel of transgenic zebrafish lines that express a series of disease-linked dominant *ACTA1* mutations on the zebrafish acta1a transcript, the predominant actin expressed in the majority of zebrafish skeletal muscle. We initially tested whether three of these variants, *acta1a H42Y*, *M134V*, and *V165M*, would lead to a muscle phenotype by injecting the mutant mRNAs into zebrafish embryos. Fish overexpressing mutant actins displayed myopathic phenotypes characterized by delayed hatching from the chorion and curved bodies to variable degrees that correspond with the severity of disease in patients with these mutations. We next generated an *acta1a V165M* transgenic line that stably expresses mutant actin in all fast myofibers, and characterized its neuromuscular phenotypes by morphological and histological analysis at different time points. The *acta1a V165M* fish display muscle weakness as evidenced by their thin bodies and curved tails at 2 days-post-fertilization (dpf) and reduced motility in touch-evoked escape response assay at 5 dpf. Whole-mount phalloidin staining at 2 dpf revealed actin aggregates in the affected fish muscle, while electron microscopy demonstrated Z-line thickening and severe myofibrillar disorganization. These results provide proof of concept that zebrafish models of actin mutations recapitulate the human disease and have robust myopathic phenotypes that may be amenable for high-throughput chemical screening.

3177W

Zebrafish *ptk7* loss-of-function mutants reveal useful genetic models for human congenital and idiopathic scoliosis. *M. Hayes[1, 2], B. Ciruna[1, 2].* 1) Program in Developmental & Stem Cell Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Ontario, Canada.

Scoliosis refers to three-dimensional curvatures of the spine and is typically broken-down into two sub-categories: congenital scoliosis (CS) that presents with vertebral anomalies associated with developmental defects, and idiopathic scoliosis (IS) diagnosed by curvature with no underlying abnormality. IS typically develops postnatally, most commonly in adolescents. Despite decades of research into the causes of scoliosis, the lack of appropriate animal models has made the investigation of possible pathological mechanisms difficult. We have generated loss-of-function *ptk7* mutant zebrafish and have observed a highly penetrant spinal curvature phenotype. Maternal-zygotic (MZ) *ptk7* mutant embryos, which represent a complete loss-of-function model, display non-canonical Wnt/PCP and canonical Wnt/β-catenin signaling defects. *Ptk7* expression in the tail bud and defects in Wnt/β-catenin-dependent mesodermal specification in MZ*ptk7* suggest a role for *ptk7* in vertebral patterning *in vivo.* Indeed, we observe vertebral malformations in MZ*ptk7* larvae that closely mimic human CS phenotypes. We suggest a role for early *ptk7* expression and Wnt/β-catenin signaling in axial development and CS pathology. Interestingly, zygotic loss of *ptk7* does not affect vertebral patterning and we do not observe vertebral malformations in mutant larvae. However, *ptk7* mutant juveniles develop severe axial curvatures that progress up until sexual maturity. We suggest that with no underlying vertebral abnormalities, *ptk7* mutant zebrafish represent a model of adolescent idiopathic scoliosis (AIS) and we are currently using these zebrafish to test possible signaling mechanisms involved in pathogenesis. Our evidence suggests that depending on the timing of loss-of-function, similar gene pathways may be associated with both CS and IS. Common pathological mechanisms may be associated with both types of abnormalities and we are using loss of *ptk7* as a model to test possible contributing factors to the human disease.

3178W

FGF Ligands Regulate Chondrocyte Differentiation in the Proximal Limb. *I.H. Hung[1,2], D.M. Ornitz[3], G.C. Schoenwolf[4], M. Lewandoski[2].* 1) Pediatrics/Medical Genetics, University of Utah, Salt Lake City, UT; 2) Cancer and Developmental Biology Lab, National Cancer Institute, Frederick, MD; 3) Dept. of Developmental Biology, Washington University School of Medicine, St. Louis, MO; 4) Dept. of Neurobiology and Anatomy, University of Utah, Salt Lake City, UT.

Activating mutations in fibroblast growth factor (FGF) receptors result in chondrodysplasia and craniosynostosis syndromes, highlighting the critical role for FGF signaling in skeletal development. Although the roles of the FGFRs in bone development have been relatively well-characterized, only two FGF ligands, FGF9 and FGF18, have been shown to regulate embryonic skeletogenesis. Our previous analyses of FGF9 and FGF18 single knockout mice suggested that these FGF ligands may be functionally redundant. To test this hypothesis and to further elucidate their roles, we have generated and analyzed the limb phenotypes of an Fgf allelic series. Here we demonstrate novel roles of these FGF ligands in chondrogenesis of the proximal limb.

3179W

The maternal polymorphism rs2236131 in ITPK1 gene is associated with neural tube defects in a high-risk Chinese population. *z. Guan, JH. Wang, J. Guo, F. Wang, XW. Wang, GN. Li, Q. Xie, X. Han, B. Niu, T. Zhang, Wang JH.* Department of biotechnology, Capital Institute of Pediatrics, beijing, Beijing, China.

Neural tube defects (NTDs) are common and severe malformations that are multifactorial, involving the combined action of both genetic and environmental factors. Inositol as maternal nutritional facts and its related genes are suggested to be implicated in Neural tube defects (NTDs), but the mechanisms are not clear. Inositol, 1,3,4-trisphosphate 5/6-kinase (ITPK1) is a key enzyme in inositol metabolism, and has been studied for gene mutation in the mouse but not single nucleotide polymorphism (SNP) in NTD-affected pregnancies. A case-control study of women with NTD-affected pregnancies (n=200) and controls (n=320) from a high-risk area for NTDs in China was carried out to investigate the association of the polymorphisms in ITPK1 gene with NTDs. The 13 tag SNPs of ITPK1 chosen based on the minor allele frequency (MAF) >20% were genotyped by the Sequenom MassArray system, and we found that 4 tag SNPs were statistically associated with NTDs ($p<0.05$). After stratifying participants by NTD phenotypes, the significant association only existed in cases with spina bifida. We predicted the binding capacity of transcription factors in the 4 tag SNPs using the bioinformatics method. Only the rs2236131 is located in the sequence transcription factors (SP-1). EMSA (electrophoretic mobility shift assay) was applied to verify the binding activity between wild and mutated oligonucleotides probes for the positive SNP (rs2236131), and showed a different allelic binding capacity of specificity protein-1 in the intron region of the ITPK1, which is affected by an G→A exchange. The RT-PCR showed that the expression decreased significantly in mutant type with rs2236131 compared with wild type in the health pregnancy ($P<0.05$). These results suggested that the maternal polymorphism rs2236131 of ITPK1 was a potential risk factor for NTDs in a high-risk area of China and the allele A of rs2236131 in ITPK1 might affect the ITPK1 gene expression level. These can supply one of the worthwhile predictor of NTDs. Foundation: National Natural Science Foundation (81070491).

3180W

SPECC1L deficiency causes neural crest cell delamination and migration defects in orofacial clefting. *I. Saadi[1], N.R. Wilson[1], A.J. Olm-Shipman[1], E. Kosa[1], D.S. Acevedo[1], K.M. Stumpff[1], G. Smith[1], L. Pitstick[2], B.C. Bjork[2], A. Czirok[1].* 1) Dept. of Anatomy & Cell Biology, University of Kansas Medical Center, Kansas City, KS; 2) Dept. of Biochemistry, Midwestern University, Downers Grove, IL.

Orofacial clefts are among the most frequent birth defects, affecting 1/800 births, in the U.S. alone. While a number of contributory genes have been identified, there is continued need to understand underlying pathogenetic mechanisms. Previously, we identified *SPECC1L* as the first gene mutated in a severe cleft that extends from the oral cavity to the eye, termed Oblique Facial Cleft (ObFC). Although less common, insights into cellular and molecular mechanisms underlying ObFC directly impact our understanding of more common facial malformations, including cleft lip. We have now created a mouse model of Specc1l deficiency using two independent genetrap alleles. Homozygous *Specc1l* mutants are embryonic lethal with defects in delamination and migration of neural crest cells (NCCs). Cranial NCCs delaminate from early embryonic neural folds and migrate to the first and second branchial arches, which give rise to majority of craniofacial structures. Specc1l is expressed in the neural folds at E8.5 (site of premigratory NCC delamination) and in the branchial arches at E10.5 (destination of migratory NCCs). In rare cases, heterozygous *Specc1l* mutant embryos show regressed facial prominences late in development, consistent with the human phenotype. *In vitro* and *in vivo* cellular analyses indicate increased actin filaments and decreased AKT signaling upon SPECC1L deficiency. SPECC1L protein is stabilized at cell-cell boundaries upon confluence and interacts with both β-catenin and E-cadherin, two canonical components of adherens junctions (AJs). Indeed, upon confluence, SPECC1L-kd cells show drastic alteration in β-catenin and E-cadherin staining that suggests *stronger* AJs. *In vitro* live-imaging analyses of wound-repair assays confirm that *SPECC1L*-kd cells move normally prior to AJ change (confluence), but significantly slower than control cells after AJ change (confluence). Lastly, AKT is proposed to directly inhibit E-cadherin levels. Thus, reduced AKT upon SPECC1L deficiency is also consistent with *stronger* AJs in kd cells. Modulation of cell-cell contacts is important not only for NCC delamination from the neural folds, but also for collective migration of NCCs to their defined destinations. We have confirmed altered β-catenin staining in epithelial and mesenchymal tissue *in vivo*. Thus, SPECC1L is an entirely novel modulator of AJ strength - the first to actually weaken AJs normally - and of AKT signaling affecting NCC function in facial morphogenesis.

3181W

Mutations in MAP3K1 tilt the balance from SOX9/FGF9 to WNT/β-catenin signaling. *J.C. Loke, A. Pearlman, H. Ostrer.* Pathology, Albert Einstein College of Medicine, Bronx, NY.

In-frame mis-sense and splicing mutations (resulting in a 2 amino acid insertion or a 34 amino acid deletion) dispersed through the MAP3K1 gene tilt the balance from the male to female sex-determining pathway, resulting in 46,XY disorder of sex development (DSD). These MAP3K1 mutations affect the balance by enhancing WNT/β-catenin/FOXL2 expression and β-catenin activity and by reducing SOX9/FGF9/FGFR2 expression. These effects are mediated at multiple levels involving MAP3K1 interaction with protein co-factors and phosphorylation of downstream targets. In primary lymphoblastoid cells and NT2/D1 cells transfected with wild type or mutant MAP3K1 cDNAs under control of the constitutive CMV promoter, these mutations increased binding of RHOA, MAP3K4, FRAT1 and AXIN1 and increased phosphorylation of p38 and ERK1/2. Overexpressing RHOA or reducing expression of MAP3K4 in NT2/D1 cells produced phenocopies of the MAP3K1 mutations. Reducing expression of RHOA or overexpressing MAP3K4 in NT2/D1 cells produced anti-phenocopies. Furthermore, the effects of the MAP3K1 mutations were rescued by co-transfection with wild type MAP3K4. Although MAP3K1 is not usually required for testis-determination, mutations in this gene can disrupt normal development through the gains of function demonstrated in this study.

3182W

Losartan increases bone mass by direct inhibition of osteoclasts. *S. Chen[1], T. Sibai[2], N. Rianon[3], T. Yang[4], J. Black[5], E. Munivez[1], T. Bertin[1], B. Dawson[1], Y. Chen[1], B. Lee[1,6].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Boston University School of Medicine Orthopaedic Surgery, Boston University, Boston, MA, USA; 3) Department of Internal Medicine, University of Texas Medical School at Houston, TX, USA; 4) Laboratory of Skeletal Biology, Center for Skeletal Disease and Tumor Metastasis, Van Andel Research Institute, Grand Rapids, MI, USA; 5) Department of Pathology, Microbiology and Immunology, Vanderbilt University, Nashville, TN, USA; 6) Howard Hughes Medical Institute, Houston, TX, USA.

Osteoporosis and hypertension are two major chronic diseases of advanced age. Although traditionally these diseases are viewed as separate entities, increasing evidence suggests an overlapping etiology. One important determinant for both hypertension and osteoporosis is renin-angiotensin signaling. Inhibition of this signaling pathway lowers blood pressure and has demonstrated potential for bone loss prevention. We investigated the molecular mechanism underlying the regulation of bone mass by the renin-angiotensin pathway through animal studies and observational human data. Wild type mice were treated with Losartan, an antihypertensive drug that inhibits the angiotensin type 1 receptor, from birth until 6-weeks of age, after which bones were collected for microCT and histomorphometric analyses. Elderly hypertensive women treated with Losartan were examined by Dual-energy X-ray absorptiometry and bone turnover markers at 6 and 12 months after the onset of treatment. Losartan increased trabecular bone volume vs. tissue volume (a 98% increase) and cortical thickness (a 9% increase) in 6-week old wild type mice. The bone changes were attributed to decreased osteoclastogenesis as demonstrated by reduced osteoclast number per bone surface in vivo and suppressed osteoclast differentiation in vitro. At the molecular level, RANKL-induced ERK1/2 phosphorylation was attenuated by Losartan, suggesting a convergence of RANKL and angiotensin signaling at the level of ERK1/2 regulation. This also suggests that Losartan acts on ERK1/2, the essential mediator of RANKL signaling, and influences osteoclast differentiation. Two women (Age 81 and 74) were prospectively treated for idiopathic hypertension with Losartan. Interestingly, both showed evidence of decreased bone resorption as measured by the urinary N-telopeptide. Altogether, inhibition of the angiotensin pathway by Losartan has beneficial effects on bone beyond reducing blood pressure. Our study adds evidence to support the relationship between angiotensin receptor blockers (Losartan) and age-related bone loss that increases the risk of fractures in the elderly.

3183W

The role of the ELOVL gene family in neurodevelopmental disorders. *J. Gerard[1], A. Moreno-De-Luca[1,2], D.W. Evans[1,2].* 1) Neuroscience Department, Bucknell University, 701 Moore Avenue, Lewisburg, PA 17837; 2) Geisinger-Bucknell Autism and Developmental Medicine Center, 120 Hamm Drive, Lewisburg, PA 17837.

Very long-chain fatty acids (VLCFAs) are essential for basic cell structure and multiple cellular functions including intercellular signaling. Fatty acid elongases, encoded by the elongation of very long-chain fatty acids (ELOVL) gene family, catalyze the first and rate-limiting step in VLCFA synthesis, condensing acyl-CoA and manoyl-CoA to produce 3-ketoacyl-CoA. Mammals have seven elongases (ELOVL1-7), each of which exhibits a characteristic pattern of expression and substrate specificity. Previous research suggests that abnormalities in the ELOVL genes may cause a wide range of disorders. Heterozygous mutations in ELOVL4 result in autosomal dominant Stargardt macular dystrophy 3, whereas homozygous mutations cause a severe neurodevelopmental disorder characterized by ichthyosis, spastic quadriplegia, and intellectual disability. Elovl3-null mice display a tousled and sparse hair coat, irritated eczematous-like skin, and increased transepidermal water loss. Elovl6 -/- mice present with obesity, hepatic steatosis, and insulin resistance. ELOVL7 overexpression is thought to play a role in prostate cancer growth. By searching through the International Standards for Cytogenomic Arrays (ISCA) Consortium database, a repository of whole genome chromosomal micro array (CMA) data from patients with clinically diagnosed neurodevelopmental disorders, we identified two individuals with overlapping deletions in chromosome 5q12.1 with a smallest region of overlap including a single gene, ELOVL7. Case 1 was referred for autism and developmental delay and found to have a 135 kb deletion including only ELOVL7. Case 2 presented with oromotor apraxia and developmental delay and has a 275 kb deletion spanning ELOVL7, DEPDC1B (DEP domain containing 1B), and ERCC8 (excision repair cross-complementing rodent repair deficiency, complementation group 8). These findings highlight the importance of the ELOVL family of genes and their potential to cause disease if disrupted. We propose that haploinsufficiency of the ELOVL7 gene increases the risk for neurodevelopmental disorders including autism.

3184W

AKT1 gene mutation levels are correlated with the type of dermatologic lesions in patients with Proteus syndrome. *M.J. Lindhurst[1], J. Wang[2], H. Bloomhardt[1], A.M. Witkowski[1], L.N. Singh[1], D.P. Bick[3], M.J. Gambello[4], C.M. Powell[5], C.R. Lee[6], T.N. Darling[2], L.G. Biesecker[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Department of Dermatology, Uniformed Services University of the Health Sciences, Bethesda, MD, USA; 3) Department of Pediatrics and Obstetrics & Gynecology, Medical College of Wisconsin, Milwaukee, WI, USA; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 5) Departments of Pediatrics and Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 6) Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD.

Proteus syndrome (PS) is characterized by progressive, mosaic, segmental overgrowth and occurs sporadically. PS is caused by a post-zygotic somatic activating mutation c.49G>A, p.Glu17Lys in *AKT1*. To date, all patients who meet the clinical diagnostic criteria for PS and have been tested in our laboratory have this mutation. Skeletal overgrowth and dermatologic lesions are the most common manifestations of PS. Cerebriform connective tissue nevi (CCTN) are a highly specific and common lesion in patients with PS and are characterized by a massively expanded dermis filled with thick collagen bundles. Epidermal nevi (EN) can occur sporadically or as part of several syndromes including PS. They have a rough surface, are dark in color, usually follow the lines of Blaschko, and exhibit hyperkeratosis, papillomatosis and acanthosis. It is unknown which cells determine the formation of these lesions. Based on the histology, we hypothesized that CCTN were generated by mutation positive cells in the dermis and that EN were generated by mutation positive cells in the epidermis. To test this hypothesis, we isolated fibroblasts and keratinocytes from CCTN, EN and apparently normal skin samples and measured the level of the mutant allele in each cell type. The mutation level in the fibroblasts isolated from seven CCTN biopsies from the feet of three patients ranged from 9-32%. The mutant allele was not found in any of the keratinocyte cultures from these samples. In four biopsies from apparently unaffected skin in two patients, the mutation level in the fibroblasts was 6-27% and there was no evidence of the mutant allele in the keratinocytes. The mutation level in fibroblast cultures isolated from nine EN samples from seven patients ranged from 0-38%, whereas in keratinocytes isolated from these EN, the mutation level was 0-44%. We conclude that the *AKT1* p.Glu17Lys activating mutation in keratinocytes is a key determinant of EN formation. The inability to detect the mutation in two of the EN keratinocyte cultures could be due to a sampling artifact or could indicate that there are multiple mechanisms that lead to EN formation. That mutant cells were found in dermal fibroblasts of both CCTN and normal-appearing skin, suggests that the presence of mutant cells in the dermis is necessary, but not sufficient, to drive the formation of CCTN.

3185W

Function of miR-199a-5p in Stage-specific Osteogenesis of Human Mesenchymal Stem Cells. *S. Gu[1], X. Chen[2], B.F. Chen[1], G. Li[1], H.W. Ouyang[2], C. Wan[1], T.L. Lee[1], W.Y. Chan[1].* 1) Rm623,LKS Integrated Biomedical Sciences Building, The Chinese University of Hong Kong, HKSAR, Hong Kong; 2) Zhejiang Provincial Key Laboratory of Tissue Engineering and Regenerative Medicine, Zhejiang University, Hangzhou, Zhejiang Province, China.

Elucidating the regulating mechanisms of osteogenesis of human mesenchymal stem cell (hMSC) is important for the development of cell therapies for bone loss and regeneration. MSC differentiation involves complex pathways that are regulated at both transcriptional and posttranscriptional levels. However, the key regulator(s) of MSC differentiation has not been identified. microRNAs (miRNAs) have been shown to regulate almost every biological process, including stem cell differentiation. Here, we show that miR-199a-5p modulates osteogenic differentiation of hMSCs. miR-199a expression detected by qPCR revealed its up-regulation during osteogenesis of hMSCs. Over-expression of miR-199a-5p but not -3p enhanced differentiation of hMSCs both *in vitro* and *in vivo*, whereas inhibition of miR-199a-5p by miR-199a-5p siRNA reduced osteogenesis of hMSCs. In order to study the underlying mechanisms of miR-199a-5p function during osteogenesis, a cyclic pathway, Twist1-miR-199a-HIF1α, was identified to have dual functions in osteogenesis both *in vitro* and *in vivo*. At early stage of differentiation, HIF1α-Twist1 pathway enhanced osteogenesis by up-regulating miR-199a-5p, while at late stage of differentiation, miR-199a-5p enhanced osteogenesis maturation by inhibiting HIF1α-Twist1 pathway. In conclusion, our findings demonstrated for the first time that Twist1-miR-199a-5p-HIF1α cyclic pathway could regulate MSC osteogenesis at different stages, which could represent a therapeutic strategy for enhancing bone formation.

3186W

DFLAT: Functional Annotation for Human Development. *H.C. Wick[1], D.P. Hill[2], H. Drabkin[2], H. Ngu[1], M. Sackman[1], C. Fournier[1,3], J. Haggett[1], J.A. Blake[2], D.W. Bianchi[4], D.K. Slonim[1,3].* 1) Department of Computer Science, Tufts University, Medford, MA; 2) Bioinformatics and Computational Biology, The Jackson Laboratory, Bar Harbor, ME; 3) Tufts University School of Medicine, Boston, MA; 4) Mother Infant Research Institute, Tufts Medical Center, Boston, MA.

Recent technological advances have led to a notable increase in genomic studies of the developing human fetus and neonate. Interpreting these studies requires widespread characterization of the functional roles of genes in different organs at appropriate developmental stages. The Gene Ontology (GO), a valuable and widely-used resource for characterizing gene function, offers perhaps the most suitable functional annotation system for this purpose. However, due in part to the difficulty of studying molecular genetic effects in humans, even the current collection of comprehensive GO annotations for human genes and gene products is often inadequate for scientists wishing to study gene function during human fetal development.

The Developmental FunctionaL Annotation at Tufts (DFLAT) project aims to improve the quality of analyses of fetal gene expression and regulation by curating human fetal gene functions using both manual and semi-automated GO procedures. Eligible annotations are then contributed to the GO database and included in GO releases of human data. In addition to augmenting GO directly, DFLAT has produced a considerable body of functional annotation that, although too preliminary for incorporation in GO, may provide valuable information about developmental genomics. This includes developmentally relevant annotations transferred from mouse orthologs to human genes. A collection of gene sets combining existing GO annotations with the 13,344 new DFLAT annotations is available for use in novel analyses. Gene set analyses of expression data on several data sets, including amniotic fluid RNA from fetuses with trisomies 21 and 18, umbilical cord blood, and newborns with bronchopulmonary dysplasia, were conducted both with and without DFLAT annotation.

The results demonstrate widespread changes with DFLAT and an increase in the number of implicated gene sets. Blinded literature review supports the validity of newly significant findings obtained with the DFLAT annotations. Newly implicated significant gene sets suggest specific hypotheses for future research. Overall, the DFLAT project contributes new functional annotation and gene sets likely to enhance our ability to interpret genomic studies of human fetal and neonatal development.

3187W

Transcriptome and pathway analysis of fetal and adult human retina, RPE, and choroid. *A.S. Boleda, M. Brooks, A. Maminishkis, S. Miller, A. Swaroop.* National Eye Institute, National Institutes of Health, Bethesda, MD.

Current knowledge of the human retina, retinal pigment epithelium (RPE), and choroid is derived from a variety of in vitro and in vivo models. Thus, functional differentiation of human tissue transcriptomes was carried out using RNA-Seq in fetal and adult tissues. By determining the expression profile of different ocular tissue types in adult and developing retina we build a foundation of knowledge for disease investigation. In this study, RNA-Seq was performed for human fetal samples (18-19th week post-conception) of three ocular tissue types (retina, RPE, and choroid) as well as for 3 adult foveal and peripheral retinal tissues (86, 87, and 92-year-old healthy eyes). Directional RNA-seq libraries (27 fetal and 6 adult) were prepared following a modified version of the standard Illumina TruSeq RNA protocol. Illumina GAIIx technology was employed for single-end sequencing to 76 nucleotides. Parallel alignments and analyses of sequencing output were performed using Genomatix software and the Tuxedo Suite. The JMP 10 package was implemented for computing principal components (PCA), creating hierarchical clusters and filtering lists of differentially expressed genes. Further pathway and network analysis was performed using GOrilla. Approximately 32 million reads per sample mapped uniquely to the Hg19 reference genome. About 28,000 transcripts were expressed at an RPKM value greater than 1.0 in any tissue, and of these between 10-15,000 transcripts showed differential expression between tissue types. PCA demonstrated that samples cluster by tissue type, and initial pathway analysis was in concordance with our predictions. Three sets of differential analysis comparisons were made. The first compared the fetal tissue types, the second compared the adult and fetal retinas, and finally we compared the adult tissue types. Genes up-regulated specifically in the fetal retina are involved in neuron differentiation, synaptic transmission and visual perception pathways. Those up-regulated in the fetal RPE are important to pigment biosynthetic processes, vitamin A metabolic processes, and light detection. Finally, up-regulated genes in fetal choroid were found to be involved in pathways for anatomical and structural morphogenesis, cellular developmental processes, response to stimuli, and multiple pathways important to immune response.

3188T

RING CHROMOSOMES ABERRATIONS AT A PEDIATRIC MEXICAN HOSPITAL. TWO CASES WITH MOSAISISM OF CHROMOSOME 13, 46XY / 46, XY, r (13) AND CHROMOSOME 18, 46, XY / 46, XY, r (18).
M. Hurtado-Hernadez[1], J.M. Aparicio-Rodriguez[2,4], M. Barrientos-Perez[3], S. Chatelain-Mercado[5]. 1) Cytogenetics; 2) Genetics; 3) Endrocrinology, Hospital para el Nino Poblano, Puebla, Puebla; 4) Estomatology, Benemerita Universidad de Puebla; 5) Biothecnology, Universidad Autónoma Metropolitana, Mexico.

The autosomic alteration due to a ring formation is a rare aberration of either chromosome 13 and 18 which is in relation with phenotipic malformations, neurologic problems and genital abnormalities. Two clinical polymalformed cases with skull treboliform dismorfies with early seizures and malformed genitals with micropenis is presented from four of the total patients found in this study. Among chromosomic alterations, the ring of autosomic chromosome 13 and 18 are not frequent, the main phenotipical alterations in this study are in relation to neurological, genital and craniofacial malformations. Taking in consideration that mutations or chromosome aberrations are alterations in the chromosome number or structure. They are mainly considered due to gametogenesis inborn error (meiosis) or during the zygote first cellular divisions. All these alterations might be observed during metaphase from the cellular cycle, where DNA loses are seen due to DNA repair processes deficiency o total absence, among others. 4617 chromosomal studies were performed at Hospital Para El Nino Poblano (Pediatric Hospital) in Mexico (from 1992 to 2011) were 34.6% (1596 patients) showed different chromosomal alterations and only two patients showed ring chromosome aberrations. These chromosome changes are classified as structural alterations. Both pediatric patients with these genetic diseases are described in this study analyzing their clinical characteristics, medical or surgical treatments according to the phenotypic alterations.

3189F

Inverted Duplication with Terminal Deletions: Variations on a Theme.
N. Christacos[1], S. Schonberg[1], C. Lau[2], J. Kelly[1], P. Mowrey[1]. 1) Dept Cytogenetics, Quest Diagnostics, Chantilly, VA; 2) Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD.

Chromosomal rearrangements resulting in an interstitial inverted duplication with associated terminal deletion, first described in 1976, have become well-established as recurrent underlying causes of genomic imbalance in man, and have been described in association with multiple chromosome regions. The application of newer molecular cytogenetic methods, in particular microarray, has allowed for a significantly higher level of resolution for detecting and characterizing these and other chromosomal abnormalities. Three mechanisms are commonly proposed to explain the origin of inverted duplications with terminal deletions, all involving formation of a dicentric chromosome intermediate that subsequently breaks during meiosis to form a monocentric duplicated and deleted chromosome. Reported cases in the literature nearly all follow a pattern of a relatively distal deletion breakpoint with end stabilization at breakpoint associated repeat sequences, likely due to ascertainment bias as others would typically result in fetal lethality. Here we report the cytogenetic results of three cases that do not follow this typical pattern and may not be completely explained by previously proposed mechanisms. Case 1 involves a derivative chromosome 14 with a proximal 14q breakpoint secondary to inverted duplication with terminal deletion and capture of the deleted region of chromosome 14 by the 7p telomeric region. Inverted duplication of chromosome 14 (14q11.2q13.1) was confirmed with FISH using the T-cell receptor LSI TRA/D probe (Abbott Molecular). As the entire deleted region was captured intact, a dicentric intermediate may not fully explain this result. Case 2 involves a derivative chromosome 15, also with a proximal long arm breakpoint secondary to inverted duplication with terminal deletion and fetal rescue by chromosome 15 nondisjunction. Case 3 has the commonly encountered inverted duplication of 8p (8p11.2p23.1) in which the inv dup (8p) chromosome was stabilized at its breakpoint by telomere capture of distal 4p resulting in duplication bands 4p15.2 to the 4p terminus. These cases illustrate how the use of newer molecular cytogenetic technologies are crucial to identifying these uncommon variations of interstitial duplication with associated terminal deletion and present a cautionary note that many such rearrangements may not follow the most common pattern described in the literature.

3190T

Infertility related to a rearrangement of the Y chromosome.
P.A. Delgado[1], S. Iyer[2], A. Jarrin[1], N. Rao[1], C.A. Tirado[1]. 1) UCLA Dept. of Pathology and Laboratory Medicine, Los Angeles, CA; 2) UT Southwestern Medical Center Dallas, Texas.

Structural abnormalities of the Y chromosome are the most frequent chromosomal aberrations in infertile male. Herein, we present a 30 year-old male with a history of infertility. A physical examination reveals no characteristics no characteristics associated with any gonosomal aneuploidy. Further studies also showed oligopermia. Chromosome analysis of peripheral blood shows what seems to be a deletion of the Y chromosome first described as 46, X, del(Y)(q11.23). C-banding was negative for the heterochromatic region normally found on the distal long arm (q) of the Y chromosome. It couldn't be determined by this analysis if part of the euchromatic band q11.23 has been deleted. FISH analysis using the SRY probe showed two signals for the **SRY** gene. Molecular genetic studies showed that there was no deletion of the AZ regions. This patient had a derivative (der) Y chromosome which results in partial monosomy of the long (q) arm and partial disomy for the short (p) arm. The ISCN was modified as: 46,X,der(Y)(-pter->q11.23::p11.2- >pter).ish der(Y)(DYZ3+,SRY++). Most of these cases are mosaic. Structural rearrangements of the Y chromosome are very common. Their influence on gonadal and somatic development is extremely variable depending on the genomic sequences duplicated or deleted and the variable degrees of mosaicism. However, rearrangements of the Y chromosome will often show a clinical picture of azoospermia or oligospermia. Further molecular genetic testing for microdeletions of the long arm of the Y chromosome was recommended. Genetic evaluation of the patient was also suggested.

3191F

A Further Case of de novo Isochromosome 18q with additional skeletal system abnormalities.
E. Karaca, T.R. Ozdemir, A. Durmaz, F. Ozkinay, O. Cogulu. Ege University Faculty of Medicine, Department of Medical Genetics, Izmir, Turkey.

Isochromosome 18q is a rare chromosomal disorder presenting a variety of phenotype ranging from mild facial dysmorphism to severe malformations. Findings in isochromosome 18q overlap with isochromosome 18p and trisomy 18, and the clinical picture has been reported from foetuses. Because very few cases have been described in the literature, the phenotypical features have not been fully well-defined because of the rarity of those cases. Main features can be summarized as holoprosencephaly, heart defects, defects of the gastrointestinal and genitourinary systems and extremity anomalies. Here, we present a new case of isochromosome 18q. The proband is a 16 months old female who was born to nonconsanguineous parents. She was first admitted to the hospital at 15 days old because of pulmonary insufficiency and has been hospitalized twice because of pulmonary infections and developmental delay. On admission her weight was 4880 g (3p), height: 53 cm (<3p), and her head circumference was 38 cm (25p). Her physical examination showed short stature, exophthalmos, right eye ptosis, blue sclera, low-set ears, long philtrum, high palate, micrognathia, pectus excavatum, short extremities, fusiform fingers, pes equinovalgus, hallux valgus and club foot. Immunological and biochemical laboratory analysis were normal. Echocardiography revealed ventricular septal defect. Visual electrophysiology showed bilateral partial conduction defects. Abdominal and cranial USG were normal. She had sensorineural hearing loss (right ear 60dB, left ear 75dB). Cranial MRI showed corpus callosum hypoplasia and dilated anterior subarachnoid space. Bone survey revealed distinctive skeletal system abnormalities such as small jaw, pectus excavatum, coxa valga, delayed ossification of hip and bilateral pes equinovalgus. Her karyotype was 46,XX, i(18)(q10). Her parents' karyotypes were normal. Skeletal system abnormalities such as pectus excavatum and coxa valga were described in trisomy 18 but not in isochromosome 18q syndrome. In conclusion our case contributes to the literature in two ways; first additional skeletal anomalies, and second reporting a further case of a very rare chromosomal disorder.

3192T

Tetrasomy 13q32.2qter Due to an Apparent Inverted Duplicated Neocentric Marker Chromosome in an Infant With Hemangiomas, Failure to Thrive, Laryngomalacia, and Tethered Cord. *J. Liu[1], M. Del Vecchio[2], D. Pezanowski[1], H. Punnett[1], J.P. de Chadarévian[1].* 1) Dept Path & Lab Med., St Chris Hosp for Children, Drexel University College of Medicine, Philadelphia, PA 19134; 2) Department of Pediatrics, Temple University School of Medicine, Philadelphia, PA 19140.

Fewer than 100 small supernumerary marker chromosomes (sSMCs) with a non-α-satellite neocentromere structure have been reported in the literature. The major morphology of a neocentric sSMC is an inverted duplication of a distal chromosome end, while ring chromosomes and centric minutes represent a small percentage. Fourteen percent were derived from segments on 13q with four breakpoints (13q14, 13q21, 13q31, 13q32) described. To date, only 5 neocentric sSMCs containing inverted duplicated segment 13q32qter have been reported, and three of the 5 cases were deceased when the manuscripts were published. Here we report a 9 week old African American male who was admitted with laryngomalacia and failure to thrive. He had multiple hemagiomas on the tip of the nose, the mid-forehead, the scalp, and the sacrum. Other abnormalities observed included micrognathia, posterior rotation of the ears, sacral dimples, tag, and cleft, tethered cord, mild splenomegaly, elevated hemoglobin/hematocrit and thrombocytopenia. He was a 34-week preemie with a normal newborn screen and hearing test. Family history was negative, consanguinity was denied. A month after the hospital visit, the proband was found dead in his crib. Chromosomal Microarray Analysis (CMA) with Affymetrix's CytoScan™ HD SNP array revealed a terminal triplication on the long arm of chromosome 13 from 13q32.2 to 13qter (tetrasomy 13q32.2qter), spanning approximately 15.6 Mb in size and encompassing 16,495 markers/probes. Reflex chromosome analysis demonstrated a small supernumerary marker chromosome (sSMC) in 95 of 100 metaphases examined that was consistent with a high percentage (95%) mosaicism karyogram. The morphology of the identified sSMC was suggestive of an inverted duplication (inv dup) of a chromosomal segment and, based on CMA findings, most likely comprises two extra copies of the segment 13q32.2qter. Reflex FISH study was consistent with the structure of an inverted duplication with a neocentromere. Genotype pattern of the involved SNPs by CMA suggests a 'U' type exchange at meiosis I. Parental cytogenetic studies were negative. We have described in this report the sixth case of an apparent neocentric sSMC containing an inverted duplicated chromosomal fragment 13q32qter characterized by genome wide SNP array, conventional cytogenetic and FISH studies. The documented clinical details will assist the medical genetics community to better understand this rare chromosomal disorder.

3193F

Prenatal diagnosis of a complex 9 break rearrangement requires karyotype, microarray and whole-genome sequencing. *M.J. Macera[1], A. Sobrino[1], B. Levy[2], V. Jobanputra[2], V. Aggarwal[2], A. Mills[1], C. Esteves[2], C-Y. Yu[3], C. Hanscom[4], V. Pillalamarri[4], M. Talkowski[4], D. Warburton[3].* 1) Department of Genetics, New York-Presbyterian Hospital, Columbia University Medical Center, New York, NY; 2) Department of Pathology and; 3) Genetics and Development and Pediatrics, Columbia University, New York, NY; 4) Center for Human Genetic Research, Massachusetts General Hospital, Department of Neurology, Harvard Medical School, Boston, MA.

Complex chromosomal translocations, defined as apparently balanced constitutional structural rearrangements involving three or more chromosomes or more than two breakpoints, are rarely detected in prenatal testing. Only 0.03% from a survey of 269371 prenatal studies were determined to be de novo (Giardino et. Al. 2009). We report a case with 5 chromosomes involved in a four way translocation and a separate two way translocation; both arms of the same chromosome 18 were involved in separate translocations. A 28 year old primi-gravida woman presented for amniocentesis sampling at 21 weeks gestation. Ultrasound and MRI revealed bilateral ventriculomegaly (13mm and 15mm) and colpocephaly, with partial agenesis of the corpus callosum. Her prior family history was unremarkable with no unusual environmental exposure. Cytogenetic and FISH analysis with telomere probes on amniocytes revealed a 46,XX,t(3;18;5;7)(p25;p11.2;q13.3;q32), t(9;18)(p22;q21) karyotype in all cells examined. SNP oligonucletide microarray analysis (SOMA) on fetal DNA showed no loss or gain of chromosomal material at any breakpoints. The pregnancy was terminated because of ultrasound findings. The unusual complex karyotype was confirmed in fetal kidney cells. Both parents had normal chromosomes. Next generation sequencing of fetal genomic DNA using large-insert jumping libraries, followed by PCR and Sanger validation identified minimal chromosomal losses and or gains. However 7 OMIM annotated genes were disrupted at the breakpoints. CNTN6 and TBC1D5 are interrupted on chromosome 3, CNTNAP2 on chromosome 7, PTPRD on chromosome 9 and L3MBTL4, LOC100130480 and WDR7 on chromosome 18. The chromosome 5 breakpoint did not involve any gene disruption. Sequencing revealed even more complexity. It was determined that small portions of chromosomes 3 and 7 were inserted into the chromosome 5 breakpoint and the p arm of chromosome 18 has a 184.5 kb inversion at the chr3/chr18 junction. This brings the total number of breaks in this chromosome complement to 9. The characterization of this extremely complex abnormality illustrates the necessity of both cytogenetic and molecular testing. G-banding coupled with telomere and painting probes detected the initial rearrangements. While microarray analysis showed no pathogenic gain or loss of material at the breakpoints of the translocations, sequencing found 7 genes disrupted by this rearrangement, as well as an even more complex chromosomal rearrangement.

3194T

Familial translocation t(4p;8p) associated with a phenotype of combined hyperlipidemia. *N. Quaresemin[1], M.L.M. Castro[1], C.H.P. Grangeiro[1], J.A. Josahkian[1], C.M. Lourenço[4], L.A.F. Laureano[2], J. Huber[1,3], E.S. Ramos[1,3], L. Martelli[1,2,3].* 1) Serviço de Genética Médica, Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto- USP, Ribeirão Preto, São Paulo, Brazil; 2) Laboratório de Citogenética - Hospital Das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 3) Departamento de Genética - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil; 4) Departamento de Neurociências e Ciências do Comportamento - Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo, Ribeirão Preto - SP, Brasil.

Combined hyperlipidemia affects around 2% of individuals and its etiology remains unknown. Lipoprotein lipase gene is located on short arm of chromosome 8 to 8p21.3interval which is considered a candidate genomic region for combined hyperlipidemia. We report four members of a family with rearrangements between chromosomes 4 and 8, two apparently balanced translocations and two derivatives of chromosome 4. The first patient (P1) was referred to Medical Genetics Division due to seizures and microcephaly. Chromosomal analysis by GTG banding showed a karyotype 46,XX, der(4)t(-4;8)(p16;p23.1)pat. The paternal karyotype was 46, XY, t(4;8)(p16;p23.1) and maternal karyotype was normal. Her lipidogram revealed hypercholesterolemia and hypertriglyceridemia. Even with a specific corrected diet, inadequate control of dyslipidemia persisted. The serum copper values were also elevated and she presented with persistent proteinuria. The 2nd patient is her half brother who presented with hypotonia at birth, failure to thrive and seizures. Interestingly, evaluation of lipids did not detect dyslipidemia or proteinuria, but showed elevated serum copper values. His karyotype was 46, XY, der(4)t(4;8)(p16;p23.1)pat. Cytogenetic investigation was performed on the extended family. Cholesterol and triglycerides levels were elevated in their father (P3), who was the carrier of the translocation. Another daughter (P4) carried the same 4;8 translocation and her lipidogram confirmed combined hyperlipidemia. The paternal grandparents both had normal karyotypes. Metabolic tests suggested an incomplete oxidation of long-chain fatty acids for P1. It has also been hypothesized defect in beta oxidation of lipids in P2. Our results indicate the need for cholesterol levels investigation in patients with aberrations involving this region of chromosome 8. In this family, overexpression of the candidate gene, 1-acylglycerol-3-phosphate-O-acyltransferase 5(AGPAT5), that maps to 8p23.1 may be correlated to the hyperlipidemia phenotype. Genomic investigation including aCGH, expression analysis and NGS are necessary for establishing a precise karyotype/phenotype/genotype correlation between this region of chromosome 8 and combined hyperlipidemia.

3195F

Mosaicism of a supernumerary marker ring chromosome six in a patient with Turner phenotype. *J. Rojas Martínez, Y.P. Guatibonza, J.C. Prieto, O.M. Moreno.* Pontificia Universidad Javeriana, Bogota, Bogota D.C, Colombia Carrera 7 No. 40 - 62.

Supernumerary marker chromosomes (SMCs) are abnormal chromosomes present in about 0.05% of the human population and in approximately 30% of carriers, an abnormal phenotype is observed. Supernumerary marker chromosomes derived from the chromosome 6(SMC 6)are rare, this anomaly has been reported in less than 1% of the cases with SMCs, associated with a phenotype highly variable related to their content of euchromatin, often involving the regions 6p21.2 to 6q11.2. We report the case of a female patient of 17 years old, with short stature, and polycystic ovary syndrome. Product from the third maternal pregnancy, non-consanguineous parents, with a history of low birth weight. On physical examination were found some minor anomalies as high nasal root, anteverted nares, flat philtrum, left palpebral ptosis, facial hirsutism, short and wide neck, low hairline in the posterior line, cubitus valgus, clinodactyly of the fifth finger in hands, hypogonadism with oligomenorrhea, along with a lack of development of secondary sexual characteristics. Renal gammagraphy shows a horseshoe kidney and the neuropsychology valuation reports a mild cognitive deficiency. Suspecting chromosomopathy, is performed a high resolution banding G Karyotype (BG) which reports: mos 47, XX, + r [61] / 48, XX, + r1 + r2 [30] / 46, XX [9], with occurrence of a mosaic with cell lines with presence of one and two small supernumerary marker ring chromosomes of unknown origin and a cell line with normal female karyotype. The array-CGH analysis revealed a gain in the 6p12.2-q12 region, with a size of 15,082 Mb, was performed high-resolution BG karyotype and FISH in the mother, which were normal, the father was not available for analysis. The partial trisomy 6p by supernumerary marker chromosomes are rare and have been reported with different phenotypes including psychomotor retardation, low birth weight and genital anomalies as in the case of this patient, as well as craniofacial, brain and limbs abnormalities, microcephaly and cardiac abnormalities. It is proposed that the phenotypic abnormalities associated with an SMC 6 are due to the presence of dosage sensitive genes in the trisomic region. Our case also presents excess genetic material in q12 region, (this region contains uncharacterized genes with respect to the phenotype) apparent in a mosaic karyotype, adding other variables in the expression of SMC 6, that could explain the Turner phenotype in this patient not previously reported.

3196T

Clinical and molecular characterization of a subtelomeric deletion of 19p13.3 including *STK11* with Peutz-Jeghers syndrome. *S. Ishimaru[1], T. Kudo[2], T. Murakoshi[2], R. Fukuzawa[3], T. Kuchikata[4], H. Yoshihashi[4].* 1) Division of Hematology and Oncology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 2) Division of Gastroenterology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 3) Division of Pathology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 4) Division of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

The subjects with a deletion of 19p have been infrequently reported. Additionally, there have been only a few studies describing full clinical features of subjects with a subtelomeric deletion of 19p on the basis of molecularly defined breakpoints. Peutz-Jeghers syndrome (PJS;MIM175200) is an autosomal dominant disorder characterized by pigmented macules on the lips and oral mucosa and multiple gastrointestinal hamartomatous polyps. The majority of cases with a clinical diagnosis of PJS are the result of disease-causing sequence variants in *STK11* (serine-threonine kinase 11) located on 19p13.3, however some are due to exons and whole-gene deletions. Here, we present the clinical manifestations on the subject with a pure subtelomeric deletion of 19p13.3 including *STK11* with PJS phenotype and review the current literatures looking for genotype-phenotype correlations. Case: The propositus was a 19-year-old female. She was delivered at term without any critical problems during the perinatal period. In childhood, she required treatment and education for congenital heart defect (ASD), seizure, repetitive middle ear disease, hearing impairment, mental retardation. At 17 years, her features included a distinctive facial appearance with relative macrocephaly, scoliosis, joint laxity, and moderate intellectual disability. The spotty pigmentation of lips and buccal mucous membrane were noted. The karyotyping was normal. The subtelomeric FISH analysis identified a cryptic deletion of 19p13.3. The oligo-array CGH (ISCA 180K manufactured by Agilent) was performed to determine the size of deletion and the precise breakpoints, which was found to be 1.1Mb(chr19: 259395-1406597 from pterminus). The deleted region encompassed at least 49 genes, including *SKT11*. Esophagogastroduodenoscopy showed multiple flat stomach polyps of varying size in the stomach. Conclusions: To our knowledge, there are a few studies reported on the subject with a pure subtelomeric deletion of 19p[Archer *et al.*(2005), Souza *et al*. (2011), Scollom *et al*. (2012)], delineating clinical features and breakpoints. Some clinical findings (distinctive facial appearance, congenital heart disease, seizures, intellectual disability) in our subject shared common manifestations with those previously reported with 19p13.3 deletion. The clinical findings and cytogenetic results found in this study support that a pure subtelomeric deletion of 19p13.3 causes the emerging contiguous gene deletion syndrome.

3197F

Association between abnormal phenotype and chromosome heteromorphisms. *J.D. Grzesiuk[1], C.S. Pereira[1], F.G.O. Gennaro[1], L.A.F. Laureano[2], S.A. Santos[1], J. Huber[2], L. Martelli[1,2].* 1) Dept of Genetics, University of São Paulo, Ribeirão Preto, SP, Brazil; 2) Clinical Hospital, School of Medicine, Ribeirão Preto, SP, Brazil.

Heteromorphisms are variations among individuals in heterochromatin regions such as inversions, duplications and deletions and they are considered benign changes. However, literature suggests its association with several disorders as infertility, cancer, mental retardation, schizophrenia and Walker-Warburg syndrome. This study aimed to describe the frequency of different chromosome heteromorphisms in patients referred for diagnostic investigation and the phenotype of the carriers. We have reviewed GTG banding karyotypes of 2,892 patients referred to the Medical Genetics Division at the Clinical Hospital (HCRP) between 2006 and 2013. Among them, 92 patients (3.18%) presented heteromorphic chromosomes, similar to the frequency of 2-5% observed by Gardner and Sutherland (2004) in the general population. This finding corroborates with the literature suggesting that heteromorphisms are not involved with phenotypic changes. However, among the heteromorphic patients, other phenotypic findings, as mental retardation and facial dysmorphisms, were described in 61 patients (66,30%), four times more frequent than the reproductive disorders phenotype diagnosed in 16 patients (17,39%). Increase in length of the heterochromatin on the long arm of chromosomes 1, 9, 16 and Y represented 29,35% of the heteromorphisms, while the satellites and stalks on the short arm of chromosomes 14,15, 21 and 22 were detected in 16,30% and chromosome 9 inversions were seen in 34,78%. We have also noticed a high rate of co-occurrence of heteromorphisms, 18 of the 92 heteromorphic patients (19,57%) presented two heteromorphisms or one heteromorphism concomitant to other type of chromosome abnormality, that reinforces the hypothesis of heteromorphism interchromosomic effect. The most frequent variants were located on chromosome 9 and the pericentric inversion of chromosome 9 was the main finding associated with positive phenotypes. The differences in size of the segments, the occurrence of two chromosomal breaks and the complexity of the inversion mechanisms could be directly related to small losses or gains of critical genetic material and consequent more severe phenotypes. Genomic studies and gene expression analysis should be used for characterization of the heteromorphisms and better understanding about its phenotypic effects.

3198T

Ring chromosome 11: Familial case with normal development.and short stature. Further delineation of this rare cytogenetic abnormality.
A. Singer[1], R. Berger[2], R. Segal[3], C. Vinkler[4]. 1) Dept Clinical Gen Unit, Barzilai Med Ctr, Tel Aviv, Israel; 2) Cytogenetic Lab. Maccabi Health Services Rehovot Israel; 3) Institute of Medical Genetics, Shaare Zedek Medical Center Jerusalem Israel; 4) Institute of Medical Genetics,Wolfson Medical Center, Holon. Israel.

Autosomal ring chromosomes are uncommon cytogenetic aberrations identified in prenatal and postnatal diagnosis. It is estimated to occur in less than 1:30,000. Rings have been reported for all chromosomes, although those involving autosomes 13 and 18 are among the most common. The classic mechanism of ring formation is breakage in both arms of a chromosome followed by fusion of the two broken arms and loss of the distal segments. Therefore, phenotypic abnormalities associated with partial deletions, can be found among patients with ring chromosomes. Only a few examples of parental transmission are known and it has been estimated that about 99% of all ring chromosomes arise sporadically. Ring chromosome 11 is rarely observed and only a few cases have been reported, almost all have growth failure, variable phenotypes and some degree of intellectual disability. We present a case of a mother and daughter with ring chromosome 11. The girl is 8 years old and was referred to our clinic for evaluation because of short stature. Her parents are healthy and nonconsanguineous. The father's height is normal and mother's height is 153 cm (3rd centile). The girl's height is 112 cm (-2.5 SD), she has three hypopigmented skin lesions, myopia, no significant dysmorphic features and normal development. Cardiac echo was normal. GTG-banded chromosome analysis on peripheral blood of the proband showed a karyotype 46, XX, r(11)(q15;p25) in all metaphases examined. Paternal chromosomes were normal and maternal chromosomes showed equal mosaic of 45XX-11/46,XX,r(11)(q15;p25). This is the first description of familial ring chromosome 11 with short stature and normal development lacking any other phenotypic abnormality. The phenotypic differences in previously reported cases are most likely due to the differences in the size of the terminal deletion(s) at one or both arms or the result of mitotic instability. Further evaluation is being done with SNP array in order to define the deleted area addressing the mild phenotype in our family. Detailed molecular characterization and documentation of chromosomal imbalances and their clinical presentation is important for accurate genotype phenotype correlations and appropriate genetic counseling.

3199F

Mosaic Turner Syndrome with unilateral absence of digital phalanges and renal agenesis.
h. ulucan, a. koparir, g. guven, a. celebi, e. koparir, m. seven, m. ozen. department of medical genetics, istanbul university cerrahpasa medical school, istanbul , Turkey.

Turner's syndrome (TS) is a sex chromosome disorder occurring in 1 in 2,500 female newborns and in approximately 50 in 100,000 adult females and is due to partial or total loss of the second sex chromosome. There is great variability in cytogenetic findings, including the 45,X karyotype, mosaics without structural abnor¬malities (as mos 45,X/46,XX and mos 45,X/ 46,XY) and structural abnormalities with or without mosaicism (as isochromosomes and marker chromosomes). The 45,X/46,XX chromosomal pattern is the most frequent mosaic type of this disease (36%). TS is characterized by retarded growth, gonadal dysgenesis and infertility. Renal and/or collecting system malformations have been found in 30-40% of cases; including horseshoe kidney, renal malrotation and collecting system malformations. Skeletal anomalies are frequent but severe hand anomalies were not reported. Here, we report five-year-old female with unilateral absence of digital phalanges and unilateral renal agenesis whose choromosomal analysis showed 45X, 46XX. TS is known to cause multisystemic disorders. Here we present a case with unilateral renal kidney which is reported as a rare anomaly. Skeletal anomalies are frequent with TS but severe hand anomalies were rare. Our patient had also absence of fingers of hand which has not been reported up to now. We discuss here; the absence of fingers in TS may be a new clinical finding or it is a sequelae of amniotic band.

3200T

A *de novo* case of 5q33.3-q34 interstitial triplication together with uniparental isodisomy of 5q34-qter.
A. Fujita[1], H. Suzumura[2], N. Harada[3], N. Matsumoto[1], N. Miyake[1]. 1) Department of Human Genetics, Yokohama City Graduate School of Medicine, Yokohama, Kanagawa, Japan; 2) Department of Pediatrics, Dokkyo Medical University, Tochigi, Japan; 3) Cytogenetic Testing Group, Molecular genetic Research and Analysis Department, Clinical Development Service Center, Mitsubishi Chemical Medience Corporation, Tokyo, Japan.

De novo triplication together with uniparental disomy (UPD) is a rare genomic rearrangement, except for those that occur on chromosome 15q11-q13. To our knowledge, co-occurrence of these two events has previously only been reported in two individuals with *de novo* intrachromosomal triplication at 11q23.3-q24.1. Here, we report the first report of a patient with *de novo* triplication together with uniparental isodisomy of chromosome 5q who with suspected karyotype of 46,XX,del(5)(q33.1q33.3),dup(5)(q31.3q33.3) or (q33.1q35.1). Our cytogenetic analysis with high resolution SNP array and FISH analysis revealed tetrasomy of 5q33.3-q34 (6.5-Mb) caused by *de novo* middle inverted triplication and uniparental isodisomy of 5q34-qter (18-Mb). Furthermore, haplotype analysis showed two informative microsatellite markers *D5S412* and *D5S2118* within the tetrasomic region indicated biparental inheritance, whereas the other two informative microsatellite markers *D5S400* and *D5S2034* within the UPD region originated from unipaternal allele inheritance. Majority of her clinical features were observed in previously reported cases of duplication overlapping with 5q33.3-q34, except the presence of hearing loss in the proband. The *FOXI1* gene, which causes autosomal recessive deafness (OMIM 600791, DFNB4) when mutated, was contained within the uniparental isodisomy region (5q34-qter). However, no mutations were identified following Sanger sequencing of *FOXI1*. As segmental isodisomy is a post-fertilization error, it is thought to have occurred during mitosis just after fertilization via a U-type exchange, while inverted duplication could have occurred during meiosis or mitosis. The final karyotype was interpreted as 46,XX,trp(pter→q34::q34→q33.3::q33.3→qter), arr 5q33.3q34(156,244,462-162,761,495)×4,5q34qter(162,761,596-180,592,321)×2 hmz. The paternal UPD was first noticed as the LOH detected by the SNP array. Thus, following only a single experiment, the SNP array can be used for genotyping as well as the detection of copy number alterations. This study reaffirms that the SNP array is a powerful tool to screen for UPD in a single experiment, especially in cases of isodisomy.

3201F

Xq28 duplication including MECP2 in 11 French affected females: what can we learn for diagnosis and genetic counselling? *S. El Chehadeh[1], R. Touraine[2], F. Prieur[2], M. Doco-Fenzy[3], T. Bienvenu[4], S. Chantot[5], C. Philippe[6], N. Marle[7], P. Callier[7], A.L. Mosca-Boidron[7], F. Mugneret[7], N. Lemeur[8], A. Goldenberg[9], P. Chambon[10], V. Satre[11], C. Coutton[11], P.S. Jouk[12], F. Devillard[12], A. Afenjar[13], M.C. Addor[14], A. Ferrarini[15], S. Lebon[16], D. Martinet[17], J.L. Alessandri[18], S. Drunat[19], C. Thauvin-Robinet[1], J. Thevenon[1], N. Perreton[20], V. Desps Portes[21], L. Faivre[1].* 1) Centre de référence Anomalies du développement et syndromes malformatifs,CHU de Dijon, Dijon, France; 2) Service de génétique clinique chromosomique et moléculaire- CHU de Saint-Etienne, Saint-Etienne, France; 3) Service de génétique, CHU de Reims, Reims, France; 4) Laboratoire de biochimie et biologie moléculaire, CHU Paris Centre, Hôpital Cochin, Paris, France; 5) Service de génétique et embryologie médicales, CHU Paris Est, Hôpital d'Enfants Armand-Trousseau, Paris, France; 6) Service de cytogénétique et génétique moléculaire, CHU de Nancy, Vandoeuvre-lès-Nancy, France; 7) Service de cytogénétique, Plateau technique de biologie, CHU de Dijon, Dijon, France; 8) Laboratoire de Cytogénétique, Etablissement français du sang - Normandie, Bois-Guillaume, France; 9) Unité de génétique clinique, CHU de Rouen, Hôpital Charles Nicolle, Rouen, France; 10) Laboratoire de cytogénétique, CHU de Rouen, Rouen, France; 11) Unité de génétique chromosomique, CHU de Grenoble, Grenoble, France; 12) Service de génétique et procréation, CHU de Grenoble, Grenoble, France; 13) Unité de génétique Clinique et neurogénétique, Hôpital d'Enfants Armand Trousseau, Paris, France; 14) Service de génétique médicale, CHUV Lausanne, Lausanne, Suisse; 15) Unité de génétique médicale et neuropédiatrie, Bellinzona, Suisse; 16) Unité de neuropédiatrie, CHUV de Lausanne, Lausanne, Suisse; 17) Laboratoire de cytogénétique constitutionnelle et prénatale, CHUV de Lausanne, Lausanne, Suisse; 18) Pôle enfants, CHU de la Réunion, Saint-Denis-de-la-Réunion, France; 19) Laboratoire de biochimie génétique, CHU Paris, Hôpital Robert debré, Paris, France; 20) CIC/EPICIME, Hospices Civils de Lyon, Groupement Hospitalier Est, Bron; 21) Service de neurologie pédiatrique, CHU de Lyon-GH Est, Bron, France.

Duplication of chromosome Xq28 including the MECP2 gene has been described primarily in male patients with severe developmental delay, spasticity, epilepsy, stereotyped hand movements and recurrent infections. Females carriers are most of the time asymptomatic and display an extreme or complete skewing of chromosome X-inactivation. Symptomatic females have been reported in 15 cases in the literature, the majority with a milder phenotype. The duplications resulted from a X-autosome translocation resulting in functional Xq disomy or were de novo or inherited intrachromosomal Xq duplications associated with a random X inactivation. The duplications were maternally inherited in only 3 cases, the others were de novo. We carried out in 2012 a national study that permitted to identify in France 11 affected females, including 2 twins, aged 4.5 to 40 years, among the 100 patients carrying a MECP2 duplication. Five were intrachromosomal duplications of Xq, 3 resulted from an unbalanced X-autosome translocation, and 2 were small size intragenic MECP2 duplications, which causality determination is on going. The size of the duplications ranged from 5.5 Kb to 11.7 Mb. The duplications were de novo in all 7 patients with available parents. Together with the literature, we showed that females with X-autosome translocations display the most severe phenotype, similar to males, while others display an unspecific phenotype. In our study 22% of affected female patients presented with spasticity, 33% with epilepsy, 66% with stereotyped hand movements, and 56% with recurrent infections. These results are of importance for genetic counselling since an abnormal phenotype in females born from carrier mothers is very rare.

3202T

Clinical features in a pediatric population due to chromosome deletions at a third level pediatric Mexican hospital in 19 years period of time: Five case reports. *S. Chatelain[1], J.M. Aparicio-Rodriguez[2,6], M.L. Hurtado-Hernandez[3], M. Barrientos-Perez[4], R. Zamudio-Meneses[5], M. Palma-Guzmán[6], H. Chavez-Ozeki[6].* 1) Biotechnology, Universidad Autónoma Metropolitana. , Mexico City; 2) Genetics,; 3) Cytogenetics; 4) Endocrinology; 5) Cardiology, Hospital Para El Niño Poblano Bolv Del Nino Poblano; 6) Estomatology, Benemérita Universidad Autónoma de Puebla, Mexico.

Chromosome aberrations are considered as alterations in the chromosome number or structure. They are mainly due to gametogenesis inborn error (meiosis) or occur during the zygote first cellular divisions where DNA repair processes are deficient. Two Wolf Syndrome patients and two criduchat patients or patients with deletion of the short arm of chromosome 4 and 5, respectively, and deletion of chromosome 9 in one patient, were observed among 4617 karyotype studies performed from 1992 to 2011, at the Hospital Para El Niño Poblano (Pediatric Hospital) in Mexico. These chromosome structural alterations or deletions at chromosomes 4, 5 and 9, observed among five patients from three different families were chosen to analyze their clinical characteristics, medical or surgical treatments and their medical evolution according to the genetic disease.

3203F

Syndromic intellectual disability in a patient with 3.5 Mb deletion at 1p13.3. *H. Cho, K. Ha, L.C. Layman, H. Kim.* OB/GYN, Georgia Regents University, Augusta, GA., United States.

1p13.3 deletion syndrome is characterized by psychomotor retardation, hearing loss, short stature, and abnormalities of the digits. Here we report on a 35-year-old Asian woman with intellectual disability, bilateral hearing loss, short stature, high arched feet, sandal gap toes, hyperactivity, seizure, low-set ears, astigmatism, ptosis, hypotonia, kyphosis, lordosis, high arched palate, and constipation. A 244K Agilent comparative genomic hybridization (CGH) revealed a deletion of at least 3.5 Mb (chr1:107,682,629-111,177,029: GRCh37/hg19) at 1p13.3 in this patient. At least 30 genes are deleted in this region and 14 genes are overlapping with one unreported case. Some genes deleted or disrupted by a balanced translocation in this region have been reported (Borg et al. 2005; Kuilenburg et al. 2009). The Netrin G1 (NTNG1) gene providing axon guidance cues, G protein signaling modulator 2 (GPSM2) regulating neuroblast self-renewal as well as four potassium channel genes -KCNA2, KCNA3, KCNA4, KCNA10- might have contributed to the psychomotor phenotype and seizure in this patient. VAV3 reported to regulate osteoclast function and bone mass might be involved in short stature and the bone phenotype of this patient. Phenotypes such as low-set ears, kyphosis, lordosis, and high arched palate have not been previously reported in the 1p13.3 deletion syndrome region, suggesting the presence of causative gene(s) in the new genomic region deleted in this patient. The candidate genes within deleted region have been described and the phenotype /genotype relationship has been discussed.

3204T

Molecular cytogenetic analysis of 8p23.1: about two Tunisian patients. *H. Ben Khelifa[1], M. Kammoun[1], I. Ben Haj Hmida[1], H. Hannechi[1], N. Soyah[2], H. Elghezal[1], A. Saad[1], S. Mougou-Zrelli[1].* 1) Laboratory of Human Cytogenetics, Molecular Biology and Biology of Reproduction, CHU Farhat Hached - Sousse-Tunisia; 2) Department of Pedriatrics, CHU Farhat Hached -Sousse- Tunisia.

To date more than 50 cases of interstitial or terminal 8p23.1 deletions have been reported. This aberration is especially prone to various genomic rearrangements mainly because of the existence of the two olfactory receptor gene clusters (REPD and REPP) which is associated with a spectrum of anomalies including especially congenital heart malformations microcephaly and mental retardation. We report on two Tunisian patients, with different clinical features but they have in common mental retardation, similar dysmorphic features, microcephaly otherwise without heart malformation. The conventional karyotype on R banded chromosomes at a resolution of 450-550 bands, revealed a terminal deletion of the short arm of chromosome 8 in only one patient. Using CGH array on a 75 Kb resolution average, we characterized this 8p deletion of 8,883,636 bp: arr 8p23.3- 8p23.1 (191,530-9,075,165) X1. The second patient carried a smaller cryptic interstitial deletion of 1,181,530 pb at 8p23.1: arr 8p23.1 (8,229,404 - 9,410,934) X1. Fluorescence in situ hybridization (FISH) analysis confirmed the deletions and the parental analysis returned normal. Taken together, these data allowed us to define a critical deletion region of 845 Kb for the major features of a deletion 8p. Indeed, the absence of cardiac malformations in these cases, often associated with the 8p deletions, could be explained by the retention of SOX7 and GATA4 genes, implicated in heart development, in both patents. Also, we suggest that loci for the microcephaly and mental retardation are within the subregion of the commonly deleted region. Indeed, the different sizes of the reported deletions could otherwise explain the non-similarity of the clinical features observed in our patients. Even though, further studies are needed to establish more clearly the role of the discussed candidate genes in carriers of 8p23.1 deletions.

3205F

Silver and Titanium dioxide Nanoparticles role in Genotoxicity. *A. Ganapathi[1], L. Koude[1], V. Mariganti[1], P. Upendram[2], Q. Hasan[2], B. Priyanka[1], R. Rohit[3], R. Devaki[1].* 1) Biochemistry, Kamineni Institute of Medical Sceinces, Hyderabad, Andhra Pradhesh, India; 2) Departments of Genetics, Kamineni Institute of Medical Sciences, Sreepuram, Nalgonda, Hyderabad, Andhra Pradhesh, India; 3) Indian Institute of Chemical Technology, Hyderabad, Andhra Pradhesh, India.

Objectives We investigated the genotoxic effects of 10-20 nm silver (Ag) and Titanium dioxide (TiO2) nanoparticles (NPs) by evaluating chromosome aberrations, polyploidy status, Apoptosis and micronucleus (MN) assay. Methods Before testing, we confirmed that the Ag-NPs were completely dispersed in the ex-perimental medium by sonication (three times in 1 minute) and filtration (0.2 µm pore size filter), and then we measured their size in a zeta potential analyzer. After that the geno-toxicity was measured. Results After incubation with the defined concentration and size of the Silver and TiO2 nanoparticles, 100 metaphases were analyzed under microscope, 9%, 12% structural aberrations and 3%, 2% numerical aberrations were observed respectively. There was no aberration found in the control sample. After incubation with the defined concentration and size of the Ag & TiO2 nanoparticles, 1000 bi-nucleated cells were analyzed under microscope, 3% and 1% of these cells were observed with micronuclei respectively. However both nanoparticles did not show any apoptosis. Conclusions All of our findings indicate that Ag-and TiO2 NPs show genotoxic effects in mammalian cell system. In addition, present study sug¬gests that the genotoxicity effect of Ag and TiO2 nanoparticles is concentration and size dependent. Keywords Chromosomal aberrations, Polyploidy status, Genotoxicity, Micronucleus assay, Nanoparticles, Apoptosis.

3206T

Mosaicism involving a normal cell line and an unbalanced structural rearrangement. *B.M. Shaw[1], S. Elrefai[1], B. Wolf[1], L. Whiteley[2], M. Strecker[3], K. Hovanes[3], M. Cankovic[2], S. Michalowski[1].* 1) Medical Genetics, Henry Ford Health System, Detroit, MI; 2) Pathology, Henry Ford Health System, Detroit, MI; 3) CombiMatrix Diagnostics, Irvine, CA.

Mosaicism for a normal cell line and a cell line with an unbalanced structural arrangement is a rare occurrence. We report a small-for-gestational age, female infant with cleft palate, low-set ears, ventriculomegaly and joint contractures born to a 28-year-old G2P2 mother. The parents are from Yemen and are reportedly non-consanguineous. The infant was intubated at birth for respiratory distress and eventually expired at 25 days of life due to respiratory complications. Chromosome analysis on cultured lymphocytes revealed a 46,XX,der(1)t(1;1)(p36.3;q25.2)[6]/46,XX[15] karyotype. Microarray analysis using a custom Illumina SNP array showed a female with a mosaic 1q25.3-q44 duplication and a 1p36.33-p36.31 deletion. These results were confirmed by both metaphase and interphase FISH analysis. These findings may be explained by one of four potential mechanisms: 1) chimerism; 2) an unbalanced zygote, loses the abnormal chromosome 1 and undergoes monosomy rescue, resulting in isodisomy for the normal chromosome 1 in a subpopulation of cells; 3) a zygote with two normal chromosome 1s and a derivative chromosome 1 is formed via 3:1 segregation from parent carrying a balanced inversion, and as a result of the abnormal cell line's instability, secondary loss of the normal chromosome 1 and the derivative chromosome 1 occurs, resulting in two different cell lines; 4) a translocation occurs between chromatids during embryogenesis, resulting in either two unbalanced cell lines (one of which was non-viable), or a normal and a balanced cell line. Variable number tandem repeat (VNTR) analysis was performed and ruled out chimerism in this patient. Monosomy rescue is considered unlikely given that no stretches of homozygosity suggestive of isodisomy were detected by SNP array and the mother has a normal karyotype. Although possible, A 3:1 segregation of chromosome 1 seems unlikely, since trisomy 1 has only once been reported in an eight-cell pre-embryo, and this configuration would also result in tetrasomy of a large portion of 1q. Based on these results, we speculate that the chromatid translocation during embryogenesis is the most likely mechanism.

3207F

High resolution copy number analysis of genes involved in gonadal differentiation in patients with disorder of sexual development. *S.A Yatsenko[1,2], F.X. Schneck[3,4], J. Fox[3], S. Madan-Khetarpal[4,5], S.F Witchel[4,6], U. Surti[1,2,7], A. Rajkovic[1,2,7].* 1) OB/GYN & Reproductive Sci, Univ Pittsburgh, Magee-Womens Hosp., Pittsburgh, PA; 2) Department of Pathology, Univ Pittsburgh, School of Medicine, Pittsburgh, PA; 3) Department of Urology; Univ Pittsburgh; 4) Children's Hospital of Pittsburgh of UPMC, PA; 5) Department of Pediatrics and Department of Medical Genetics; Univ Pittsburgh; 6) Division of Pediatric Endocrinology; Univ Pittsburgh; 7) Department of Human Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Cytogenetic and molecular genetic analyses are essential elements in clinical diagnosis and gender assignment in children with disorders of sexual development (DSD). In patients with multiple congenital anomalies the whole genome array comparative hybridization (aCGH) has become the first tier test in the clinical diagnosis. A significant number of genes have been recognized to play a role in male and female sex development, however small deletions and duplications involving an individual gene have remained beyond the detection resolution by routine aCGH. To detect copy number variations (CNVs) with the resolution of 0.5-1 kb we have constructed a custom CGH microarray with oligonucleotide probes covering 397 clinically relevant and candidate genes associated with gonadal differentiation and genitourinary tract development. We studied 37 patients diagnosed with sex-reversal or genitourinary tract abnormalities including seven XX males, seven XY females, and 23 individuals with 46,XY chromosome complement and partial testicular differentiation, ambiguous external genitalia, cryptorchidism, or hypospadias. Pathogenic CNVs involving a single gene were detected in 4 patients (~10%) with nonsyndromic abnormalities. These include heterozygous and hemizygous imbalances within the *CYP11B1*, *DAX1*, *POU3F*, and *FGF13* genes, ranging from 1 to 160 kb in size. The application of high-resolution custom DSD microarray in patients with DSD conditions provides substantial advantages in the detection of small genomic and intragenic aberrations that are beyond the resolution of currently available aCGH analysis and other molecular techniques such as FISH analysis and sequencing of the gene of interest. As the number of genes implicated in sex development continues to grow, high resolution genome-wide approaches provide valuable information for the clinical diagnosis and management of patients with nonsyndromic DSD.

3208T

22q11.2 deletion in patients with syndromic cleft lip and palate. *S. Carpeta[1], I. Zarante[2], M.C Martinez[3], O.M Moreno[2].* 1) Biological science student Master. Institute of Genetics, Pontifical Javeriana University, Bogota, Colombia; 2) Teacher Institute of Genetics, Pontifical Javeriana University, Bogota, Colombia; 3) Teacher Empowered Dentistry, Pontifical Javeriana University, Bogota, Colombia.

Craniofacial abnormalities are one of the most common features of all birth defects, orofacial disorders being the most frequent. Cleft lip and palate are congenital malformations that affect 1 of every 700 live births and arise from variable deficiencies in tissues that form the palate and upper lip during the first weeks of gestation. In 30% of cases, genetic factors are found. Craniofacial abnormalities are described in 22q11.2 microdeletion syndrome as part of the 180 clinical signs known to contribute to the phenotypic spectrum. The aim of this study was to determine the presence of chromosomal abnormalities and 22q11.2 deletions in patients with cleft lip and palate and clinical suspicion of 22q11.2 deletion syndrome in patients from ECLAMC program and various hospitals in Bogotá. The patients analyzed had cleft lip, cleft palate and cleft lip accompanied by other clinical features associated with 22q11.2 deletion syndrome, such as congenital heart disease, dysmorphic features, inmunologic abnormalities or behavioral and learning disorders. We performed high-resolution G-banding karyotype and molecular analysis of the 22q11.2 region by MLPA (SALSA MLPA KIT P250-B1 DiGeorge). We evaluated 33 patients, 18 men and 15 women. 15 patients had palate clefts, 14 had cleft lip and palate and 3 had cleft lip. None of the patients had abnormal karyotype, 5 patients had the 22q11.2 deletion detected by MLPA, 4 of them had a deletion within the typically deleted region of ~ 3 Mb, and one had a atypical distal deletion expanding between LCR D and LCR F (~ 1.8 Mb). Patients with the confirmed deletion presented in addition to the cleft lip / palate, mainly cardiac abnormalities, immune abnormalitites and abnormal facies. The patient with the atypical deletion presented with hypothyroidism, anxiety, inguinal hernia and herpes zoster. These results suggest the importance of evaluating patients with cleft lip / palate in a genetics service and applying molecular tools as MLPA for analysing the 22q11.2 region in order to establish the genetic etiology in this group of abnormalities and provide appropriate genetic counseling.

3209F

Co-occurrence of 22q11 Deletion Syndrome and HDR Syndrome. *R. Fukai[1,2], N. Ochi[3], N. Matsumoto[1], N. Miyake[1].* 1) Department of Human Genetics, Yokohama City Graduate School of Medicine, Yokohama, Kanagawa, Japan; 2) Department of Neurology and Stroke Medicine, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 3) Department of Pediatrics, Aichi Prefectural Hospital and Rehabilitation Center for Disabled Children, Dai-ni Aoitori Gakuen, Okazaki, Japan.

22q11 deletion syndrome [MIM 188400] is one of the most common chromosomal deletion syndromes and is usually caused by a 1.5-3.0 Mb deletion at chromosome 22q11.2. It is characterized by hypocalcemia resulting from hypoplasia of the parathyroid glands, hypoplasia of the thymus, and defects of the cardiac outflow tract. We encountered a Japanese boy presenting with an unusually severe phenotype of 22q11 deletion syndrome, including progressive renal failure and severe intellectual disabilities. Diagnostic testing using fluorescent in situ hybridization revealed deletion of the 22q11 region, but this did not explain the additional complications. To seek the additional genetic causing, we performed copy number analysis using whole genome single nucleotide polymorphism (SNP) assay was performed. The detection threshold of the copy number alterations was set as follows: a confidence value of 90%, 20 contiguous probes, and larger than 100 kb for duplications; and a confidence value of 89%, 20 contiguous markers, and larger than 10 kb for deletions. As a result, the *de novo* pathological deletions at 10p14 (7.4 Mb-8.9 Mb) were newly identified in the proband in addition to the *de novo* deletion at 22q11.2 (19.0 Mb-21.5 Mb). The deleted 10p14 region is the locus for hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome [MIM 146255] caused by haploinsufficiency of *GATA3*. Conclusively, these two syndromes sufficiently explain the patient's phenotype. As far as we know, this is the first report of the co-occurrence of 22q11 deletion syndrome and HDR syndrome. As the two syndromes overlap clinically, this study indicates the importance of carrying out careful clinical and genetic assessment of patients with atypical clinical phenotypes or unique complications. Unbiased genetic analysis using whole genome copy number SNP arrays is especially useful for detecting such rare double mutations.

3210T

Generation of a Custom Array for Copy Number Variation in Congenital Diaphragmatic Hernia. *M. Russell[1,2], K. Darvishi[2,3], M. Longoni[1,2], CH. Hsieh[2,3], F. High[1], A. Kashani Pour[1], A. Tracy[1], C. Coletti[1], C. Zhang[2,3], B. Pober[1,2], K. Lage[1,2], C. Lee[2,3], P. Donahoe[1,2].* 1) Dept Pediatric Surgery, Massachusetts General Hosp, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Department of Pathology, Brigham and Women's Hospital, Boston, MA.

Congenital Diaphragmatic Hernia (CDH) is a common birth defect (1/3,000 live births in the United States), characterized by a diaphragm defect often with severe lung hypoplasia. CDH is genetically heterogeneous, usually occurring in isolation, and on occasion as part of a syndrome. Recurring chromosomal imbalances or point mutations have been implicated in CDH; however, the etiology remains unknown for the majority of cases. We have used array-based comparative genomic hybridization (aCGH) to detect congenital diaphragmatic hernia (CDH) candidate loci, hypothesizing that they contain CDH-causative genes. We analyzed aCGH data on 57 CDH patients from our cohort with either isolated or complex CDH using two different platforms, Agilent 244k and Agilent 1M arrays. A total of 32 rare CNVs (defined as <5% frequency reported in the Database of Genomic Variants) were selected for follow up. We also combined our CNV findings with an additional 48 CDH-associated CNV regions reported in the literature, as well the chromosomal regions of known CDH associated or CDH candidate genes identified via different techniques in our laboratory. With these data a custom CNV 'morbid map' was created using statistical/bioinformatic integration. This 'morbid map' of CDH-associated CNVs and gene regions will be validated by designing custom arrays on 350+ CDH patients within our cohort and 1,000+ controls, to identify CNVs that are significantly associated with CDH risk. In a collaborative effort with other CDH investigators, we will replicate our findings in a second large cohort of cases and controls. Using aCGH data, we have developed a morbid map and a custom array to test statistically valid associations between rare CNVs and CDH.

3211F

Multiplex Ligation-dependent Probe Amplification (MLPA) for detection of copy number variation in early-onset glaucoma patients from the US. *K. Allen, M. Janessian, K. Linkroum, W. Abdrabou, E. DelBono, J.L. Wiggs.* Massachusetts Eye and Ear Infirmary, MA.

Early-onset glaucoma can be caused by mutations in CYP1B1, FOXC1, and PITX2. We have previously identified mutations in these genes in approximately 25% of patients from the United States using Sanger sequencing methods. The purpose of this study is to determine the frequency of copy number variations in CYP1B1, FOXC1, and PITX2 in this US clinic-based population using Multiplex Ligation-dependent Probe amplification (MLPA). MLPA testing was performed using genomic DNA samples from 50 early onset glaucoma patients using the MRC Holland SALSA MLPA probemix kits for CYP1B1 (P128-B2 Cytochrome P-450) and FOXC1 and PITX2 (P054-B1 FOXL2-TWIST1).

MLPA raw data analysis was analyzed using the MRC Holland protocol for internal quality controls. Intra-sample normalization was calculated by comparing the relative florescence intensity of each peak within a sample. Inter-sample normalization was calculated by dividing the relative peak area of each experimental sample to that of reference samples included within the same experiment. In this study, a normal dosage quotient was 0.85-1.15. A dosage quotient of 0.35-0.65 was considered a heterozygous deletion.

MLPA analysis for this study revealed heterozygous deletions in CYP1B1 (1 patient), FOXC1 (3 patients) and PITX2 (1 patient). The heterozygous CYP1B1 deletion found in this study spanned the entire coding region for this gene. Insertions or homozygous copy number variants were not identified. The results of this study show that deletions in CYP1B1, FOXC1, and PITX2 are relatively frequent mutations in this patient population and should be included as part of the genetic diagnostic evaluation.

3212T

Autoimmune Disorders as Late Onset Feature of 16p11.2 Duplications. *C. Lowther[1,9], DJ. Stavropoulos[2,3], S. Dyack[4,5], AL. Lionel[6], CR. Marshall[6], SW. Scherer[6,7], AS. Bassett[8,9].* 1) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 2) Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Laboratory Medicine and Pathology, University of Toronto, Ontario, Canada; 4) Department of Pediatrics, Dalhousie University, Halifax, Nova Scotia, Canada; 5) Izaak Walton Killam Centre, Division of Medical Genetics, Halifax, Nova Scotia, Canada; 6) The Centre for Applied Genomics and Program in Genomics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 8) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 9) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.

The use of genome-wide microarrays to detect pathogenic copy number variations (CNVs) is now routine in many postnatal settings, and there is growing interest in the aplication to prenatal diagnosis. This has placed additional pressure on the medical genetics community to delineate the variable expressivity of CNVs such as recurrent 600 kb 16p11.2 duplications to inform genetic counselling and anticipatory care. Notably, there is limited information as yet on the adult phenotype apart from the known enrichment of 16p11.2 duplications in schizophrenia (SZ). We assessed multiple relatives of probands previously identified to carry the 16p11.2 duplication from two cohorts comprised of adults diagnosed with SZ or tetralogy of Fallot. Each individual, regardless of duplication status, underwent phenotyping including comprehensive medical, psychiatric and physical assessment to identify major lifetime features and illnesses. We report on the first 19 adults from five families, 12 of whom had the 16p11.2 duplication (median age 58, range 32-88). Eight (66.6%) of these 12 had some form of neuropsychiatric illness: three with major depression and one with intellectual disability, in addition to the four ascertained with SZ. None of the seven individuals without the 16p11.2 duplication had a neuropsychiatric phenotype. With respect to the physical phenotype of the 16p11.2 duplication, two individuals had spinal abnormailities: spina bifida occulta and spinal stenosis. There were also five with autoimmune disorders including rare conditions: lichen sclerosus and Addison's disease. One individual had developed severe blepharospasm, a rare form of dystonia, in middle age. To our knowledge this is the first study to report that late onset autoimmune and movement disorders may be features of the extended adult 16p11.2 duplication phenotye. Of the ~25 genes in the 16p11.2 region, candidates for autoimmune disorders include SPN, a sialoglycoprotein found on the surface of T lymphocytes, and CORO1A, an actin regulatory gene previously associated with a reported case of severe combined immunodeficiency. Variants of PRRT2, a potential candidate gene for dystonia, have been associated with paroxysmal kinesigenic dyskinesia. Later onset features associated with pathogenic CNVs can only be observed in well phenotyped adults. Further study of genetic modifiers within and outside of the 16p11.2 region may shed light on the variable phenotypic expression.

3213F

Evaluation of newborns with multiple congenital anomalies by array-CGH: the experience of a public hospital from Southern Brazil. *M. Riegel[1,2,3], L. Dorfman[1], R. Mergener[2], K. de Souza[1], JC. Leite[3].* 1) Post Graduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil; 2) Molecular Cytogenetics Laboratory, Gene Therapy Center, Hospital de Clínicas de Porto Alegre, Porto Alegre-RS, Brazil; 3) Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre-RS, Brazil.

Aim and methods: We analyzed DNA samples isolated from individuals with multiple congenital abnormalities born in a public hospital from Southern of Brazil to identify copy number variants involving critical chromosome regions using array comparative genomic hybridization (array-CGH). Results: In a series of samples from 32 newborns with Multiple Congenital Anomalies (MCA) we identified six deletions and four duplications. We identified one subject with a previously unknown 12 Mb deletion at 4p16.3-p15.33 associated with the Wolf-Hirschhorn syndrome region and one subject with a complete trisomy of chromosome 13. Thirty of the syndromic cases had MCA without a specific clinical diagnosis. The deletions were pathogenic and associated with the phenotypes in two cases and the duplications were classified as pathogenic and associated with the phenotypes of two subjects. In two further cases, an observed genomic imbalance was classified as variation of unknown significance (VOUS) as there was insufficient evidence to conclude if the CNV was either pathogenic or benign. In general, gain involving CNVs was more common than loss, accounting for ~70% of the imbalances. Conclusion: We have shown that array-CGH analysis of DNA samples obtained from newborns with MCA of unknown etiology is an efficient method to identify candidate chromosomal loci and/or genes, complementing the genetic evaluation of newborns with MCA. Cytogenetic testing is expected to use mostly array-CGH based technology in the near future. However, the high cost associated with the cytogenetic molecular methods and the present lack of technical skills and professional experience needed for its application are major challenges for public hospitals in developing countries such as Brazil. Financial support: FIPE/HCPA 10-560; CNPq 214906/2012-4; CNPq 402012/2010-0.

3214T

Germline DNA Copy Number Variation investigation in individuals with Argyrophilic Grain Disease reveals CTNS as a plausible candidate gene. *C. Rosenberg[1], L. Kimura[1], D. Schlesinger[2], A. Gonçalves[1], P.L. Pearson[1], C.K. Suemoto[4], C. Pasqualucci[5], A.C. Krepischi[3], L.T. Grinberg[6], D. Vilella[1].* 1) Genetics Evolutionary Biology, Universidade de Sao Paulo, sao paulo, sao paulo, Brazil; 2) Israel Institute for Teaching and Research Albert Einstein,, São Paulo, Brazil; 3) National Institute of Science and Technology in Oncogenomics, AC Camargo Hospital, São Paulo, Brazil; 4) Discipline of Geriatrics, Department of Internal Medicine, University of São Paulo Medical School,São Paulo, Brazil; 5) Department of Pathology, University of São Paulo Medical School, São Paulo, Brazil; 6) Brazilian Aging Brain Study Group - LIM22, Department of Pathology, University of São Paulo Medical School, São Paulo, Brazil.

Argyrophilic grain disease (AgD) is a late-onset dementia morphologically characterized by the presence of argyrophilic grains (ArG) in neuronal processes and coiled bodies in oligodendrocytes. Several studies have confirmed that the main protein constituent of ArG is the abnormally and hyperphosphorylated tau protein and that the grains are found in both cortical and subcortical structures. It also became apparent from recent clinicopathological studies that AgD shows a significant correlation with advancing age and it might account for approximately 5% of all dementia cases. However, the cause of AgD is not known. The disease seems to be sporadic but genetic studies have failed to discover a link of AgD with a particular gene locus. An important and recent advance in human genetics was the recognition that our genome presents copy number variations (CNVs) that involve gain or losses of genomic DNA among phenotypically normal individuals. Indeed, it has become well established in the literature that CNVs represents a significant proportion of the total genetic variability in the human population. Since the discovery of the existence of CNVs it has increased the number of studies that demonstrate its critical role in the phenotypes of complex diseases. However, there are no reports of specifics CNVs related to AgD. Therefore, the aim of this work was to identify CNVs, using the array-CGH, that possibly contributes to the development of AgD. The samples were obtained from the Brain Bank of the Brazilian Aging Brain Study Group. The array-CGH analysis revealed a 315 Kb genomic imbalance in the 17p13.2 region that includes a gene involved in age- related memory deficit, the CTNS. There is a study showing that a knockout mice Ctns -/- presents a severe age-related memory deficit, which makes CTNS a good candidate gene for susceptibility to AgD.

3215F

DUF1220 Domain Copy Number Associated with Extreme Brain Size in Individuals with 1q21 Copy Number Variations. *V.B. Searles[1,2,3], J. Davis[1], S.W. Cheung[4], S.C.S Nagamani[4], J. Sikela[1,3].* 1) Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Denver, CO; 2) Medical Scientist Training Program, University of Colorado School of Medicine, Denver, CO; 3) Human Medical Genetics and Genomics Program, University of Colorado School of Medicine, Denver, CO; 4) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Background: Copy number variations (CNVs) in chromosome 1q21 are associated with recurrent neurodevelopmental disorders. DUF1220, a protein domain on 1q21, is hyper-amplified specifically in the human lineage and has been associated with an evolutionary increase in human brain size. Previous work has suggested DUF1220 copy number reduction is associated with microcephaly in individuals with 1q21 CNVs, but obtaining accurate DUF1220 copy number measurements has been limited by technological methods. The present study was designed to confirm the association between DUF1220 copy number and brain size in an expanded number of individuals with 1q21 CNVs using a novel method for accurately measuring copy number of highly duplicated sequences. Methods: DNA and clinical data from 47 pediatric individuals with 1q21 CNVs was used for genotype-phenotype analysis. 41 of these individuals had previously been analyzed by arrayCGH. We estimated brain size using fronto-occipital head circumference Z-scores (FOCZ), a standard measure used in pediatric studies. We analyzed DUF1220 copy number for two DUF1220 clades, conserved clade 1 (CON1) and human-lineage specific clade 1 (HLS1). Copy number was determined using digital-droplet PCR (ddPCR), a novel PCR-based technique. This method entails digesting genomic DNA then dispersing the digested DNA into over 12,000 oil droplets. DNA within each droplet is amplified with DUF1220-specific primers and the presence or absence of the target sequence is read via fluorescence measurements on a droplet reader. Results: Copy number of DUF1220 clades CON1 and HLS1 was associated with brain size in this pediatric population with 1q21 CNVs (CON1: $R2=0.41$, $p<0.00001$; HLS1: $R2=0.20$, $p<0.01$; total DUF1220: $p<0.0001$). This confirms previous results using arrayCGH data in a slightly larger sample size. ddPCR copy number estimates correlated with previous high-density 1q21 arrayCGH signal intensities from the same samples (HLS1: $R2=0.98$, $p=3.3e-42$; CON1: $R2=0.75$, $p=3.56e-15$). Conclusions: We confirmed previous results showing that DUF1220 copy number is associated with brain size in a pediatric population with 1q21 CNVs, and extended this association to a larger population. We also demonstrate that ddPCR is an effective technique for determining copy number of highly duplicated sequences such as DUF1220 and may be used for similar studies in future research where arrayCGH analysis or qPCR do not meet accuracy needs.

3216T

Delineation of the genomic structure in the human 2q13 region. *B. Yuan[1], P. Liu[1], K. Potamousis[2], D. Schwartz[2], J.R. Lupski[1,3,4].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Laboratory for Molecular and Computational Genomics, University of Wisconsin-Madison, Madison, WI 53706, USA; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 4) Texas Children's Hospital, University of Texas Health Science Center at Houston, Houston, TX 77030-1501, USA.

Familial juvenile nephronophthisis is an autosomal recessive kidney disorder, which is the most frequent inherited cause of chronic renal failure in children. Loss of function of the gene *NPHP1* is responsible for approximately 85% of the nephronophthisis cases. About 80% of these patients carry a large homozygous deletion including *NPHP1*. *NPHP1* maps to 2q13, a region with extremely enriched low copy repeats (LCRs). Previous studies have demonstrated that deletion of the *NPHP1* can be mediated by nonallelic homologous recombination (NAHR) between two 45kb flanking repeats. One non-pathologic inversion haplotype involving *NPHP1* region flanked by two 358kb large inverted repeats has also been reported. Moreover, extensive evidence suggest that the human 2q13 region is structurally highly polymorphic. Therefore, we hypothesize that different historical genomic rearrangements occurred within the 2q13 region and generated various structural haplotypes observed in the human population today. Individual susceptibility to the *NPHP1* deletion can be affected by the haplotypes they carry in their personal genomes. By exploring the different haplotypes from various individuals, we may glean a more detailed understanding of the correlation between disease susceptibility and structural variation mediated by LCRs. We aim to find evidence to support the various predicted structural haplotypes. By mining the data from the literature, we observe the polymorphic copy number variation of the 45kb repeats in the 2q13 region in different populations. Optical Mapping analysis of one HapMap individual reveals the inversion haplotype of *NPHP1* region as well as the deletion of one 45kb repeat. In this study, we will also use fosmid clones constructed in the Human Genome Structural Variation (HGSV) project to infer potential haplotypes. The discordant fosmid clones whose mapping cannot be explained by the 2q13 genomic structure in the reference genome imply alternative haplotypes. In total, 226 discordant fosmids in the 2q13 region are called from 17 HapMap individuals involved in HGSV project. By examining all the available discordant fosmids, we may be able to identify novel haplotypes. The various predicted haplotypes identified by this study will promote our understanding of the complex structure in the 2q13 region and other regions with similar high-level LCRs.

3217F

Designing Custom Oligo FISH Probes for the Detection of Chromosomal Rearrangements in FFPE Tissues. *A. Bergstrom Lucas[1], M. Ruvolo[1], V. Kulkarni[1], S. Chen[1], B. Mullinax[2], J. Venneri[2], J. Barboza[3], S. Happe[3], S. Fulmer-Smentek[1], M. Srinivasan[1].* 1) Genomics R&D, Agilent Technologies, Santa Clara, CA; 2) Genomics R&D, Agilent Technologies, La Jolla, CA; 3) Genomics R&D, Agilent Technologies, Cedar Creek, TX.

Cancer cells frequently contain chromosomal rearrangements that result in oncogene activation, and the genes involved in these rearrangements are increasingly being identified using molecular technologies. We have developed a new generation of fluorescently labeled *in situ* hybridization (SureFISH) probes for the detection of these rearrangements. SureFISH probes are comprised of thousands of unique long oligonucleotides that are tiled across the targeted chromosomal region avoiding non-unique portions of the genome. The oligonucleotides are synthesized using Agilent's Oligonucleotide Library Synthesis (OLS) technology. Using knowledge of translocation breakpoints, SureFISH probes are designed to detect the translocated sequences using both break-apart and dual fusion strategies. The *in silico* design methodology and *de novo* synthesis of the SureFISH probes enables the optimization of design characteristics so that each probe provides balanced signals, facilitating the detection of chromosomal rearrangements. The flexibility afforded by the SureFISH design pipeline also enables rapid probe customization. Custom designs can be generated that target almost any genomic region, allowing for the production of probes that are not possible using other methods. We demonstrate the performance of both catalog/routine and custom probes on cytological samples and tissues that have been preserved in formalin and embedded in paraffin (FFPE).

3218T

Rearrangement of MLL gene (11q23) in pediatric patients with Acute Leukemia at Hospital Civil de Guadalajara `Dr. Juan I. Menchaca'. *H.J. PIMENTEL-GUTIÉRREZ[1, 2], L. BOBADILLA-MORALES[1, 2], S. GALLEGOS-CASTORENA[1], B. GONZÁLEZ-QUEZADA[2], F. SÁNCHEZ-ZUBIETA[1], F. GÁLVEZ-GÁLVEZ[1], A. CORONA-RIVERA[1, 2].* 1) Unidad de Citogenética, Serv. Hematol. Oncol. Pediát., Div. Pediatría, Hosp. Civil "Dr. Juan I. Menchaca";; 2) Laboratorio de Citogenética, Genotoxicidad y Biomonitoreo, Inst. Genét Hum 'Dr.ECR', CUCS, Universidad de Guadalajara.

Introduction: The Mixed-Lineage Leukemia gene (MLL, 11q23) is essential for normal hematopoiesis in acute leukemias. Although MLL disruption is generally related to unfavorable patient prognosis, more than 73 different chromosomic translocations that generate fused genes involving MLL have been reported, posing a point at issue on clinical prognosis classifications. Statement of Purpose: To present cytogenetic findings that involve MLL/11q23 in pediatric patients with acute leukemia and to correlate them with treatment response. Methods Used: Unstimulated bone marrow samples of pediatric acute leukemias were cultured 24 hours. After culture chromosome harvest was done for slide preparation and Wright staining. Fluorescence in situ hybridization was done using MLL break apart probe supplied by Abbott Molecular. Summary of Results: 11 cases were studied: 2 cases of acute myelogenous leukemia (AML) M4, 5 cases of AML M5 and 3 cases of acute lymphoblastic leukemia (ALL). Rearrangements found: 5 cases with t(9;11), 2 with t(4;11), 1 with t(6;11) and 1 with del(11q23). The patients where 5 male and 6 female (one of which was an infant with t(4;11)). Central nervous system infiltration was found in 3/11 cases, while visceromegaly on 4/11. Treatment was chemotherapy-based according to leukemia type. Survival in 18% of the patients (2/11), from which 2 received bone marrow transplant. Conclusions: The t(11q23;variable) are found on 10% of ALL (70% of infants) and 3% of AML as Slany and cols. demonstrated. Five different rearrangements where found, from which t(9;11) was the most common. Poor response to treatment was found in the majority of cases. The patient with t(1;11) had good outcomes, which has been reported previously.

3219F

A derivative chromosome involving the q arms of chromosome 1 and 15 is associated with myeloproliferative neoplasm. *P. Hord, C. McCoin, C. Curtis, S. Singh.* Cytogenetics, CSI Laboratories, Alpharetta, GA.

A derivative chromosome, der(1;15), is rare however, it is recognized as a recurrent finding in myeloid disorders. To date, only 10 case studies have been reported in *Atlas of Genetics and Cytogenetics in Oncology and Haematology*, affecting patients with a median age of 47 years. The aim of this study is to report the cytogenetic profile of three new cases referred to our laboratory with clinical histories of a myeloproliferative neoplasm. Bone marrow specimens received from the patients were cultured and harvested using standard procedures. Twenty GTW-banded cells were analyzed using CytoVision (Genetix) system and the karyotypes were written according to the International System of Human Cytogenetic Nomenclature (ISCN) 2009 Standards. All of these patients were cytogenetically identified to have a derivative chromosome, possibly dicentric, consisting of the q arms of chromosomes 1 and 15, in place of one normal chromosome 15, resulting in a net gain of 1q (trisomy 1q). In each of these cases, the der(1;15) was accompanied by additional chromosomal abnormalities, indicative of clonal evolution and/or disease progression. We conclude that our cases are very similar to those reported previously and demonstrate a clear association of the der(1;15) with myeloproliferative neoplasms. The mechanisms that result in leukemogenesis or its progression are not clear. However, several factors such as rearrangements which silence genes at the chromosomal breakpoints, partial trisomy resulting in a gene dosage effect, the development of acquired uniparental disomy, and juxtapositioning of both centromeric regions may play a role. Additionally, chromosome 1 may harbor oncogenes or tumor suppressor genes that are pathogenetically relevant to both chronic and advanced phases of MPN. It is therefore important to generate a database of these less frequently seen abnormalities that may be potential markers for disease progression or targets for newer generation drugs.

3220T

Effect of methamidophos on bone marrow cells of mice in vivo: Micronucleus assay. *I. Aranha.* Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil.

The use of pesticides is still the main strategy to fight plagues in agriculture. Methamidophos is an organophosphorous pesticide largely used in the world crop due to its efficiency. The purpose of the present work was to study the effect of methamidophos on chromosomes of mice bone marrow cells in vivo, using the micronucleus assay. Animals (ICR mice) were separated into four groups. In the first group, 6 animals received methamidophos intraperitoneally during five consecutive days in a concentration equivalent to 25% of the LD50. In the 6th day, animals were sacrificed, their femurs removed, the bone marrows collected and smears were made for slides preparation. After 24h cells were stained with Giemsa Gurr (2%) and analyzed under optical microscope. As positive control, 6 animals received cyclophosphamide (50mg/mL), once and 6 animals not exposed to any drug served as negative control for the experiment. In the test group, 13353 cells were observed and 183 showed micronuclei. In the positive control group, 12037 cells were observed and 238 had micronuclei and in the negative control group, of 12156 cells observed, none had micronucleus. The chi-square test for independence showed that our results were extremely significant (P<0.0001). They suggest that methamidophos is responsible for the micronuclei observed.

3221F

Mood disorders associate with a gene-disrupting chromosomal translocation spanning three-generations. *C. Ernst[1,2], F. Jollant[2], C. Hanscom[3], M. Stone[3], I. Blumenthal[3], G. Turecki[1,2], C. Cruceanu[1], M. Talkowski[3,4].* 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Department of Psychiatry, McGill University, Montreal, Quebec, Canada; 3) Center for Human Genetic Research, Harvard Medical School, Boston, MA, USA; 4) Medical and Population Genetics program, the Broad Institute, Cambridge, MA, USA.

We identified a family that carried a translocation across three generations [46,XX,t(11;14)(p11.2;p12)]. Due to a profound history of mood and anxiety disorders associated with the anomaly, we sequenced the translocation to basepair resolution using a mate-pair library preparation followed by massively parallel sequencing. The translocation breakpoints were localized to an unmapped contig, chrUn_gl000220, which we can now locate to Chr 14p12, and to intron 3 of LRRC4C. LRRC4C is a binding partner for Netrin G1, an axon guidance molecule important in neurodevelopment. To understand if LRRC4C is important in mood disorders, we analyzed exome sequencing data from a large bipolar family study and identified additional subjects with very rare amino-acid altering mutations in this gene, suggesting that LRRC4C mutations might have a role depressed mood and anxiety. These data demonstrate the benefits of using whole-genome sequencing to detect structural variation for psychiatric disorders and for localizing unmapped contigs to the human genome.

3222T

Effectiveness of comprehensive cytogenetic investigations in the diagnosis and genetic counseling of 8,131 patients. *L. Martelli[1,2], J. Huber[2], S.A. Santos[1], L.A.F. Laureano[2], A.C. Laus[1], J.A. Squire[3], E.S. Ramos[1,2].* 1) Dept Genetics, University of Sao Paulo, Ribeirao Preto, SP, Brazil; 2) Clinical Hospital, School of Medicine, Ribeirao Preto, SP, Brazil; 3) Dept Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada.

Aberrations involving gain or loss of chromosomal segments are particularly important in clinical cytogenetics. Classical karyotyping systematically supported by companion molecular cytogenetics has been a powerful tool for genome scanning, detecting aberrations involving gains and losses of chromosomal regions, as well as intra and inter-chromosomal rearrangements. The main objective of this study was to demonstrate the importance of comprehensive chromosomal and molecular cytogenetic analysis for clinical diagnosis over the last 15 years at the Department of Medical Genetics (HCFMRP). Initially, we evaluated 8,131 karyotypes from patients, of these 7,469 were informative, corresponding to 91.85% of the samples. Chromosomal aneuploidies were detected in 1,363 karyotypes (18.25%), structural aberrations in 405 karyotypes (5.42%) and 483 karyotypes (6.47%) had normal karyotypes with variants of normality. Among the translocations, 26 were inherited and these findings were used to determine the risk of recurrence. A subset of 17 patients required extensive molecular cytogenetic analysis including SKY and FISH techniques, to evaluate supernumerary marker chromosomes. The definitive diagnosis was established for 71% of these patients. Seven markers were originated from acrocentric chromosomes - five derivatives of chromosome 15 and two from 22 -, two from chromosome 9, and one from each pair 2, 4, 10 and 18. For four cases the origin was inherited. For two cases mBand analysis provided additional characterization of the chromosomal rearrangements. In one case of reciprocal translocation (1q;11p), small deletions were excluded in 1q31 and 11p13 cytobands. In the second case, the karyotype with homozygous inversion in chromosome 12 was characterized as 46, XY, der (12) inv (12) (p11.2p12.3) inv (12) (q21.1q24.1) x2. This patient was also investigated by aCGH technique which showed genomic gains within cytoband 12p12.3 that seemed to be related to the clinical phenotype. In conclusion, our results combining classical/molecular/cytogenomics methods have established the diagnosis of chromosomal disorders in 24% of the patients with a positive phenotype and were essential to define the etiologic diagnosis of 1,768 patients, determining the genetic counseling for their families.

3223F

Characterizing the cytogenetic dose-response relationship at very low doses of gamma radiation using structural chromosomal aberrations. *I. Seth[1], M.C Joiner[2], J.D Tucker[1].* 1) Biological Sciences, Wayne State University, Detroit, MI; 2) Department of Radiation Oncology, Wayne State University, Detroit, MI.

The shape of the dose-response curve at very low doses of radiation has long been a topic of debate in the field of radiation biology. The linear-no-threshold (LNT) model is widely used to assess risks associated with low dose exposures. According to the LNT model, the risks of genetic damage increase linearly with dose without any threshold. However, lack of knowledge of the shape of the dose-response curve at low doses, where most exposures occur, is a major problem because if low dose risks are not understood, then risk estimation and establishing scientifically valid dose limits for radiation protection are not possible. The low-dose hypersensitivity (LDH) phenomenon, in which cells are especially sensitive to low doses but then show increased radio-resistance at higher doses, provides evidence of non-linearity in the low dose region. LDH is more prominent in the G2 phase of the cell cycle than the G0/G1 or S phases and others have shown that G2 phase cells with sub-threshold DNA damage enter mitosis due to failure of the ATM-dependent G2 arrest. We hypothesized that the cytogenetic dose-response relationship at very low doses is also non-linear, and hence, should not be extrapolated from the effect at high doses. To test our hypothesis, experiments were performed in which human peripheral blood lymphocytes from one normal healthy blood donor were acutely exposed to cobalt-60 gamma rays at doses 0-4 Gy, with dose groups below 1 Gy spaced at intervals of 0.1 Gy. Damage was assessed using structural chromosomal aberrations. Preliminary results indicate that the effect per unit dose is up to 6-fold higher between 0 and 0.3 Gy than between 0.4 and 1.5 Gy. These data provide the first cytogenetic evidence for the existence of LDH in human peripheral blood lymphocytes and suggest that the LNT model is not optimal for making radiation risk assessments in the low-dose region.

3224T

Double Robertsonian translocation in pregnancy loss. *S. Wenger.* Dept Pathology, West Virginia Univ, Morgantown, WV.

The incidence of balanced Robertsonian translocation is 1/1,000 with rob(13;14) being the most common at 76%, followed by rob(14;21) at 10%. Data from our laboratory during the past 26 years identified 64 cases, two of which were acquired in leukemia and one which was mosaic with a normal karyotype. The most common Robertsonian translocation was 13;14 (52%) followed by 14;21 (16%). Absent from our records were 14;22, 21;22 and 22;22. Fourteen families were identified in which eleven demonstrated maternal inheritance, two were paternal and one case had a noted family history of trisomy 14 with no further detail. It was not possible to determine the de novo rate since not all parents were evaluated for all of the Robertsonian translocations. Adverse outcomes involving trisomies related to the Robertsonian translocation included 13, 14, and 21, although one case had an additional three way reciprocal translocation and another case had a single X chromosome. Within these fourteen families, not all pregnancy losses were karyotyped, however one pregnancy loss had both trisomy 13 and 15 as the result of two copies of rob(13;15)mat. Only one other case has been reported in the literature with a duplicated Robertsonian translocation; a karyotype of 44,XX,rob(14;21)(q10;q10)matx2 in a 17 year old woman with primary amenorrhea (Kopakka et al., Int J Gynecol Obstet 2012;116:253-257). The most likely explanation for two copies of an inherited Robertsonian translocation is nondisjunction of the Robertsonian translocation in meiosis II or first mitotic division of the zygote.

3225F

Detection of Interstitial 9q34.2-q34.3 Deletion Including the COL5A1 Gene by Array-CGH in a Patient With Classical Ehlers-Danlos Syndrome and Developmental Delay. *E.C. Beltran[1,2,3], D.W. Stockton[1,2,3], M. Hicks[1,2,3], M. Hankerd[4,5], D. Schloff[4,5], M. Kristofice[4,5], S. Ebrahim[4,5].* 1) Children's Hospital of Michigan-Specialty Center, Detroit, MI; 2) Division of Genetic and Metabolic Disorders, Detroit, MI; 3) Department of Pediatrics, Wayne State University School of Medicine, Detroit, MI; 4) Department of Pathology, Wayne State University School of Medicine, Detroit, MI; 5) Detroit Medical Center University Laboratories, Cytogenetics Laboratory, Detroit, MI.

Ehlers-Danlos syndromes (EDS) are a heterogeneous group of heritable connective tissue disorders. The classic type is characterized by hyperextensible skin, atrophic cutaneous scars due to tissue fragility and joint hyperlaxity. Mutations within the COL5A1 (9q34) or COL5A2 (2q31) genes have been found in about half of patients with Classical Type EDS. We report a 13 year old female who was referred for a genetic consultation because of EDS, developmental delay with mild cognitive impairment and partial complex seizures. She was diagnosed with EDS at 8 years by a skin biopsy. Genetic workup for a unifying diagnosis identified an interstitial 9q34.2-q34.3 deletion encompassing the COL5A1 gene. Array comparative genomic hybridization (CGH) testing detected a 1.95 Mb interstitial deletion involving chromosome 9 long arm region: arr[hg18] 9q34.2q34.3 (136,405,205-138,358,850)x1. The deleted segment harbors 35 genes including COL5A1 gene implicated in Ehlers-Danlos syndrome. However, the deleted region does not include the EHMT1 gene suggested to be associated with 9q subtelormeric deletion also known as Kleefstra syndrome. Partial deletions of chromosome 9 long arm are rare, and are associated with a wide range of phenotypes depending on the size of the deleted region. A MEDLINE search for reports of chromosome 9q34.2-q34.3 deletions encompassing the COL5A1 gene did not reveal similar cases. However, there are a few reported cases with deletions that overlap this region described in individuals with variable degrees of developmental delay, cognitive impairments, facial dysmorphisms, and other congenital malformations. Deletion of chromosome 9q encompassing the COL5A1 gene in our patient is likely satisfactory explanation for her classic EDS presentation plus her neurological symptoms. The specific contribution of the other genes in the region to her neurologic symptoms is not clear. She has some phenotypic similarities to the 9q subtelomeric deletion syndrome despite excluding the EHMT1 gene.

3226T

Use of cytogenetic tools to detect oxidative damage markers in diabetes mellitus and cardiomyopathy patients. *R. Saraswathy[1], C.R.A. Moses[2], K.T. Babu[3].* 1) Biomolecules and Genetics, VIT University, Vellore, India; 2) Madras Medical College, Chennai; 3) Heartline Medical and Research centre, Vellore.

Oxidative DNA damage is known to be strongly correlated with age related diseases and cancers. Various in vitro and in vivo cytogenetic methods have been developed to assess the DNA damages in human and animals. In this study a correlation between the DNA damages in diabetes mellitus and cardiomyopathy patients due to oxidative stress developed during pathogenesis was determined. The cytogenetic tools such as chromosomal aberration analysis and cytokinesis block micronucleus cytome assays were utilized. The results revealed a significantly higher frequency of DNA damage in the patients than in the controls. The interaction between the defect in the patient's genome and various environmental contaminants they are exposed to may lead to DNA damage and genomic instability. This study will not only lead to the use of cytogenetic tools to assess the oxidative damage markers in other human diseases but develop interventional strategies for prevention.

3227F

Partial Trisomy 21 in a Patient without the Down Syndrome Phenotype. *A. Umrigar[1], T.J. Chen[2], F. Tsien[1].* 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) Hayward Genetics Center, Tulane School of Medicine, New Orleans, LA.

We present a 24-year old female who was found to have trisomy 21 by routine karyotype, but did not present with the typical Down Syndrome phenotype. The patient had short stature, strabismus, and mental deficiency, but with none of the common dysmorphic features of Down syndrome. Previous results indicated that the patient had three copies of chromosome 21, one of which appeared smaller in size. Subsequent fluorescence *in situ* hybridization (FISH) using the Down Syndrome Critical Region (DSCR) probe showed a deletion of the 21q22.11-22.2 region on the third copy of the 21 chromosomes. Microarray comparative genomic hybridization (aCGH) revealed a deletion of approximately 13.721 Mb at 21q22. Thus, it is concluded that the patient's lack of the Down Syndrome phenotype is due to this deletion. This occurrence is extremely rare, requiring both the deletion of the DSCR and a meiotic non-disjunction event of the normal chromosome. Additionally, the proband later miscarried a fetus with Potter syndrome, but we were unable to perform cytogenetic and molecular analyses of the products of conception (POC). Our case demonstrates the importance of molecular confirmation in atypical patients previously diagnosed by karyotype.

3228T

Age-related decrease of meiotic cohesins in human oocytes. *M. Tsutsumi[1], R. Fujiwara[1], H. Nishizawa[1], H. Kogo[2], H. Inagaki[1], T. Ohye[1], T. Kato[1], H. Kurahashi[1].* 1) Fujita Health University, Toyoake, Japan; 2) Gunma University, Maebashi, Japan.

Chromosomal segregation error, e.g. nondisjunction or predivision, during meiotic cell division produces aneuploid gametes. Aneuploidy of fetal chromosomes causes miscarriage or newborns with congenital birth defects such as Down syndrome (DS). It is well known that the majority of the extra chromosome of trisomy 21 in DS originates from segregation error of maternal meiosis I, and the risk of DS increases with maternal age. Also, age-related aneuploidy causes elevated risk of miscarriage in older women. However, the etiology of maternal age-related increase of segregation error remains unclear.

Meiotic cohesins have a key role in correct segregation of chromosomes in meiosis. In mammalian oocytes, the cohesion between sister chromatids is established by the cohesin complex at fetal stage, and then the oocytes become arrested at prophase I for a prolonged period until they are ovulated after the sexual maturation. However, it has been shown that the meiotic cohesin does not undergo turnover after birth in female mice. It is possible that this might cause the gradual decrease of cohesins, and then results in increased risk of age-related segregation error.

To verify this hypothesis, we examined the amount of meiotic cohesins in oocytes by quantification of immunofluorescence signal intensity on ovarian tissue sections. Oocytes were obtained from ovaries surgically resected from eight women having ovarian cancers (age range: 19-49 years) as well as those from normal C57BL/6N female mice. It was demonstrated that the immunofluorescence levels of meiosis-specific cohesin subunits, REC8 and SMC1B, decreased in women in their 40s compared to around 20 years old. Age-related decrease of meiotic cohesins was also shown in mice. In contrast, the signal levels of subunits shared by mitotic and meiotic cells, SMC3 and RAD21, did not change. These results suggest that the decrease of the meiotic cohesin subunits impairs the cohesion between sister chromatids with age to induce segregation error. To elucidate the mechanism of decrease of meiotic cohesins will contribute to develop a novel strategy to prevent age-related increase of aneuploidy.

3229F

Microcephaly and chromosomal abnormalities: review of four years study. *M. Kammoun[1], S. Dimassi[1], H. Hannachi[1], I. Bel hadj Hmida[1], N. Soyah[2], H. El Ghzel[1], A. Saad[1], M. Soumaya[1].* 1) department of cytogenetic and reproductive biology, Farhat hached, university teatching Hospital, Sousse, Tunisia; 2) Department of pediatrics, Farhat hached, university teatching Hospital, Sousse, Tunisia.

Microcephaly is clinically a heterogeneous disorder caused by several conditions including genetic defects. We analyzed cytogenetic profiles of 135 children referred for microcephaly during the period from January 2009 to December 2012. Standard karyotyping was abnormal in 14% of cases. Conventional and molecular Cytogenetic analysis using Fluorescent In Situ Hybridization and Comparative Genomic Hybridization array showed us that microcephaly is mostly related to Angelman syndrome(42%) , chromosome 18 rearrangements(26%) and Wolf Hirschorn syndrome (WHS) secondary to 4p16.3 deletion(21%). Down syndrome, Williams and Turner syndrome, 8p23.1 and 5p21 deletion, 17q23 duplication, trisomy X and Xq28 duplication were also associated to microcephaly. Microcephaly is commonly induced by autosomal recessive mutations. Nevertheless, it is associated to the majority of chromosomal rearrangements. UBE3A gene, which is involved in AS, is proved to contribute to cellular proliferation and its invalidation was shown to be significantly associated to increased cerebellar apoptosis. Microcephaly is probably related to DOK6 haploinsufficiency located at 18q 22 and shown to be involved in cortex neurite outgrowth and in neural precursor cell differentiation during cortical development. Otherwise, Xq28 duplication is an emerging syndrome that involves MECP2 gene, a transcriptional regulatory factor strongly involved in the central nervous system development. Microcephaly is frequently associated to chromosomal rearrangement. CGH array is of great importance to detect cryptic anomalies and to narrow critical regions of microcephaly and so identifying new candidate microcephaly genes.

3230T

Constitutional 560.49 kb chromosome 2p24.3 duplication including the MYCN gene identified by oligonucleotide and SNP chromosome microarray analysis in a child with multiple congenital anomalies and bilateral Wilms tumor. *M.A. Micale[1,3], B. Embrey IV[1], J.K. Macknis[1,3], C.E. Harper[2], D.J. Aughton[2,3].* 1) Dept Pathology and Laboratory Medicine, Beaumont Health System, Royal Oak, MI; 2) Beaumont Children's Hospital, Royal Oak, MI; 3) Oakland University William Beaumont School of Medicine, Rochester, MI.

Less than 100 patients with partial chromosome 2p trisomy have been reported.Clinical features are variable and depend on the size of the duplicated segment, but generally include psychomotor delay, facial anomalies, congenital heart defect, and other abnormalities. A subset of patients has also developed neuroblastoma. We report a 560.49 kb duplication of chromosome 2p in a 13 month-old male with hydrocephaly, ventricular septal defect, suspected partial agenesis of the corpus callosum, and bilateral Wilms' tumor. He underwent neoadjuvant chemotherapy followed by right radical nephrectomy that revealed triphasic Wilms' tumor with favorable histology and skeletal muscle differentiation. A needle core biopsy on one of two lesions on the left kidney revealed Wilms' tumor. A partial left nephrectomy revealed focally positive margins that necessitated left flank radiotherapy. The tumor karyotype was 46,XY,t(7;8)(q36;p11)[8]/46,XY[12]. DNA sequencing of the WT1 gene revealed no mutation. His constitutional karyotype was 46,XY, suggesting that the t(7;8)(q36;p11) was associated with the malignancy. Cytogenomic evaluation of peripheral blood utilizing a 135K NimbleGen oligonucleotide array revealed a 539.57 kb chromosome 2p24.3 duplication involving four OMIM genes: NBAS, DDX1, MYCNOS, and MYCN. SNP array analysis of the tumor utilizing the CytoScan HD array identified a 560.49 kb chromosome 2p24.3 duplication with no other pathogenic genomic abnormalities. This is, to our knowledge, the first reported partial trisomy 2p case with constitutional MYCN duplication in a child with bilateral Wilms' tumor. As of May 2013, the 2-year-old boy continues to do well without clinical or radiographic evidence of recurrent disease. This case is also instructive because the child's health insurer initially denied authorization for CMA on the basis that 'there is insufficient evidence in the peer-reviewed literature to demonstrate the clinical/therapeutic utility' of CMA, and it was more than one year before such authorization was granted. That initial decision to deny coverage could have had untoward health implications for this child, as the identification of constitutional MYCN duplication necessitated surveillance for neuroblastoma. While the child did not develop neuroblastoma during the time it took to get insurance authorization for CMA, it is obvious how such a delay could have precluded early detection of a second malignancy for which this child is at high risk.

3231F

Incorporation of flanking probes reduces truncation losses for fluorescence in situ hybridization analysis of recurrent genomic deletions in tumor sections. *M. Yoshimoto[1,2], O. Ludkovski[3], J. Good[1], R.J. Gooding[4], A. Boag[1], A. Evans[3], M.S. Tsao[3], P. Nuin[1], J. McGowan-Jordan[2,5], J.A. Squire[1].* 1) Department of Pathology and Molecular Medicine, Queen's University, Kingston, Canada; 2) Cytogenetics Laboratory, Children's Hospital of Eastern Ontario, ON, Canada; 3) University Health Network, Princess Margaret Hospital, Division of Applied Oncology, Toronto, ON, Canada; 4) Department of Physics, Engineering Physics and Astronomy, Queen's University, Kingston, ON, Canada; 5) Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada.

The establishment of quality control standards in clinical laboratories using routine fluorescence in situ hybridization (FISH) analysis of formalin-fixed paraffin-embedded (FFPE) sections is crucial as more genomic biomarkers are used for predictive and prognostic oncology. For gene fusions, break-apart and gene amplifications there are guidelines for both analysis and reporting, but at the present time there are no comparable guidelines for the application and analysis of genomic deletions using FFPE sections. The use of FISH on archival FFPE samples is technically demanding and becomes more challenging when applied to paraffin-embedded tissue microarrays. In this study we report a generalizable four-color deletion FISH approach to assist interpretation problems arising when evaluating FISH signals. The guidelines will help address interpretative dilemmas associated with overlapping and truncated nuclei in FFPE prostate cancer sections. The four-color FISH approach was developed using the PTEN gene deletion model in prostate cancer. The PTEN assay was based on a robust bioinformatics analysis of 311 published human genome array datasets and comprises a centromeric 'chromosome enumeration' probe, a specific PTEN gene probe, and flanking probes either side of the target probe. The sensitivity and specificity parameters of the four-color PTEN probe set were further characterized using a large number of well-characterized tumors and stringent scoring criteria. The incorporation of flanking probes allowed the analysts to determine if the chromosomal region was subject to truncation loss. A minimum threshold for apparent deletion frequency was set to address the heterogeneous and homogeneous nature of tumor histology. In addition the approach facilitated analysis of genotypic heterogeneity and varying clonality within different foci of tumor in the prostate. Overall the approach provided robust and highly reproducible results that minimized inter- and intra-assay variability. The four-color FISH deletion assay reduced the frequency of misinterpretation and improved both the quality and throughput of FISH analyses using clinical samples. Moreover the use of established controls and conservative cut-offs for assigning deletions will facilitate more coherent approach to developing reporting standards for deletion assays as more tumor suppressor genes of clinical importance are discovered by next generation sequencing methods.

3232T

High resolution genome profiling in Li-Fraumeni patients without germline mutations in TP53 gene. *T.R. Basso[1], R.A.R. Villacis[1], M. Pinheiro[2], M.I.W. Achatz[1,3], S.R. Rogatto[1,4].* 1) CIPE, AC Camargo Cancer Center, São Paulo, SP., Brazil; 2) Bioscience Institute, UNESP, Botucatu, SP., Brasil; 3) Oncogenetics Department, AC Camargo Cancer Center, São Paulo, SP., Brazil; 4) Department of Urology, Faculty of Medcine UNESP, Botucatu, SP., Brazil.

Li-Fraumeni (LFS) and Li-Fraumeni like (LFL) syndromes predisposes a variety of different tumors occurring over a wide age range and are associated with germline mutations in TP53. Around 80% of LFS and 40% of LFL are predictive to carry germline TP53 mutations. The aim of this study was to assess germline copy number variations (CNVs) in a cohort of 18 patients with tumors typical of LFS/LFL, history of multiple primary tumors and without pathogenic mutations in the TP53 gene. Peripheral blood samples were obtained from patients at least one year after finished the chemotherapy and or radiotherapy. Genomic alterations were evaluated using the Affymetrix high-density microarray platform (CytoScan™HD). The data were analyzed by Affymetrix Chromosome Analysis Suite (ChAS) software. The CNVs detected were confirmed using the single nucleotide polymorphic (SNPs) probes present in the same platform. It was found 21 CNVs: Genomic gains on 1p31.3, 11q25 and 21q21.1 were detected in one case each. Nine cases presented 18 genomic losses mapped on chromosomes 1, 2, 4, 6, 7, 8, 9, 13 and 14. One female patient, who developed three malignant neoplasms (melanoma at age 40, breast at age 43 and bladder carcinoma at age 47), presented five genomic losses. Interestingly, one patient with history of three tumors (breast at age 58, non-Hodgkin lymphoma at age 61 and lip carcinoma at age 65) exhibited a large deletion (43 Mb) on 7q22.1-34. One patient was diagnosed as Turner Syndrome. In overall these findings reveal putative candidates associated with multiple primary cancer predisposition in these families.

3233F

A patient with Angelman-like features due to deletion of chromosome 15q26.1q26.2 encompassing *CHD2* and *RGMA*. *S. Kantarci[1], J.A. Martinez[2]*. 1) Deparment of Pathology and Laboratory Medicine, UCLA; 2) Depatment of Human Genetics, UCLA, Los Angeles, CA.

Angelman syndrome (AS) is a neurodevelopmental disorder caused by a variety of genetic abnormalities involving chromosome 15q11-13 region that is subject to regulation by genomic imprinting. Characteristic clinical features include severe developmental delay, speech impairment, gait ataxia, seizure, inappropriate laughter, microcephaly, strabismus, subtle dysmorphic facial features, and hyperactive lower limb deep-tendon reflexes. About 10% of patients with a presumptive clinical of Angelman syndrome (AS) have an unknown etiology. Here, we report a 4.5-year-old female patient with a history of global developmental delay. The patient was born to a 39-year-old G4 P3 mother with an ectopic pregnancy at 40 weeks' gestation. Prenatal chromosome testing was normal. At birth, the weight was 6 pounds and the length was 19 inches. She developed strabismus at 4 months of age and wears eyeglasses. By the age of 2 years, she presented speech delay, hypotonia, and increased tone on the left lower extremity. She walked at 14 months old, but toilet trained at 2.5-year-old. Although she had previously some ambulatory issues, her walking improved immediately after bilateral eye repair but she has some spasticity of the left lower extremity. She had her first words at the age of 2.5 years and sentences by 3.5-year-old. She has short stature, microcephaly, dysmorphic features (facial hypotonia with a flat midface, posteriorly rotated ears, micrognathia, and down turned corners of the mouth), and bilateral 5th finger clinodactyly. The patient was recently found to have absence seizures by EEG and they remain uncontrolled. Her parents and two siblings are healthy. Chromosomal microarray analysis (CMA) using Affymetrix CytoHD SNP array revealed a 2.7 Mb deletion of chromosome 15q26.1q26.2 ranging from genomic position 93,510,242 to 96,199,396 (GRCh37/hg19). This deletion interval includes 5 RefSeq genes: *CHD2* (exons 17-39), *RGMA*, *MCTP2*, *LOC400456*, and *LOC145820*. Parental CMA studies were not available. Chromosome 15q26.1q26.2 deletions (>5Mb) including *CHD2*, *RGMA*, and additional genes have been previously reported in patients with congenital diaphragmatic hernia. Recently, a de novo ~500Kb deletion of 15q26.1 encompassing only *CHD2* and *RGMA* in a female patient with Angelman syndrome-like phenotype without diaphragmatic hernia was described. Our report supports the significance of chromosome 15q26.1 deletion encompassing *CHD2* and *RGMA* in Angelman syndrome-like phenotype.

3234T

De novo complex unbalanced chromosome rearrangement, a case of ``constitutionalchromothripsis''. *M. Haddadin[1], M. ElNaggar[1], T. Sahoo[1], R. Owen[1], LM. Bird[2], L. Ross[1], S. Wang[1], L. Hinman[1], F. Boyar[1]*. 1) Deptartment of Genetics, Quest Diagnostics, Nichols Institute, San Juan Capistrano, CA; 2) Department of Pediatrics, Division of Genetics/Dysmorphology, Rady Children's Hospital San Diego.

The term complex chromosome rearrangement (CCRs) is used when 3 or more breaks involving two or more chromosomes are present. Chromothripsis was first coined by Stephens PJ, et al. (2011) to describe massive clustered chromosomal rearrangements (breakage and illegitimate repair) in cancer cells. Subsequently a similar phenomenon leading to CCRs in the germline of patients with congenital disorders was described (Kloosterman WP et al., 2011). The wide utilization of Oligo-SNP array for the study of congenital disorders and developmental abnormalities has allowed for the discovery of a wide range of genomic micro-rearrangements that were not detectable by conventional chromosome studies. Here we report a case of a newborn child with ventricular septal defect, Dandy Walker malformation, craniosynostosis, solitary right kidney and cryptorchidism. Chromosome analysis revealed an unbalanced rearrangement involving chromosomes 2 and 4. Oligo SNP-array study combined with FISH studies targeting the involved regions revealed a CCR involving chromosomes 1, 2 and 4, with duplicated segments of distal 1q, fragmented into five segments and inserted randomly into 2p and 4p. A pericentric inversion of chromosome 2; a reciprocal translocation between 2p and 4p and a microdeletion on 7q were also observed: (46,XY,der(2)ins(2;1)(p23;q41q32.1)inv(2)(p23q14.2)t(2;4)(p23;p16),der(4) ins(4;1)(p15.3;q41q42.1)t(2;4)(p23;p16,arr1q32.1q41(203,434,243-222,304,374)x3,1q42.11q42.12(224,577,514-225,253,174)x3, 1q42.12q42.2(226,610,463-231,048,716)x3,7q36.1(150,301,556-15 1,452,199) x1). Parental chromosome analysis and Oligo-SNP array analysis were normal. Since partial trisomy of chromosome 1 is very rare, segmental duplication of distal 1q may be the upper limit of what an embryo can tolerate and still survive. Although the other chromosome breaks and possible disruption of critical genes may contribute to the phenotype of this patient, the cardinal anomalies of skull deformation, CNS, cardiovascular and genitourinary anomalies in this patient may be attributed mainly to partial 1q duplications. A minimum of 17 break and reunion events involving 3 chromosomes having multiple translocations, duplications and an inversion characterize this CCR. Until recently, the finding of a de novo CCR of a child was attributed to germline mosaicism for one or more rearrangement in a parent. Germline chromothripsis may be involved in this case.

3235F

Telomerase gene copy number is increased in IBD and PSC. *A. Amiel[1,2], Y. Sulayev[1,2], H. Katz[1,2], A. Stein[3], M. Liberman[1], F. Konikof[3], I. Laish[3]*. 1) Genetic Institute, Kfar-Saba, Israel; 2) Bar-Ilan University, Ramat-Gan; 3) Gastroenterology and Liver Institute,.

Background: Telomerase is an enzyme complex that lengthens telomeres. It is composed of the catalytic subunit of human telomerase reverse transcriptase (hTERT) and the telomerase RNA component (TERC) encoded by TERC gene. Gene amplifications involving the TERC gene (3q26) are frequent in human tumors. Telomere capture (TC) is an alternative process by which broken chromosomes can aquire new telomeres. Primary sclerosing cholangitis (PSC) and inflammatory bowel disease (IBD) are pre-malignant conditions. The aim of this study was to evaluate the TERC gene copy number and TC status in peripheral lymphocytes of patients with PSC and IBD, as a possible surrogate marker for increased tendency for malignancy. Methods: By applying fluorescence in situ hybridization (FISH) to leukocytes of 14 PSC patients, 13 IBD patients (8 with Crohn's disease, 5 with ulcerative colitis) and 12 healthy controls, we estimated gene dosage of the TERC gene at 3q26.3. We used the SNRPN and 13q14.3 genes (red) and the 13qter and 15qter (in green) to evaluate the TC phenomenon. Results: The percentage of cells with more than two copies of the TERC gene was significantly higher in PSC patients (mean (3 x red signals) 44.7) than in IBD patients (mean 29.6) and controls (mean 3.9; p-value < 0.0001). The TERC gene copy number was also higher in PSC patients with concomitant colitis (64%) despite significantly lower disease activity indices. Significantly more cells were observed with TC with both genes in IBD lymphocytes than in control cells. Conclusion: TERC gene copy number is increased in IBD and even more in PSC lymphocytes, while TC is higher in IBD lymphocytes. These findings may be related to the different tendencies and predispositions of these conditions to become malignant. In addition, TC is another mechanism for telomere elongation and is probably higher when the telomerase gene copy number is lower.

3236T

Structural anomalies of the Y chromosome versus sexual ambiguity or infertility. *I. Ben Hadj Hmida[1], S. Dimassi[1], M. Kammoun[1], H. Kairi[2], H. Marmouche[3], A. Saad[1], S. Mougou-Zrelli[1]*. 1) Laboratory of Cytogenetics and Reproductive Biology and Human Genetics, Farhat Hached University Teaching Hospital, Sousse, Tunisia; 2) Department of Gynecology, Farhat Hached University Teaching Hospital, Sousse, Tunisia; 3) Department of Endocrinology, University Hospital Fattouma Bourguiba, Monastir, Tunisia.

Human Y chromosome evolution has progressively been directed towards a role in sex determination and reproduction. Among the structural anomalies of chromosomes, Robertsonian translocations between acrocentric chromosomes and reciprocal translocations are found with a frequency of 8 to 10 times higher in infertile men. Given the rather random chromosomal breakpoints encountered in reciprocal translocations, it is unlikely that achieving gametogenesis is due to the systematic breaking of a major gene in spermatogenesis. Genetic disorders of chromosomal origin could give rise to abnormal karyotypes or germinal mosaic figure. They could involve gene abnormalities affecting numerous genes localized on several chromosomes, in particular the Y chromosome. The physiopathologic identification of male infertility is interesting because of the risk of the genetic factors involved being transmitted to the offspring The isochromosome Y of either can be detected in female subjects with a classic Turner syndrome, or in subjects with gonadal dysgenesis with ambiguous genitalia or infertile male subjects. We identified, in the laboratory of Human Cytogenetics and Reproductive Biology, during January 2010 to January 2012, two patients with karyotype 46, XisoYp/45, X in one woman with a hypogonadotropic hypogonadism with Karyotype 46,XisoYp[5]/45, X[7] and in a man having azoospermia and sexual ambiguity with Karyotype 46,XisoYp[12]/45, X[6]. A fluorescent in situ hybridization was performed for these two cases in search of the SRY gene was present showing the form isochromosomique. The relationship between chromosomal abnormalities and infertility are not yet clear, since subjects with alterations may be highly variable in their spermatogenesis. Chromosomal rearrangements observed could either delete one or more genes involved in spermatogenesis or prevent proper chromosome pairing and initiate a shutdown of meiosis. Therefore individuals carrying isochromosomes Y, are described with variable phenotypes of infertility. Clinical manifestations range from a kind female phenotype Turner, a male phenotype with azoospermia and sexual ambiguity. Azoospermia may be accompanied by a deletion of the region Azoospermia factor (AZF).

3237F

Molecular Diagnosis of Deletions and Duplications Associated with Intellectual Disability and Obesity. *C.S. D'Angelo[1], M.F.M. Santos[1], C.A Kim[2], C.M. Lourenço[3], C.P. Koiffmann[1].* 1) Human Genome and Stem Cell Center, Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, Brazil; 2) Genetics Unit, Department of Pediatrics, Children Institute, School of Medicine, University of Sao Paulo, Sao Paulo, Brazil; 3) Neurogenetics Unit, Department of Medical Genetics, School of Medicine, University of Sao Paulo, Ribeirao Preto, Brazil.

Some rare genetic syndromes lead to obesity, often associated with intellectual disability (ID), abnormal development, dysmorphic features and congenital malformations. Prader-Willi syndrome (PWS) is the most common syndromic form of obesity. It is caused more specially by deletions of the SNORD116 (H/MBII-85) gene cluster in the paternally derived PWS/ Angelman syndrome (AS) region of chromosome 15. Other genomic disorders associated with copy number variants (CNVs), such as chromosomal deletions of 1p36, 2q37, 6q16, 9q34, 11p13 and 17p11.2, have an increased prevalence of obesity. In most cases, haploinsufficiency of one gene in the deleted region explains the majority of phenotypes of the disorders (e.g. HDAC4 [2q37], EHMT1 [9q34], RAI1 [17p11.2]), whereas in other syndromes haploinsufficiency of one gene in the deleted region explains only some specific feature(s); haploinsufficiency of SIM1 and BDNF explains the obesity in PWS-like patients with interstitial 6q deletions and WAGR patients with larger 11p13 deletions, respectively. The recent implementation of chromosomal microarray analysis (CMA) into the field of clinical genetics has revealed an increasing number of CNVs that are factors, or causes of obesity co-morbid with ID, e.g. deletions at 1p21.3, 2p25.3, and 16p11.2. Nearly 300 patients with obesity in the context of an ID syndrome were evaluated for chromosomal imbalances by MLPA after DNA methylation analysis ruled out PWS. Half of these patients were further investigated with different array platforms. We identified syndromic genomic disorders, such as deletions of 1p36 (7 cases), 2q37 (HDAC4; 5 cases), 6q16 (SIM1; 2 cases), 9p24 (3 cases), 9q34 (EHMT1; 1 case), and 17p11.2 (RAI1; 5 cases), as well as specific CNVs already linked to obesity, such as 1p21.3 and 2p25.3 deletions and an unbalanced translocation der(8)t(8;12) . Additionally, we identified 22q11.2 deletions causing DiGeorge syndrome (6 cases), 22q11.2 duplication syndrome (1 case), and 22q11.2 distal deletion syndrome (2 cases). Known genomic disorders with incomplete penetrance and variable expressivity included 15q11.2 deletion of NIPA1 (2 cases), 16p11.2 deletion of SH2B1 (1 case), 16p11.2 duplication of TBX6 (1 case), 16p13 deletion of MYH11 (1 case), 17q11.2 duplication of NF1 (1 case), and 17q21.31 duplication of MAPT (1 case). Clinical variability in well-known syndromes may facilitate the identification of disease genes.Financial Support: CEPID-FAPESP, CNPq.

3238T

Mechanisms of concurrent deletions and duplications at 1p36. *M. Gajecka[1,2], J.A. Karolak[1,2], J. Shen[3], K. Wakui[4], C. Glotzbach[5], L.G. Shaffer[6].* 1) Institute of Human Genetics, PAS, Poznan, Poland; 2) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poland; 3) Children's Hospital Central California, Madera, CA, USA; 4) Department of Medical Genetics, Shinshu University School of Medicine, Nagano, Japan; 5) Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA, USA; 6) Paw Print Genetics, Genetic Veterinary Sciences, Spokane, WA, USA.

In Monosomy 1p36, four classes of rearrangements are identified: pure terminal deletions, interstitial deletions, unbalanced translocations and complex rearrangements. Here we present seven cases of complex rearrangements involving terminal or interstitial deletions and duplications. To characterize the aberrations and identify mechanisms of concurrent deletion and duplication formation, rearrangements were analyzed using array CGH, SNP arrays and other molecular cytogenetics and molecular biology methods. For each individual, deletion and duplication sizes and parental origin of the rearrangements were determined. Fiber FISH analysis demonstrated either tandem or direct duplications. Rearrangement breakpoints were determined and characterized and we found that deletion and duplication sizes as well the breakpoints locations were different in each patient. We were not able to identify all junctions in the rearrangements. In four cases, apparently simple junctions were found to be more complex at the sequence level. Both duplications and deletions were interrupted with insertions of material from 1p or other chromosomal regions. Our results show higher than accepted complexity of the studied rearrangements and indicate involvement of multiple mechanisms in the DNA breakage and repair process during rearrangement formation. Scenarios of the rearrangement formation based on the DNA sequence level examination and computational analyses will be presented. Support: Polish Ministry of Science and Higher Education, Grant NN301238836.

3239F

Unexpected 10.6 Mb 19q13.33q13.43 duplication due to a de novo translocation: Contribution for phenotype-genotype correlation. *F.T. Lima, G.M.G. Carvalheira, S.S. Takeno, M.M. Oliveira, V.F.A. Meloni, M.I. Melaragno.* Disciplina de Genética, EPM-UNIFESP, Sao Paulo, SP, Brazil.

The current use of high-resolution techniques, such as genomic arrays, has revealed an unprecedented number of different cytogenomic unbalances. One of them, partial trisomy 19q, a rare aneusomy, had only a few cases reported in literature, with different breakpoints and an inconsistent clinical delineation. We report on a female patient with mild phenotypic findings and an important cognitive alteration, the only child of a non-consanguineous couple. At eight years old, she presents some facial dysmorphisms, astigmatism, short stature, long fingers, Raynaud's phenomenon, skeletal alterations, neuropsychomotor developmental delay and intellectual disability. She had recurrent infections, urolithiasis and seizures beginning at age four years old. Karyotype using G-banding at 550 band-resolution was normal. Subsequent studies with the MLPA (Multiplex Ligation-dependent Probe Amplification) technique, using SALSA MLPA kit P070 Human Telomere-5 containing probes for subtelomeric regions, revealed three copies of 19q. Genomic array using the Affymetric Genome-Wide Human SNP Array 6.0 showed a 10.6 Mb duplication of the long arm of chromosome 19 resulting in: arr19q13.33q13.43(48,463,121-59,097,842)×3. FISH with a BAC probe for 19q13.43 (RP-11-359B7), used to determine the genomic duplication position, revealed that the duplicated region was attached to the short arm of chromosome 21. The distal region of the chromosome19 corresponds to a genomic segment with a very high transcriptional activity, with more than 600 known genes encompassed by the duplicated region, 30 of them associated with known diseases, including epilepsy, intellectual disability and speech delay, as presented by the patient described. This rearrangement is unprecedented in literature, and its size is unexpected in view of the phenotype observed. The phenotypic consequences of duplications are complex to be studied because the chromatin context can modify the transcriptional status of trisomic segments, especially when attached near DNA repetitive regions. Therefore, localization and orientation of the extra copy may interfere in its expression levels and, consequently, in the clinical features. This report adds important data for the genotype-phenotype correlation of this region and emphasizes the importance of a cytogenomic assessment in the presence of a normal karyotype. Financial support: FAPESP, CNPq, Brazil.

3240T

Small supernumerary marker chromosome 17: molecular characterisation of a new case and review of the literature. *J.M. Turbitt[1], J. Murphy[1], R. McGowan[2], P. Batstone[1].* 1) Cytogenetics department, North of Scotland Regional Genetics Service, Aberdeen; 2) Clinical Genetics Unit, Aberdeen Royal Infirmary, Aberdeen.

Small supernumerary marker chromosomes (sSMC) are structurally abnormal chromosomes that cannot be thoroughly characterised by conventional G-banding analysis and are present in 0.288% of patients with developmental delay. The unstable, gene-rich 17p11.2-p12 chromosome region is associated with various structural anomalies including SMCs. To date there are relatively few cases of chromosome 17-derived sSMCs reported in the literature. Those that are recorded are of variable size, tissue distribution and level of mosaicism. Here we report a 30-year-old male with a de novo mosaic SMC(17). Our patient presented with learning difficulties, anxiety and tremors and is a known carrier of a paternally inherited balanced translocation between chromosomes 1 and 11, which is thought not to contribute to his phenotype. Conventional cytogenetic analysis revealed the presence of two cell lines. One cell line had 46 chromosomes and the t(1;11) and the second had 47 chromosomes with the t(1;11) and an additional sSMC. FISH and array-CGH studies revealed that the sSMC was comprised of the centromere and proximal short arm of chromosome 17. The sSMC appeared to be present in approximately 70% of cells. Parental chromosome analysis and array-CGH indicate that the sSMC appears to have arisen de novo. Phenotypically patients with the sSMC(17) have been described with mental retardation/developmental delay, hypotonia and dysphasia; features which are often found in Potocki-Lupski syndrome (PLS). Whilst this patient's duplication overlaps the PLS region it does not include the FLII/RAI1 gene region that has been implicated in the syndrome. To our knowledge, only a few cases of SMC(17) that do not include this region have been described to date. Our patient's SMC(17) will be compared to other reported cases in the literature.

3241F

Wolf Hirschhorn Syndrome with epibulbar dermoids in a patient with deletion 4p associated to Xp duplication. *S. Bragagnolo, M. Colovati, R.S. Guilherme, M.I. Melaragno, A.B. Perez.* Centro de Genética Médica, UNIFESP - EPM, São Paulo, São Paulo, Brazil.

Wolf-Hirschhorn syndrome (WHS) is a multiple malformation genomic syndrome that results from a deletion of the critical region (WHSCR) at 4p16.3 region. About 50%-60% of individual have a de novo deletion of 4p16 and about 40%-45% have an unbalanced translocation de novo or inherited from a parent with a balanced rearrangement. WHS is characterized by typical craniofacial features consisting of broad bridge of the nose continuing to the forehead, microcephaly, high forehead with prominent glabella, ocular hypertelorism, epicanthus, highly arched eyebrows, short philtrum, downturned mouth, micrognathia, and poorly formed ears with pits/tags. All affected individuals have pre- and postnatal growth deficiency and retardation, hypotonia with muscle underdevelopment and developmental delay/intellectual disability. The ocular abnormalities include strabismus, refractive errors, downslanting palpebral fissures, microphthalmos, microcornea, iris coloboma, optic nerve coloboma, ocular cyst, ptosis, glaucoma, and nystagmus. The epibulbar dermoid, hemifacial microsomia, ear malformation and vertebral anomalies are observed in patients with oculoauriculovertebral spectrum (OAVS). It has been recently suggested, based on the molecular analysis of a balanced translocation t(4;8)(p15.3;q24.1) in an OAVS patient, that abnormal expression of BAPX1, located in 4p15.3, might be involved with the etiology of OAVS. We described a six month-old female, with cleft palate, pulmonary stenosis, atrial septal defect and gastroesophageal reflux. Presented with growth deficiency, developmental delay, apparent facial asymmetry, high forehead, ears tags and left microtia, cleft palate, pectus carinatum and pre-sacral pit. The ophthalmological evaluation revealed epibulbar dermoids in the right eye. Her mother, maternal aunt, grandmother and great-grandmother have ears pits and hearing impairment. The patient' s conventional G-banded karyotype was normal at 550 band resolution but genomic array (Genome-Wide Human SNP Array 6.0,Affymetrix) revealed a ~13 Mb deletion in chromosome 4 and a ~9 Mb duplication in chromosome X (.arr 4p16.3p15.33(68,345-13,569,183)×1,Xp22.33p22.31(168,551-8,907,556)×3, GRCh37/hg19), including the BAPX1 gene in 4p15.3, a possible candidate involved in OAVs pathogenesis. This is the first patient with WHS with epibulbar dermoids, hemifacial microsomia, and other phenotypic changes common to OAVS.

3242T

Goldenhar syndrome and oculoauriculovertebral spectrum (GS/OAVS): Clinical and cytogenomic study. *M.ES. Colovati, S. Bragagnolo, R.S. Guilherme, A.G. Dantas, A.B. Perez, M.I. Melaragno.* Genetics Division, Department of Morphology and Genetics, Universidade de São Paulo, São Paulo, Brazil.

Goldenhar syndrome (GS) has been characterized as having an association with external ear deformities, facial asymmetry, epibulbar dermoid, and vertebral changes. Its etiology is multifactorial, familial cases have been described with autosomal recessive or dominant. Some authors characterizes microtia, hemifacial microsomia and Goldenhar syndrome phenotypes a oculoauriculovertebral spectrum (OAVS). In contrast, Goldenhar syndrome is part of the OAVS, but there is changes in other organs or systems, the most common being cardiovascular (50%), central nervous system or mental impairment (5 to 10%) and renal / respiratory (5%). The accuracy of a clinical diagnosis is a challenge for clinical geneticists, since there is a heterogeneous phenotype and etiology. To date, several chromosomal abnormalities have been associated with the syndrome involving most frequently 22q and 5p. In literature, five 22q11.2 microdeletions have been described involving the same region, suggesting a possible OAVS candidate gene. In this study we analyzed 30 patients with GS/OAVS by G-banding and we found three patients with positive cytogenetic results with two apparent chromosome heteromorphisms and one inversion: 14ps+, 15ps+ and inv(12)(q15q24.1). These three patients were evaluated by genomic array (Genome-Wide Human SNP Array 6.0,Affymetrix) and Multiplex Ligation-Dependent Probe Amplification (MLPA kit SALSA P250). In one patient these metodologies revealed a ~581 kb deletion in chromosome 22 (.arr 22q11.21(20,716,923-21,297,749)×1, GRCh37/hg19). The regions deleted includes 21 genes and it is flanked by the low copy repeats LCR-B and LCR-C, not including the 1.5 Mb DiGeorge critical region (LCR-A to LCR-B). This patient presents hemifacial microsomia, retrognathia, agenesis of external auditory canal and ossicular chain with right moderate hearing loss, soft palate cleft, and thoracic hemivertebrae. His neuro-psycho-motor development is normal. We present the sixth patient with OAVS and 22q11.2 microdeletion reported in literature. This study supports the previous reports indicating that patients with GS/OAVS should be screened for 22q11.2 microdeletion and other chromosomal abnormalities in order to identify genomic loci that are potentially involved in this disease pathogenesis and also to understand the phenotypic and genetic heterogeneity in this syndrome. Financial support: FAPESP, Brazil.

3243F

CGH-array characterization of an interstitial 7p deletion in two patients exhibiting features of GCPS, RSS and SCS syndromes. *H. Hannachi[1], M. Kammoun[1], S. Dimassi[1], A. Labalme[2], D. Sanlaville[2], M. Gribaa[1], S. Mougou-ZERELLI[1], A. Saad[1].* 1) Human Cytogenetics, Molecular Genetics and Reproductive Biology Departements CHU Farhat Hached, Sousse, Tunisia; 2) Laboratoire de cytogénétique constitutionnelle Centre de Biologie et Pathologie Est CHU de Lyon-GH Est - Hospices Civils de Lyon 59 Boulevard Pinel 69677 BRON CEDEX FRANCE.

Given the richness of the short arm of chromosome 7 in genes regulating development, as GLI3, TWIST and HoxA cluster, 7p interstitial deletions are well documented. We report here, two additional patients sharing common clinical features involving severe craniofacial dysmorphism, syndactyly, severe psychomotor retardation and failure to thrive. Deletions, detected in standard karyotype, benefited of a CGH-array exploration revealing losses of genomic material of 11Mb and 13Mb. Molecular investigations performed to characterize the parental origin of rearrangements showed that the small deletion is carried by the paternal chromosome; the other is carried by the maternal one. Association of haploinsufficiency of the segment 7(p12p14) in the first patient with the triad of symptoms: frontal bossing, limb abnormalities, mental retardation evoked Greig syndrome (GCPS). Alteration of GLI3 gene is detected in 75% of patients carrying this syndrome. Haploinsufficiency of this gene in our patient resulted in several phenotypic features compatible with the mild form of GCPS. However, they differ in the occurrence of microcephaly and severe failure to thrive. The deleterious region encompasses the maternally imprinted segment 7(p12p13). Uniparental maternal disomy (UPDmat) of this interval was reported in 10% of Silver-Russell syndrome (RSS) patients. This syndrome is mainly characterized by a severe failure to thrive. Then, we hypothesize that the nullosomy for the paternal copy of 7(p12p13) gave rise to the symptoms of RSS, in this patient. On the other hand interstitial deletion of the segment 7(p21.3p15.3) in the second patient alters the candidate gene TWIST, which is known to be involved in 71% of cases of Saethre-Chotzen syndrome (SCS). Hemizygosity of this gene seems to be the main cause of SCS phenotypic manifestations in this patient. The particular severity of the phenotype and the development of an umbilical hernia may be related to the loss of a large number of genes mapping on this segment. Our findings support that the clinical features of patients with del(7p) are relatively consistent and specific to the region of deletion. Using CGH-array for diagnosis purposes may refine the characterization of additional cases allowing better genotype-phenotype correlation.

3244T

Compound Heterozygous Microdeletion of Chromosome 15q13.2q13.3 Region in a Child with Hypotonia and Global Developmental Delay. *P. PRASUN[1], L. SIVASWAMY[3], L. SCUSSEL[3], M. HANKERD[4], M. KRISTOFICE[4], S. EBRAHIM[2]*. 1) Division of Genetics and Metabolism, Children's Hospital of Michigan, Detroit, MI; 2) Department of Pathology, Wayne State University School of medicine, Detroit, MI; 3) Division of Pediatric Neurology, Children's Hospital of Michigan, Detroit, MI; 4) Detroit Medical Center University Laboratories, Cytogenetics Laboratory, Detroit, MI.

Heterozygous deletion of the 15q13.3 BP4-BP5 region (OMIM #612001) is characterized by a highly variable clinical phenotype ranging from mental retardation, epilepsy, neuropsychiatric disorders, and facial dysmorphism to a complete absence of symptoms. Homozygous and compound heterozygous micrdeletions of the same region are extremely rare with only a few cases having been reported in the literature to date. We report on the molecular detection of compound heterozygous 15q13.2q13.3 deletions in a 23 month old Caucasian female with global developmental delay, generalized muscular hypotonia, and visual dysfunction. The patient was first evaluated by pediatric neurologist at the age of 3 month and found to have generalized hypotonia, no head control, no face regard or visual fixation. MRI of the brain and routine karyotype were obtained both of which were normal. There was concern of cortical visual failure and a detailed evaluation by pediatric ophthalmologist was done. Ophthalmologic examination and evaluation consisting of visual evoked potential (VEP) and electroretinogram (ERG) were normal. Patient has a 6 year old brother with global developmental delay, hypotonia, and seizure disorder. He was found to have mosaicism for ring chromosome 20 by G-banding karyotype. Chromosomal microarray detected a 1.28 Mb deletion on one chromosome 15q13.2q13.3 region, and a smaller sized deletion estimated to be 410 Kb in size on the second chromosome 15q13.3 homologue resulting in nullisomy for at least the smaller deletion within the15q13 region. The larger deletion contained 7 genes: MTMR15, TRPM1, MIR211, KLF13, OTUD7A, CHRNA7, and MTMR10. The smaller deletion contained CHRNA7 and part of OTUD7A genes. The deletions were confirmed by FISH. Thus this patient is a compound heterozygous for the 1.28 Mb deletion on one chromosome 15q13.2q13.3 and a smaller size deletion on the other chromosome 15q13.3 region. Compound heterozygous 15q13.3 microdeletion is extremely rare and to our best of knowledge only two such patients have been reported in literature thus far. The findings in our patient suggests that pathogenesis of visual dysfunction which is a consistent finding in homozygous/compound heterozygous 15q13.3 microdeletion depends upon the size of microdeletion. Homozygous loss of CHRNA7 leads to cortical visual dysfunction while homozygous loss of TRPM1 further contributes to visual failure by retinal dysfunction.

3245F

Deletion 12q12 - 12q13.11: Case Report with unusual neurologic manifestations. *A.M. Zarante[1], G. Giraldo[1], J.C Prieto[1,2]*. 1) Instituto de Genetica Humana, Universidad Javeriana, Bogota, Colombia; 2) Hospital La Victoria, Secretaria Distrital de Salud, Bogota, Colombia.

Interstitial deletions involving the long arm of chromosome 12 are rare events and few case had been reported in the literature , this deletions are characterized by : developmental delay, growth retardation, macrocephaly, facial dysmorphism, prominent forehead, hypertelorism, downslanting palpebral fissures, upturned nose, Low-set ears, pterygium colli, cardiac anomaly, 2-3 toe syndactyly, single palmar creases, genitourinary malformations, anomalies , pyloric stenosis, Ectodermal anomaly, asthma/reactive airway, ocular abnormalities and normal MRI. We present a case of a Colombian girl of 8 years old. She was born at term after a normal pregnancy, non-consanguineous parents and family history was unremarkable. The patient presented a global developmental delay. MRI showed cerebellar vermix hypoplasia. The findings found on the physical examination were: microcephaly, narrow forehead, front and back low implantation hair, bilateral palpebral ptosis, downslanting palpebral fissures, high nasal root, hypoplastic nasal wings, anteverted nares, flat philtrum, thin lips, dental crowding, with winged ears, helix folding, in the limb presents shortening and clinodactyly fifth finger, 2-3 toe syndactyly and joint hypermobility. The band R karyotype reported: 46, XX, the Comparative Genomic Hybridization (CGH) analysis showed deletion of 12q12-12q13.11 (2,695 MB). The parents were negative for deletion of 12q12-12q13.11, Then it is a novo deletion. We describe a patient with a cytogenetically 12q deletion is about 2695 MB detected by Array CGH this deletion has been associated with developmental delay, growth retardation, facial dysmorphism. The striking thing about this case is the cerebellar vermix hypoplasia and microcephaly.

3246T

Molecular cytogenetic techniques in investigation of suspected microdeletion syndromes: an experience with 330 cases. *A. Halder, M. Jain, I. Chaudhary, V. Mohan, P. Kumar*. Reproductive Biology, All India Institute of Medical Sciences, New Delhi, Delhi, India.

Background: Microdeletion syndrome is characterized by small (< 5Mb) chromosomal deletion in which one or more genes are involved. They are frequently associated with multiple congenital anomalies. The phenotype is the result of haploinsufficiency of genes in the critical interval. FISH, MLPA, QFPCR and aCGH techniques are commonly used for the diagnosis. Here this study will assess role of FISH and aCGH in the diagnosis and research on suspected microdeletion syndrome. Method: This study was comprised of 330 cases of suspected microdeletion syndromes. There were 184 cases of 22q11.2 microdeletion, 52 cases of William, 47 cases of Prader Willi/ Angelman, 18 cases of Miller Dieker, 14 cases of Retinoblastoma, 5 cases of Trichorhinophalangeal (TRP) and 10 cases of other microdeletion syndromes. FISH was carried out in all using non-commercial probes. Subsequently, aCGH was performed in 55 cases (40 cases of 22q11.2 and 15 cases of other microdeletion). In another 36 (mostly 22q11.2 microdeletion) aCGH experiment is in progress. Result: FISH was confirmatory in 29 cases (8.8%; 20 cases of 22q11.2 microdeletion, 5 cases of Prader Willi, 3 cases of William and 1 cases of TRP syndrome). There were 10 cases with mosaicism and 19 cases with pure deletion. Microarray was picked up CNV with or without LOH in 72.5% of cases, mostly involving several chromosomal loci. However, aCGH was failed to pick up mosaic cases (even 45% deleted cell lines). Clinically suspected specific deletion was detectable only in 27.3% cases by aCGH. Variation in deletion size, break point difference as well as other CNVs was observed. Conclusion: We conclude FISH should not be the first method of choice for clinically suspected microdeletion syndrome as cost, labor & time versus benefit is unjust. We think aCGH should be first line of investigation and FISH may be used for confirmation, detecting mosaicism, screening family members and prenatal diagnosis. However, microarray is likely to miss mosaic cases, if deleted cell concentration is less than 50%. Furthermore, microdeletion syndrome best fitted with genomic disorder as several chromosomal loci are involved in CNV with or without LOH and alteration in deletion size or breakpoint. We did not find identical deletion profile in any case, thus explaining reason for phenotypic variability between cases.

3247F

Array CGH in pediatrics diagnosis versus karyotype. Experience with 189 cases. *M. Pérez Sánchez[1], A. Mora.[1], J.L. Barrionuevo.[2], S. Roldán.[2], A.R. González.[3]*. 1) Análisis Clínicos, Hospital Virgen de las Nieves, Granada, Granada, Spain; 2) Pediatria. Hospital Virgen de las Nieves. Granada. Spain; 3) FIBAO. Hospital San cecilio. Granada.

Microarray-based comparative genomic hybridation (array CGH) has provide a relatively quick method to scan the genome for gains and loses of chromosomal material with significantly higher resolution and greater clinical yield than was previously possible. This new methodologies have led to identification of novel genomic disorder (microduplications and microdelections) in patients with developmental delay/mental retardation and/or multiple congenital anomalies (DD/MR/MCA), with a significant increase in diagnostic yield . In this study we present the result of array CGH obtained in 189 children with normal karyotype but DD/MR/MCA . The NimbleGen CGX Cytogenetic Microarrays platform was performed. The results has shown that in 41 patients (21,7 %) was detected a chromosomal delection or duplication previously described like pathogenic copy number variants (CNVs),. In 18 cases (9,5 %) was necesary the analysis of parental simples,showing that 7 anomalies (49 %) had ocurred de novo and was clasificated as pathogenic and in 11 cases (61 %) appeared to be inhereted from an unaffected parent. In a total of 48 patients (25,4 %) was posible to detect a pathogenic CNVs . In contrast, only in less than 2 % of cases chromosomal alterations can be detected when karyotyping studies are performed. Several recent studies suggest that when aCGH is performed with a finding of an apparently normal karyotype, the diagnostic yield increases by an additional 8-17 %, in our study, we have obtained a 25,4 % of children with pathogenic CNVs that is higher than the results obtained by other authors. This increase at the detection rate probably is due to the array type utilized or for different methods in patients selecction. As a conclusison, array CGH is the most advanced method yet for assessing genomic imbalances associated with genetic disease. It has greatly enhanced the diagnostic capabilities of the clinical cytogenetics laboratory and has led to the continued discovery of novel genetic síndromes. These technology is being implemented routinely in our laboratory when a phenotype of DD/MR/MCA autism and apparently balanced translocations are presents.

3248T

Linear Measurement of the 2p telomere in μm may be associated with dementia status of people with Down syndrome. *E.C. Jenkins[1], L. Ye[1], S.J. Krinsky-McHale[1], W.B. Zigman[1], N. Schupf[1,2], W.P. Silverman[3].* 1) Dept Hum Genetics, NYS Inst Basic Res Dev Disab, Staten Island, NY; 2) Taub Institute for Research in Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY; 3) The Kennedy Krieger Institute and The Johns Hopkins University School of Medicine, Baltimore, MD.

We have recently reported that 11 people with Down syndrome (DS) and mild cognitive impairment (MCI) exhibited shorter telomeres in short-term T lymphocyte cultures versus 11 age- and sex-matched controls with DS without MCI (Jenkins et al., 2012). Multiple methods were used to compare telomere (tel) lengths of individuals with and without MCI using an FITC-labeled PNA tel probe (DAKO). Regardless of the method used, we found a consistently strong relationship between tel length and cognitive status [i.e., MCI (shorter tel) vs. no MCI (longer tel)]. We were able to develop classification criteria for tel 'length' that corresponded to clinical status (MCI/no MCI) with either perfect or near perfect sensitivity and specificity. In fact, one method based on 'direct' linear measurement of chromosome 1 tel length referenced against the remainder of chromosome 1, showed a 4 SD separation between the highest ratio from a case with MCI versus the lowest ratio from a case with 'normal aging.' This led us to hypothesize that this method will be validated as either a biomarker of clinical status or as a predictor of individual risk. Recently, we developed a less labor-intensive method that might have equally impressive power, by testing the same digitized material with the short arm of chromosome 2, reducing measurement time by almost half. We chose chromosome 2 because: 1) it provided us with a new measurement (distinct from our original measures), 2) we had a centromere 2 probe (provided as a gift from DAKO) that allowed rapid identification of chromosome 2, and 3) the cen [2] probe facilitated measurement of the chromosome 2 short arm. Eleven people with DS and MCI have now been compared to 11 age-sex-matched DS controls thus far. Consistent with our previous findings, tel length in the two groups did not overlap. In fact, the largest ratio for an individual with MCI was 2.45 SDs smaller than the smallest ratio for a control case. This substantial separation in scores between groups was associated with a highly significant difference in group means, $t(10) = 9.64$, $p < .0001$. We now plan to extend these findings to establish a method likely to have perfect specificity and sensitivity but requiring significantly less time and effort compared to the previous procedure which we found to show the greatest separation between groups. Supported in part by NYS OPWDD, Alz. Assoc. IIRG-96-077, IIRG-07-60558; NIH PO1-HD035897, RO1-HD037425, RO1-AG014673, and P30-HD024061.

3249F

Copy Number Variations in patients with Agenesis of Corpus Callosum. *C.P. Oliveira[1,2], C.T.N. Medina[2], R.S. Heredia[2], R.O.A. Benício[2], M.T.O. Cardoso[2], S.F. Oliveira[3], A. Pic-Taylor[3], J.F. Mazzeu[1,4].* 1) Laboratory of Clinical Genetics, Faculty of Medicine and Health (FS), University of Brasilia (UnB), Federal District, Brazil; 2) Genetics Nucleus (NUGEN), Hospital de Apoio (HAB), Secretary of State for Health, Federal District, Brazil; 3) Department of Genetics and Morphology, Institute of Biologic Sciences, University of Brasilia (UnB), Federal District, Brazil; 4) Post-graduation program in Genomic Sciences and Biotechnology Catholic University of Brasilia (UCB), Federal District, Brazil.

Agenesis of corpus callosum (AgCC) is a common brain malformation that can be found isolated or in association with other systemic malformations. Individuals with brain malformations have a high rate of de novo copy number variations (CNV), including microduplications and microdeletions, which may play an important causative role. This study aimed to search for copy number variations in patients with AgCC in order to identify possible causative genes for this condition. We assessed 19 patients with partial or complete AgCC by a comprehensive radiological and clinical review. Genome cytogenetic analysis was performed using Affymetrix 750k SNP array. At least 25 consecutive probes were considered in order to call for an abnormality. From a total of 19 AgCC patients, we identified six that showed pathogenic CNVs. Two individuals presented large chromosomal rearrangements also confirmed by G-banding: a tetrasomy 9p and an unbalanced translocation t(5;15). In one patient, array data disclosed a 3p duplication (~22 Mb) and a deletion 13q (~30Mb) suggesting a possible translocation, further confirmed by re-examination of patient karyotype: t(3;13)(q26.3;q31.1). In three patients we identified small rearrangements: a 657 kb duplicated segment on chromosome 16q (chr16:77,972,993 - 78,540,503 - hg19), a 261 kb duplicated segment on 4q (chr4:73,833,080-74,094,428) and a 760 kb deleted segment in Xp (chrX:1,979,232-2,742,089). These segments were not described previously as common CNVs and genes mapped in these regions are correlated with various neurological and morphological disorders and clinical features, like intellectual disability and brain malformations. These results confirm that chromosome rearrangements are an important cause of AgCC and that chromosome microarray analysis is an useful diagnostic tool for the etiology of this condition. Financial support: CNPq, FAPDF(PPSUS).

3250T

Sex Chromosomal Mosaicism In The Gonads Of Patients With Disorders Of Sex Development (DSD). *AKader. SA[1], Kamel. AK[1], AGhany. HM[2], Desouky. N[3], Mekkawy. MK[1], Hennawy. A[4], Makhluf. M[2], Mazen. I[5].* 1) Human Cytogenetics, National Research Center, Cairo, Egypt; 2) Cytogenetics,Specialized Children's Hospital,Cairo University, Cairo, Egypt; 3) Pediatric Surgical Division, Specialized Children's Hospital, Cairo University, Cairo,Egypt; 4) Pathology,Cairo University, Cairo,Egypt; 5) Clinical Genetics,National Research Center,Cairo, Egypt.

The phenotypic variability observed in DSD patients with sex chromosomal abnormalities depends on many factors including the presence of SRY gene, the complexity of the structural rearrangement and the presence of chromosomal mosaicism, especially 45,X cell line,which can affect the threshold of SRY gene expression. The phenotypic sex strongly depends on the percentage of Y chromosome and 45,X cells in the developing gonads. This can explain differences observed between gonadal phenotype and the karyotype. We report on ten patients with variable presentations of disorders of sex development (DSD) including ambiguous genitalia, primary amenorrhea and short stature. Conventional cytogenetics studies and flourescence insitu hybridization(FISH) technique on peripheral blood showed a non mosaic chromosomal constitution in five patients. The other five patients exhibited mosaic cell pattern associated with different types of sex chromosomal abnormalities including 45,X, isodicentric Y chromosome, X;Y translocation and ring X chromosome. FISH analysis on paraffin embedded or fresh cultured gonadal tissue specimens was done for all patients and showed mosaicism in eight of them and a single cell line in two patients. Our study demonstrates the importance of studying sex chromosome mosaicism in the gonadal tissue of patients showing a discrepancy between their karyotype and gonadal phenotype. Using FISH technique on paraffin embedded or fresh cultured gonadal tissue specimens is recommmended in those patients for better understanding of the phenotype/ karyotype correlation.

3251F

Directional genomic hybridization: An improved biomarker for radiation exposure. *F. Ray[1,2], E. Robinson[2], M. Cornforth[2,3], J. Bedford[1,2], E. Goodwin[2], S. Bailey[1,2].* 1) Environmental Radiological Health Sciences, Colorado State Univ, Fort Collins, CO; 2) KromaTiD Inc, 320 East Vine Drive, Fort Collins, CO; 3) University of Texas Medical Branch, Galveston Texas.

Cytogenetic biomarkers have been a preferred choice for retrospective estimation of radiation exposure because they are sensitive, quantifiable, and relevant to biological effects of concern. The most commonly used involve the measurement of dicentrics and symmetrical translocations. Both have shortcomings that become increasingly problematic for assays carried out at long times after radiation exposure. Dicentrics in samples from peripheral blood lymphocytes decrease with time after exposure (t1/2 1-2 years), and in the case of more stable symmetrical translocations, background levels are 10-fold higher and increase with age. Another aberration type, inversions, result from exchanges within a chromosome that reverse the orientation of the broken segment. We have developed an approach based on directional genomic hybridization (DGH) that facilitates detection of inversions with a greater than 10 fold improvement in resolution over existing techniques, allowing the detection of 1Mb or smaller inversions. Bioinformatics guided design of sequence- and strand-specific probe sets, which when coupled with single-stranded hybridization, produced chromatid - rather than chromosome - paints. Inversions register simply as a signal switch from one sister chromatid to the other in the inverted region. Importantly, like chromosome paints, chromatid paints also reveal translocations and dicentrics. Modeling suggested that inversions should be more common than translocations after densely vs. sparsely ionizing radiation exposures. We irradiated human cells with high LET heavy ions or low LET gamma rays and, using chromatid painting, compared the dose-response yields for induction of inversions, translocations and dicentrics. As predicted, the slope of the dose-response curve following heavy ion irradiations was steeper, and the yields per unit dose for inversions were higher than for either translocations or dicentrics. In another application, chromatid painting of orangutan cells revealed an inversion that presumably occurred during karyotype evolution of mammals. Together, our results demonstrate that inversion detection is useful for a variety of applications and can be further developed for use as a sensitive tool to measure past exposure to ionizing radiation and/or other clastogens. Funding for this work was provided by NASA (NNX09CE42P; NNX10CB05C) and NIAID (R01AI080486-02).

3252T

A novel approach to detecting miRNA in living cells. *D. Weldon, A. Ko, G. Johnston, V. Koong.* EMD Millipore, Temecula, CA.

The regulatory role that microRNA's play in gene expression has made them a target of interest in many areas of research. Unfortunately, detecting miRNA has been limited to fixed or lysed cells, which presents the problem of looking at only a single time point while utilizing duplicate samples to mimick more dynamic study's of changes in expression levels which can be observed only in live cells. The ability to monitor miRNA levels within live cells without altering gene expression or affecting cell health allows for a more biologically relevant understanding of miRNA biology. However, monitoring RNA within intact cells can prove to be challenging with current techniques due to complex or harmful sample preparation techniques or transfection reagents. Further, amplification methods can create false positives or erroneously inflate differences. Here we describe the ability to detect miRNA levels in live intact cells without the need for transfection reagents. We compare the expression level of miR-21, miR-210, and miR-155 in live cells to that of levels seen by RT-PCR. RT-PCR data is limited to a population of cells where the live cell data is captured at the single cell level. Our technique utilizes gold nano-particles conjugated to duplexed oligonucleotides. In the presence of target RNA a fluorescent reporter is released from the proximity of the gold and fluorescence can be detected. In contrast to traditional RNA detection methods our technique allows for the detection within live cells allowing us to sort them based on their miRNA expression and subsequently re-use those cells for follow up experiments. This illustrates the advantage of profiling the RNA expression in live cells with the ability to further study those cells in downstream applications. It also allows for the ability to understand the expression levels of a given miRNA target across a population of cells at the single cell level using flow cytometry as the detection platform. Detecting RNA expression levels in live cancer cells with the ability to utilize the same cells in downstream testing gives researchers the ability to perform experiments which were previously thought to be impossible.

3253F

Mosaic Maternal UPD15 in a Newborn with Complex Heart Defect. *D. Pickering[1], W.G. Sanger[1,2], R.E. Lutz[2], J. Carstens[1], M. Wiggins[1], B.J. Dave[1].* 1) Human Genetics Laboratories, Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE; 2) Genetics Medicine Department, Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE.

Prader Willi syndrome (PWS) results from lack of the paternal copy of chromosome 15q11.2-13.2, most often due to either deletion 15q11.2-q13.2 or maternal uniparental disomy (UPD) of chromosome 15. Microarray analysis using high-density genome-wide SNP arrays detects UPD events as well as copy number changes. Based on copy-neutral segmental or whole chromosome isodisomy detected by SNP array, subsequent methylation studies for paternal and maternal allele are typically performed to confirm the diagnosis. We report here a newborn male referred for microarray with clinical findings of complex congenital heart defect (CHD) involving unbalanced AV canal with hypoplastic arch and dysmorphic facies, small rib cage, and undescended testes. The pregnancy was complicated by intrauterine growth retardation and heart defect detected by ultrasound. Prenatal karyotype and aneuploidy FISH studies were normal. High-density SNP array detected a 49,000,000 base-pair segment of copy-neutral loss of heterozygosity (LOH) at chromosome 15q14-q25. Large interstitial segments of LOH without deletion and restricted to one chromosome typically indicate uniparental heterodisomy. Confirmatory chromosome 15 methylation studies using primer sets covering SNPRN revealed the presence of the methylated maternal band and a significantly diminished unmethylated paternal band suggestive of mosaicism. Concurrent high-resolution chromosome analysis on 20 cells was normal. Subsequent FISH studies analyzing 500 interphase nuclei using SNPRN and 15q telomere probes to rule out possible concomitant low-level mosaicism of trisomy 15 in peripheral blood were also normal. There was no evidence of mosaic trisomy 15 by SNP array copy number analysis. Although fibroblast cells were unavailable, based on the faint paternal band observed in methylation studies, the 49Mb interstitial 15q segmental LOH noted by SNP microarray and the patient's CHD, it is reasonable to assume that a second cell line containing a paternal copy does exist. Additional studies on buccal mucosa tissue are pending. This case highlights the ability of SNP array to identify UPD, even uniparental heterodisomy, nevertheless, ensuing confirmatory studies may be challenging. Understanding the underlying mechanisms of UPD is essential for accurate interpretation, diagnosis, and appropriate recommendations.

3254T

Segments of Homozygosity and Uniparental Disomy Identified by Oligonucleotide SNP Array: Experience from a Reference Laboratory. *J. Wang, L. Ross, L. Mahon, R. Owen, M. Hemmat, B. Wang, M. El Naggar, A. Anguiano, F. Boyar, M. Haddadin, T. Sahoo.* Cytogenetics Laboratory, Nichols Institute, Quest Diagnostics, San Juan Capistrano, CA.

Copy neutral segments with allelic homozygosity (SOH) are frequently identified in cases interrogated by oligonucleotide-SNP microarrays. SOH may be due to parental relatedness, chromosomal recombination or rearrangements. Depending on the genomic context, it may indicate ancestral homozygosity, uniparental isodisomy (UPD), or parental consanguinity. We collected cases from consecutive specimens sent to our clinical laboratory over the past two years. The cases were reported based on the presence of SOH >10 Mb in a single region or >5 Mb in at least two regions. The percentage of the genome encompassed by SOH regions was calculated based on the total coverage of Affymetrix CytoScan™ HD array, which is about 2,700 Mb. Classification of the cases for the degree of parental relatedness was based on the 95% confident interval value recently published (Sund et al., 2013). Of 14,575 cases analyzed by SNP arrays, 67% had a normal result, 23% had results of unclear clinical significance, and 10% had clinically significant copy number variations (CNVs). For each category, about 5-6% of cases had one or more reportable SOH. Of the 872 (6% of 14,575) cases with SOH, 659 (76%) cases were interpreted as arising due to identity by descent (IBD), and 213 (24%) cases were suspected or confirmed as resulting from UPD. For the cases with IBD, an estimate of the inbreeding percentage was estimated, and 5% were suspected to have first degree or closer parental relatedness, 9% second, 19% third, 16% fourth, and 51% fifth. Suspected or confirmed UPD cases were identified involving every single chromosome. Of the 16 cases with SOH from chromosome 15 alone, six cases had isoUPD15 (4 confirmed by methylation studies), another six were confirmed to have mixed iso- and hetero-UPD15. One case had copy number neutral mosaic maternal UPD15 with modified Prader-Willi phenotype. At least ten cases with SOH were confirmed to contribute to autosomal recessive disorders, such as homozygous deletions in the MCPH1 gene (microcephaly), homozygous nonsense mutation in the growth hormone receptor gene (Laron syndrome, pituitary dwarfism), and homozygous nonsense mutation in the RFXAP gene (immune deficiency, Bare lymphocyte syndrome type III). This study demonstrates that the identification of SOH, in addition to CNVs, is much more frequent than previously recognized and often reflects close parental relatedness or unravels UPD in many cases.

3255F

The first case of mos 46,XX,+21,der(21;21)(q10;q10)[16]/ 45,X [14] with clinical evolution and parental karyotypes: 46,XY [20] and 46,XX [20]. *M. Carvalho[1], E. Carvalho[2], M. Montenegro[2], K. Carvalho[3].* 1) Medical Genetics, Universidade Estadual do Ceará-UECE/APAE-CE/UniChristus, Fortaleza, Ceará, Brazil; 2) ICrHCFMUSP; 3) Medical Genetics, Universidade Estadual do Ceará.

Double aneuploidy involving both autosomal and sex chromosomes is very rare. Down's/Turner's mosaic, occurs in about 1 in 2 000 000. We report the first case of Down's/Turner's mosaic with robertsonian translocation. The patient was the first child of non consanguineous parents. It was a female whom born a term, by uneventful cesarean section, weighting 2,220g and with length of 46 cm, without neonatal complications. At birth, the pediatrician made the diagnosis of Down syndrome. At 3 months, in consultation with routine pediatric a heart murmur was heard, so the patient was referred to a cardiologist. Echocardiography was performed and diagnosed congenital heart disease (patent ductus arteriosus). There were performed two surgeries to repair the heart defect, one with 6 months of life and another with 2 years. She also had mild developmental delay. At 3 years old she was examined at our outpatient Genetic unit. The patient showed more clinical findings of Down syndrome than Turner syndrome: low weight and height for age, microcephaly, flat facial profile, upslanting palpebral fissures, epicanthal folds, short nose with depressed nasal bridge, hypotonia with tendency to keep month open and protrude the tongue, short neck, single palmar creases, and prominent ears. Abdominal ultrasound and thyroid function were normal. Cytogenetic analysis of peripheral blood preparations by using G-banding revealed mosaicism with 2 cell lines (mos 46,XX,+21,der(21;21)(q10;q10)[16]/ 45,X [14]). Additional genetic studies (karyotypes) are required to define the cause which probably originated this double aneuploidy with this translocation or isochromosome. So we present the first case related of Down-Turner mosaicism with robertsonian translocation or isochromosome and we review the previous reports of Down-Turner syndrome.

3256T

Characterization of Expression Profile of the CER1 Gene and Two Regulatory Elements in Human Mesenchymal Progenitor and Neural Stem Cells. *X. Hauge, J. Sloan.* Dept Biol & Physics, Kennesaw State Univ, Kennesaw, GA.

Terminal deletions of the short arm of human chromosome 9 (9p-) cause mental retardation, cardiac abnormalities, abnormal genitalia, trigonocephaly and other craniofacial abnormalities. The gene that is of most interest for understanding this syndrome is the cerberus 1 gene (CER1) which plays a role in establishing the anterior-posterior axis in vertebrates. The ortholog of CER1 in mouse binds directly to bone morphogenic proteins (BMPs) and prevent BMPs from binding to their cognate receptors, acting as a BMP inhibitor. The inhibition may slow the ossification of sutures, allowing normal brain growth and development. A deletion of the CER1 gene or of its regulatory elements could cause a premature closure of sutures, leading to trigonocephaly. Little is known regarding the CER1 expression profile during embryonic development in humans due to legal and ethic issues. We seek to investigate the expression profile of the CER1 gene during early embryonic bone development using pluripotent human mesenchymal progenitor cells. We cultured mesenchymal progenitor cells in the osteogenic differentiating medium for various time periods, and then isolated RNA from these cells. To quantify the gene activities of the CER1 and several important osteogenic markers, such as RUNX2 and BSP, quantitative polymerase chain reactions were performed. We also examined the CER1 expression profile in human neural stem cells. Finally, we examined the regulatory functions of an enhancer and a repressor in mesenchymal progenitor cells using a dual luciferase assay. These regulatory elements were cloned from the critical region of 9p- syndrome previously by us. Enhancer clone 5008 demonstrated higher activity in the human mesenchymal progenitor cells than in adult human cell lines.

3257F

Four new patients with maternal UPD20: a phenotype of isolated growth retardation and feeding difficulties. *E.J. Bhoj[1], S. Mulchandani[2], B. Thiel[2], E. Zackai[1], I. Krantz[1], N. Spinner[2,3], L. Conlin[2,3].* 1) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pathology & Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

We present four new patients with maternal uniparental disomy (UPD) for chromosome 20 and propose that maternal UPD20 is associated with a consistent phenotype characterized by prenatal onset growth retardation and feeding difficulties in the absence of dysmorphic features. Uniparental disomy of chromosome 20 is a rare diagnosis. Previously eleven patients with paternal UPD20 have been reported, which were associated with pseudohypoparathyroidism. There are additionally four reports of patients with maternal UPD, all with prenatal and or postnatal growth retardation. However, in three of these patients the phenotype was complicated by mosaic or full trisomy 20, and in those patients psychomotor retardation was also recorded. In previously reported patients, the phenotype of individuals with UPD20 and mosaic trisomy 20 was hypothesized to be a result of the mosaic trisomic cells. This report expands on the description of isolated UPD20 with four new patients identified using genome-wide SNP arrays. All four patients were referred with isolated prenatal-onset growth retardation and feeding difficulties requiring direct gastric feeds. Two of these individuals had a significant increase in length upon growth hormone supplementation. Two of the four patients had complete isodisomy suggestive of monosomy rescue or post-zygotic trisomy rescue and two had iso-hetero disomy consistent with a trisomy rescue mechanism. Chromosome 20 harbors many known and predicted imprinted genes. The phenotype related to maternal UPD20 can be a consequence of known disease-associated imprinted loci (including GNAS) or from other imprinted genes on chromosome 20. All of these patients are undergoing extensive clinical testing to further delineate the physiologic mechanisms affected by UPD20 leading to their phenotype. Given the similarity between our patients and those previously reported, we propose that the growth retardation and feeding difficulties result from matUPD20 rather than mosaic trisomy 20, while the trisomic cells may be the main contributor to psychomotor retardation. This study doubles the number of previously-reported patients with UPD20, allowing for improved description of the phenotype and more accurate prognostication of the course of the disease with and without interventions such as growth hormone supplementation.

3258W

Targetted Amplicon Sequencing of Tongue cancer Genome: Indian Experience. *R. Rawal, S. Bagtharia, K.C. Kothari, S.N. Shukla, A.K. Saxena.* Cancer Biology, Gujarat Cancer & Research Institute, Ahmedabad, India.

Cancer of tongue is one of the most common malignant cancers of the oral cavity predominantly found in India and its subcontinents. The major cause of morbidity and mortality in these patients is due to the high local invasiveness & recurrence, with lymph node and distant metastasis. Lack of predictive and prognostic biomarkers leading to over- or under-treatment of patients poses significant personal and socioeconomic impact. Present study was undertaken to unravel the molecular signatures in patients with tongue cancer with and without habit of tobacco consumption. Study included two patients with cancer of oral tongue (One with habit of tobacco and one without habit of tobacco). The Genomic DNA was extracted from fresh frozen tumour tissue and peripheral blood derived Mono nuclear cell fraction in both cases. The Genomic DNA samples were segregated in to 3 pools (Pool1: Tumour genomic DNA from Patient without habit; Pool2: Tumour genomic DNA from Patient with habit and Pool3: Pooled Genomic DNA from PBMNC of both the patients). Targeted amplicon sequencing was carried out on ION-TORRENT PGM following manufacturer's instruction using 409 gene Ampliseq cancer panel 318 chip. The output data was uploaded by Torrent Variant Caller plugin and analyzed using ION REPORTER and INGENUITY independently. Sequences were compared against hg19 human genome sequence as control. Pool1 and Pool2 results were further compared with constitutional genome (Pool3). Data were filtered using frequency (35%-100%) and coverage (20X-100X) for the exonic region. The other variables used for filtering the data were zygosity, missense mutations etc. Results were plotted Chromosome wise and Venn diagrams were plotted to identify common and unique variants (genes) showing non-synonymous mutations in exonic region. There were 47 unique variants (23genes) in Pool1 and 21 variants (13 genes) in pool2 with 6 common variants (13 genes). The unique allelic variants in tobacco habitues and non habitues suggestive of a tobacco independent mechanism of malignant transformation in tongue tissue.

3259T

The relationship between breast cancer and Tetraspanin8 (TSPAN8) gene polymorphism. *M. Acar[1], T. Yilmaz[1], M. Oznur[1], Z. Unal[1], E. Gunduz[1], M. Gunduz[1,2].* 1) Medical Genetic, Turgut Ozal University Medical Faculty, Ankara, Turkey; 2) Departments of Otolaryngology Head and Neck Surgery, Ankara, Turkey.

Drugs have been improved for breast cancer but there is still a need safe and more effective treatments. Variations in BRCA1, BRCA2, CDH1, PTEN, STK11 and TP53 genes increase the risk of breast cancer. Unfortunately most risk factors of breast cancer in women are not known. TSPAN8 gene (tetraspanin8) is a member of the transmembrane 4 super family. It's cytogenetic localization is at 12q21.1. Most of the coded proteins are surface proteins. These proteins mediate signal transduction in cell development, activation, growth and motility. This gene is expressed in different carcinomas and has been proved to have a relation with ovarian cancer, gastric carcinoma, colorectal cancer, polycystic ovary syndrome, esophageal cancer and diabetes. It is also seen in cell migration and invasion. With up regulation the gene occupies surrounding tissues. It causes metastasis of the lungs and also shortens the life span of rats. Our project's aim is to find a connection between breast cancer and the polymorphism of TSPAN8 rs7961581 C>T. Proving that the polymorphism of TSPAN8 rs7951581 has a role in breast cancer's etiopathogenesis could lead to a means for early diagnosis of this cancer, use of the gene as a biomarker and development of a more effective treatment. For the experimental group, 50 patients diagnosed with breast cancer in 2008-2012 were selected. The DNA was isolated from the tumor tissue. DNA was isolated from blood collecte from the control group of 50 indivudals with no diagnosise of breast cancer. For the TSPAN gene specific primers were designed and the gene was amplified using polymerase chain reaction(PCR). After checking the quality of the amplified products using agarose gel electrophoresis, the gene was cut with Hpy166II restriction enzyme which recognizes the rs7961581 polymorphism. We identified the genotype, confirmed the PCR-RFLP(Restriction Frangment Lenght Polymorphism) results with DNA chain analysis and entered our data into SPSS16.0 and evaluated the relationship between the data using the chi-square test. 23(46%) individuals from the experimental group have the genotype TT,25(50%) CT and 2(4%) CC.14(28%) individuals from the control group have TT,25(50%) CT and 11(22%) CC. There is a recognizable statistical relation between breast cancer and TSPAN8 rs7961581 polymorphism (pearson chi-square=0.015). However a broader investigation must be done to confirm the significance of these findings.

3260F

PPM1D Mutations in Circulating White Blood Cells and the Risk of Ovarian Cancer. *M.R. Akbari[1,2], P. Lepage[3], B. Rosen[4], J. McLaughlin[2,5], H. Risch[6], M. Minden[7], S.A. Narod[1,2].* 1) Women's College Research Institute, Univ Toronto, Toronto, ON, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 3) Genome Quebec Innovation Centre, McGill University, Montreal, Canada; 4) Department of Gynecology-Oncology, Princess Margaret Hospital, Toronto, Canada; 5) Samuel Lunenfeld Research Institute, University of Toronto, Toronto, Canada; 6) Department of Epidemiology and Public Health, School of Public Health, School of Medicine, Yale University, New Haven, USA; 7) Department of Medical Oncology, Princess Margaret Hospital, Toronto, Canada.

One fifth of ovarian cancer cases are estimated to be hereditary and the genes responsible for these are being found continuously and PPM1D was the last one in this series with some unique features. We used deep amplicon sequencing of the PPM1D mutation hot spot (exon 6) in white blood cells' (WBC) DNA of a case-control group and found 20 carriers of truncating mutations among 1,295 ovarian cancer cases and one in 834 controls (OR = 13.1, 95%CI : 1.7 - 97.5, p = 0.0005). All these mutations were mosaic in WBCs and the lifetime risk of ovarian cancer among the female first-degree relatives of PPM1D mutation carriers was similar to the risk among the female first-degree relatives of non-carrier ovarian cancer patients (HR = 1.32; p = 0.78) that suggest PPM1D mutations are probably not inherited. Also the 12-year survival of the PPM1D-positive cases was lower than PPM1D-negative cases (HR = 2.0, 95%CI : 1.21 - 3.39, P < 0.007). The high odds ratio for PPM1D carriers justifies preventive oophorectomy and approximately one percent of all ovarian cancers might be prevented through screening for PPM1D mutations in healthy women regardless of their family history.

3261W

Juvenile myelomonocytic leukemia in six months old boy with breast cancer and leukemia in his Family history. *M. Akouchekian.* Medical Genetics , Tehran University of Medical Science, Tehran, Iran.

Juvenile myelomonocytic leukemia (JMML) classified as a rare childhood cancer and it usually occurs in children younger than 2 years old. It is known that certain medical conditions such as neurofibromatosis type 1 can make a child more likely to develop it. We investigate the chromosomal abnormality of six months old boy diagnosed JMML patient that his grandfather died from breast cancer in age of 57 and the grandfather's father died from leukemia in age of 60. Deletion 7(q) is diagnosed in this boy. Cluster of breakpoints in 7q11 to 7q36, is with two common minimal morphological zones in q22 and in q32-34. Using loss of heterozygocity (LOH) studies and YAC libraries, a 2 to 3 Mb segment in 7q22 has been designated as proximal common deleted area; the q33-34 zone is the consensual area for the distal deletion; LOH studies suggest that a specific mechanism, such as mitotic recombination in bone marrow stem cell leading to homozygosity in both granulocytes and lymphocytes, may be implicated. We also plan to run a whole genome study in this family to investigate the relationship between cancer family histories with this rare childhood cancer. The contribution between NF1 mutation and 7q deletion in JMML disease will also investigate.

3262T

Analyses of genome-wide linkage scan data among families with aggregation of breast and prostate cancer reveals evidence for linkage at 16q21-23. *J. Beebe-Dimmer[1,2], E. Lange[3], K. Zuhlke[4], K. Cooney[4].* 1) Karmanos Cancer Inst, Detroit, MI; 2) Wayne State University, Detroit, MI; 3) University of North Carolina, Chapel Hill, NC; 4) University of Michigan, Ann Arbor, MI.

Purpose: Epidemiologic studies have shown a co-clustering of breast and prostate cancer suggesting that there are germline variants that increase the risk of both hormonally-driven neoplasms. Mutations in BRCA1 and BRCA2 genes may explain a small portion of the observed occurrence of both breast and prostate cancer within families. The current investigation focuses on the delineation of chromosomal regions which may harbor new genes that play a role in the aggregation of breast and prostate cancer among first degree family members. Methods: A genome-wide linkage scan was conducted on 50 families participating in the University of Michigan Prostate Cancer Genetics Project. All families had at least 2 first-degree relatives diagnosed with prostate cancer and at least one female relative diagnosed with breast cancer in a first-degree relationship with one of the participating prostate cancer cases. Genome-wide multipoint nonparametric linkage analyses for the combined phenotype of prostate and breast cancer were performed using the software Merlin. Results: The strongest evidence for linkage was detected at 16q22 (LOD=3.07 at rs722579), a region previously reported to be in linked to prostate cancer. This region contains several interesting candidate genes including known prostate cancer tumor suppressor genes WWOX and ATBF1, as well as BCAR1, a gene involved in a number of critical carcinogenic processes including cell migration, growth, and differentiation. Conclusions: Next generation sequencing of the genes in this region in our linked families are in progress to identify new mutations that explain clustering of prostate and breast cancer in these families and provide us new information on shared genetic pathways between these two common cancers.

3263F

Diversity of inherited damaging mutations in all breast cancer genes in three series of breast cancer patients: young onset, triple-negative and those unselected for family history, age at diagnosis or hormone receptor status. *G. Bernier[1], J. Mandell[1], T. Walsh[1], S. Casadei[1], M. Lee[1], E. Swisher[2], M.C. King[1].* 1) University of Washington, Division of Medical Genetics, Seattle, WA; 2) University of Washington, Department of Obstetrics and Gynecology, Seattle, WA.

In addition to *BRCA1* and *BRCA2*, inherited mutations in multiple other genes predispose to breast cancer. Many of these genes are in the BRCA-Fanconi Anemia complex. We used BROCA, a targeted genomic capture and massively parallel sequencing approach, to identify all classes of mutations in all known breast cancer genes. Only truncations, complete gene deletions, splice mutations known to lead to a mutant message, and missenses demonstrated functionally to be damaging were counted. Study participants were three series of breast cancer patients: those sequentially enrolled from two local cancer centers and not selected for age at diagnosis, family history or hormone receptor status (SBC; n=253); those with invasive breast cancer diagnosed at age 40 or younger (YBC; n=533); and those with triple negative (i.e. estrogen receptor, progesterone receptor, and HER2 negative) breast cancers (TNBC; n=344). SBC patients had not been commercially tested for *BRCA1* and *BRCA2* at time of enrollment so all genes were included in their analysis. YBC and TNBC patients with positive commercial test results for *BRCA1* or *BRCA2* were not recruited, so only genes other than *BRCA1* and *BRCA2* were included in the YBC and TNBC analyses. Of the SBC patients, 7.1% (18/253) carried damaging mutations in any targeted gene: 3 in *BRCA1*, 4 in *BRCA2*, 4 in *CHEK2*, 2 in *PALB2*; and 1 each in *TP53*, *ATM*, *BARD1*, *BAP1*, and *NBN*. Of the YBC patients, 7.7% (41/533) carried damaging mutations in genes other than *BRCA1* or *BRCA2*: 21 in *CHEK2*, 4 in *PALB2*; 3 each in *ATM*, *BARD1*, *BRIP1*, *RAD51D*; and 1 each in *TP53*, *NBN*, *STK11*, and *MRE11A*. Of the TNBC patients, 7.6% (26/344) carried mutations in genes other than *BRCA1* and *BRCA2*: 8 in *PALB2*, 7 in *BARD1*, 3 each in *CHEK2* and *RAD51D*, 2 each in *BRIP1* and *RAD51C*, and 1 in *Abraxas*. The mutational spectrum of these patients was highly heterogeneous, with 53 different mutations in 13 different genes. Across all series, the most common alleles were *CHEK2* c.1100delC (8 patients, including one homozygote) and *PALB2* p.Y1183X (4 patients), consistent with the large contribution of northern European ancestry to the Pacific Northwest population. TNBC patients included surprisingly large numbers with *PALB2* or *BARD1* mutations. Carriers of mutations in all genes met NCCN criteria for *BRCA1*/*BRCA2* testing. We recommend adapting the current genetic testing guidelines for breast cancer patients to include multi-gene evaluations, such as BROCA, for those that meet testing criteria.

3264W

Germline Copy Number Variants as Genetic Risk Factors for Familial Colorectal Cancer Type X. *D.D. Buchanan[1], M. Clendenning[1], C. Rosty[1,2,3], M.D. Walsh[1,4], S. Parry[5,6], A.K. Win[7], J.L. Hopper[7], M.A. Jenkins[7], Colon Cancer Family Registry.* 1) Cancer & Population Studies Group, QIMR, Brisbane, Queensland, Australia; 2) School of Medicine, University of Queensland, Brisbane, Queensland, Australia; 3) Envoi Pathology, Brisbane, Queensland, Australia; 4) Sullivan and Nicolaides Pathology, Brisbane, Queensland, Australia; 5) New Zealand Familial Gastrointestinal Cancer Service, Auckland Hospital, Auckland, New Zealand; 6) Department of Gastroenterology, Middlemore Hospital, Auckland, New Zealand; 7) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Parkville, Victoria, Australia.

Background: Only one-third of the familial risk of colorectal cancer (CRC) is explained by variants in known CRC susceptiblity genes. Multiple-case CRC families that fulfil the Amsterdam-I criteria for Lynch syndrome but do not carry mismatch repair (MMR) gene mutations and demonstrate no evidence of MMR-deficiency in their tumors are referred to as meeting the Familial Colorectal Cancer Type X (FCCTX) criteria. Copy number variants (CNVs; deletions or duplications of DNA segments), have been detected in and around cancer predisposing genes such as the MMR genes, BRCA1 and BRCA2; however, the role of germline CNVs in CRC susceptibility remains poorly understood. The aim of this study was to identify rare or novel germline CNVs that predispose to CRC in families that meet the FCCTX criteria. Methods: Blood-derived DNA from CRC- and polyp-affected and unaffected relatives from 15 multiple-case CRC families fulfilling the FCCTX criteria and from 16 unrelated controls from the Australasian Colorectal Cancer Family Registry were genotyped using the Illumina HumanOmni2.5-8 SNP array. CNVs were identified using Partek GS v6.6 from LogR ratios. Candidate CNVs were identified that encompassed genes, segregated with CRC, were absent in unaffected relatives and unrelated controls, and either not previously reported in the Database of Genomic Variants (DGV) or were reported as rare. Results: Forty CRC-affected (mean age at diagnosis =56.7 years ± standard deviation (SD)= 11.7 years) , 33 early-onset polyp-affected (mean =45.6 years ± SD= 11.3 years) and 15 unaffected family members of 15 FCCTX families (minimum 2 CRC-affected and 1 unaffected relative/family) were genotyped. The number of CNVs that were present in all CRC-affected individuals in a family but not detected in the controls or seen in the DGV ranged from 5 to 13 deletions per family and 14 to 30 duplications per family. Candidate CNVs included a duplication that lies 5' to the PTPRJ gene (previously reported as a likely CRC-susceptibility gene) and a deletion encompassing the CTNNA3 gene on 10q21.3. Conclusion: We have identified several rare or novel germline CNVs that segregate with CRC-affected members in families that meet the FCCTX criteria. Future studies will extend the study to genotype additional FCCTX families, test a large series of CRC cases and controls for variants in these genes and further characterise the effects of these candidate CNVs on gene expression.

3265T

HLA-DPB1 polymorphisms and cervical squamous cell carcinoma risk. *T. Chang[1], Y. Yang[2,3], Y. Lee[1,4,5], T. Chen[2], W. Lin[1], S. Chang[1].* 1) Med Res Dept, Mackay Memorial Hosp, New Taipei, Taiwan; 2) Department of Gynecology and Obstetrics, Mackay Memorial Hospital, Taipei, Taiwan; 3) Department of Gynecology and Obstetrics, Taipei Medical University, Taipei, Taiwan; 4) Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan; 5) Pediatrics, Taipei Medical University, Taipei, Taiwan.

Cervical cancer is a multifactorial disease and infection by oncogenic human papillomaviruses (HPVs) represents the major environmental risk factor. However, the mere presence of HPVs is not enough for cervical cancer development and host immunogenetic background may play an important role. Variations in human major histocompatibility genes may alter the efficiency of immune response to HPV antigens and have been implicated in the risk for cervical cancer. The aim of this study is to examine the role of the human leukocyte antigen (HLA)-DPB1 gene in cervical cancer susceptibility. We used high-resolution methods to genotype HLA-DPB1 in 344 women with cervical squamous cell carcinoma (CSCC) and 250 age/sex matched healthy controls. The presence and genotypes of HPV in CSCC patients were determined by PCR. We found the HLA-DPB1*13:01 was associated with an increased risk of both CSCC (OR = 1.68, 95% CI 1.17-2.43) and HPV-16 positive CSCC (OR = 1.94, 95% CI 1.26-2.98). However, the significance remained only in HPV-16 positive CSCC patients after Bonferroni correction (*Italic Text*$P_{\text{Subscript Text}}c$ = 0.02). In conclusion, our results suggest that HLA-DPB1*13:01 is involved in the genetic susceptibility to HPV-16 positive CSCC in the Taiwanese population. Further studies with larger cohort are required to confirm the role of DPB1*13:01 in the development of CSCC.

3266F

Replication of cervical cancer susceptibility loci identified in genome-wide association study in a northern Swedish population.
D. Chen[1], J. Hammer[1], D. Lindquist[2], U. Gyllensten[1]. 1) Department of Immunology, Genetics and Pathology, Uppsala University, Sweden, Uppsala, Sweden; 2) Department of Radiation Science, Umeå University, Sweden.

In a genome-wide association study (GWAS) we have previously identified and performed the initial replication of three novel susceptibility loci for cervical cancer: rs9272143 upstream of *HLA-DRB1*, rs2516448 adjacent to MHC class I polypeptide-related sequence A gene (*MICA*) and rs3117027 at *HLA-DPB2*. The risk allele T of rs2516448 is in perfect linkage disequilibrium with a frameshift mutation (A5.1) in *MICA* exon 5, which results in a truncated protein. To validate these associations in an independent study and evaluate effect modification by age of onset and tumor stage, we genotyped the single-nucleotide polymorphisms at rs2516448, rs9272143, rs3117027 and the *MICA* exon 5 microsatellite in 961 cervical cancer patients and 1725 cancer-free control subjects from northern Sweden. Association between each variant and cervical cancer risk was estimated by logistic regression analysis. The C allele of rs9272143 conferred protection against cervical cancer (odds ratio [OR]=0.73, 95% CI=0.65-0.82; $P=1.8\times10^{-7}$), whereas the T allele of rs2516448 increased the susceptibility to cervical cancer (OR=1.36, 95% CI=1.21-1.52; $P=1.8\times10^{-7}$), with the same association shown with *MICA-A5.1*. We also identified protective effects of the *MICA-A4* (OR=0.79, 95% CI=0.67-0.93; $P=4.1\times10^{-3}$) and *MICA-A5* (OR=0.60; 95% CI=0.50-0.72, $P=5.2\times10^{-8}$) alleles. The direction and the magnitude of these associations were consistent with our previous findings. None of the variants studied showed heterogeneity by age. No association was observed between rs3117027 and risk of cervical cancer (OR=0.99, 95%CI=0.87-1.11 for allele A; P=0.80). Our results support the role of *HLA-DRB1* and *MICA* in the pathogenesis of cervical cancer.

3267W

Gene Expression Profiling of a Single Laser Capture Microdissected (LCM) Cell.
S. Chu, M. Gauthier, K. Schmidt, S. Patel. Life Technologies, South San Francisco, CA.

Laser Capture Microdissection (LCM) is a technique providing a rapid and reliable method to procure purified cell populations, either as multiple single cells or a group of cells, from a heterogeneous tissue sample allowing a targeted approach to genomic profiling. When analyzing gene expression profiles from a group of cells, the average profile may not be a true representation of the many different profiles that could exist even in a pure cell population (e.g., in different states of growth, differentiation, or activation). As a result, the transcriptional variability of individual cells and any insight into the relationship between specific genes in single cells are lost. To fully understand the complexity of tissue and cellular heterogeneity, it is necessary to measure molecular signatures at the single cell resolution. Here we developed a workflow for gene expression profiling of a single LCM cell that can also be applied to small number of LCM cells. The ArcturusXT™ LCM system was used to harvest single cells from frozen human tumor tissue samples which were then processed and analyzed with a quantitative RT-PCR assay platform.

3268T

Loss in the function and homeostasis of the PDE family might inflict on the cAMP and cGMP signaling and lead to the development of prostate cancer.
R.B. de Alexandre[1,2], A. Horvath[1], A.D. Manning[1], N. Hatipoglu[1], F. Kardauke[2], D. Carraro[3], F. Soares[3], M. Nesterova[1], C. Stratakis[1], F.R. Faucz[1,2]. 1) Section on Endocrinology & Genetics, PDEGEN, NICHD, NIH, Bethesda, MD, USA; 2) Laboratory of Molecular Genetics, NIMA, PPGCS, Pontificia Universidade Catolica do Parana, Curitiba, Brazil; 3) Laboratory of Genomics and Molecular Biology - A.C. Camargo Hospital, São Paulo, Brazil.

The phosphodiesterases (PDEs) are a family of 11 intracellular isozymes coded by 21 different genes. They are responsible for hydrolyzing cAMP and cGMP to their respective 5'-nucleoside monophosphate. Some PDEs are specific for the hydrolysis of cAMP or cGMP, while others have mixed specificity. Each PDEs has a different ability to control each of these messengers concentration in different tissues, provoking different cellular responses and functions. Cyclic AMP and cGMP intracellular levels depend on the balance between their synthesis and degradation. Variants in PDEs have been associated with different disorders. The first PDE to be implicated in the predisposition of prostate cancer (PCa) was the PDE11A, whose inactivating mutations have been reported to be frequent in its patients. Further, higher levels of cAMP have been measured in prostate cancer compared to the normal tissue. In addition, cAMP and cGMP signaling is involved in PCa cell growth and modifies androgen receptor effects. We sequenced 16 different PCa tumor DNAs on SOLiD4 platform, after targeted enrichment for the coding parts of 196 genes, including the family of the phosphodiesterases, adenylate and guanylate synthases, all genes involved in cAMP and cGMP pathways and other genes related to endocrine tumors. After alignment and variation calling, Sanger sequencing confirmation and biostatistics analysis (through comparison patients with controls of the 1000 Genome Project), we were able to select 22 different SNPs associated with the disease. These significant SNPs where found to be located in the PDE1C, PDE2A, PDE4B, PDE5A, PDE6A, PDE6B, PDE6C, PDE8A, PDE8B and PDE11A genes. Moreover, it was found two novel mutations confirmed by Sanger located in the PDE1A and PDE7B genes. Furthermore, th previously-ere are still seven novel mutations with large chances to be real positives spread through the PDEs awaiting to be confirmed by Sanger sequencing. These results demonstrate that there might be a set of disarrangement in the intracellular levels of cyclic AMP and cGMP altering its balance. This data also establish a link between the PDEs and PCa, suggesting that, not as what was report previously with only the PDE11A, but that the whole family might be involved with a higher susceptibility to PCa. However, further investigations at protein level, trying to identify what exactly each alteration might influence in the final phenotype, are needed to be accomplished previously publication.

3269F

Genetic variants in miR-499a, miR-938 and miR-1206 are associated with gastric cancer in Europeans from the EPIC-EURGAST study. *Y. Espinosa-Parrilla[1, 2], X. Muñoz[3, 4], I. Torruella-Loran[1], C. Bonet[5], N. Garcia[4, 5], E. Riboli[6], C.A. Gonzalez[5], N. Sala[4, 5].* 1) Ciencies Experimentals i de la Salut (CEXS-UPF), Institut de Biologia Evolutiva (UPF-CSIC), Barcelona, Barcelona, Spain; 2) Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile, Santiago de Chile; Chile; 3) Hereditary Cancer Program; Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain; 4) Molecular Epidemiology Group, Translational Research Laboratory, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain; 5) Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, ICO-IDIBELL, Barcelona, Spain; 6) Department of Epidemiology and Public Health, Imperial College London, London, United Kingdom.

MicroRNAs (miRNAs) are post-transcriptional gene regulators that participate in diverse biological pathways and may act as either tumor suppressors or oncogenes. Single nucleotide polymorphisms (SNPs) in miRNAs may contribute to cancer development causing changes in either miRNA expression or function. To look for a possible contribution of miRNA genetic variants to gastric cancer (GC) susceptibility we selected 40 SNPs potentially functional because of their location in the seed (4), mature (5) or precursor (31) sequence of 40 miRNAs and genotyped them in 1284 controls matched to 365 incident GC cases with different histological and tumor location phenotypes (European Prospective Investigation into Cancer and Nutrition (EPIC) cohort). Logistic regression analysis under the log-additive model showed that rs2114358T/C, located in the precursor sequence of miR-1206, was associated with the non-cardias localization of the adenocarcinoma (p value= 0.0093, OR (95% CI)= 0.73 (0.58-0.93)); this SNP could affect the dosage of the corresponding miRNA by altering its biogenesis or stability. Furthermore, two out of the four SNPs located in miRNA seed regions were associated with GC: rs12416605C/T in miR-938 associated with the diffuse phenotype (p value= 0.0281, OR (95% CI)= 0.70 (0.51-0.97)) and rs3746444T/C in miR-499a-3p associated with the cardias CG subtype (p value= 0.0308, 0.64 (0.42-0.98)). To investigate if differential gene regulation could be underlying the genetic association of GC with these SNPs in miR-938 and miR-499a we predicted target genes for the different miRNA alleles of both variants using TargetScan. We found a very low overlapping of predicted genes for the different alleles (three common target genes for the two alleles of rs12416605 and no common target genes for the two alleles of rs3746444) as well as a significant decrease in the number of genes predicted to be regulated by the minor alleles of rs12416605 (115 and 75 predicted target genes for the C and T alleles, respectively) and rs3746444 (259 and 9 predicted target genes for the T and C alleles, respectively). These results indicate that specific miRNA allele variants are associated with GC susceptibility in European populations probably influencing either miRNA dosage or the number and spectrum of target genes regulated by these miRNAs.

3270W

Bladder cancer susceptibility variants within CCNE1 are associated with mRNA expression of an alternative splicing form. *Y.P. Fu[1], I. Kohaar[1], W. Tang[1], P. Porter-Gill[1], J.D. Figueroa[2], M. Garcia-Colsas[2,3], N. Chatterjee[2], N. Malats[4], M. Kogevinas[5-8], D. Baris[2], D. Albanes[2], A. Carrato[9], A. Tardón[7,10], C. Serra[11], R. García-Closas[12], J. Lloreta[13], A. Johnson[14], M. Schwenn[15], M.R. Karagas[16], A. Schned[16], W.R. Diver[17], S.M. Gapstur[17], E. Jacobs[17], J. Virtamo[18], F.X. Real[11,19], S.J. Chanock[1], J.F. Fraumeni Jr.[2], D.T. Silverman[2], N. Rothman[2], L. Prokunina-Olsson[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 3) Division of Genetics and Epidemiology, The Institute of Cancer Research, London SW7 3RP, UK; 4) Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center, Madrid 28029, Spain; 5) Centre for Research in Environmental Epidemiology (CREAL), Barcelona 08003, Spain; 6) Municipal Institute of Medical Research, Barcelona 08003, Spain; 7) CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona; 8) National School of Public Health, Athens 11521, Greece; 9) Ramón y Cajal University Hospital, Madrid 28034, Spain; 10) Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo 33003, Spain; 11) Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona 08003, Spain; 12) Unidad de Investigación, Hospital Universitario de Canarias, La Laguna 38320, Spain; 13) Hospital del Mar-Institut Municipal d'Investigació Mèdica (IMIM), Universitat Pompeu Fabra, Barcelona 08003, Spain; 14) Vermont Cancer Registry, Burlington, Vermont 05401, USA; 15) Maine Cancer Registry, Augusta, Maine 04333, USA; 16) Dartmouth Medical School, Hanover, New Hampshire 03755, USA; 17) Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA; 18) National Institute for Health and Welfare, Helsinki 00271, Finland; 19) Epithelial Carcinogenesis Group, Molecular Pathology Programme, Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid 28029, Spain.

BACKGROUND: SNP rs8102137 located 6 Kb upstream of the cyclin E1 gene (CCNE1) on chromosome 19q12 is associated with bladder cancer in a genome-wide association study (GWAS). CCNE1 regulates cell cycle and thus is a strong candidate gene for cancer susceptibility. METHODS: Based on the bladder GWAS and the 1000 Genomes Project reference panel (phase 1 version 3, 2012 March revised), we imputed 4,650 SNPs within +/- 200 Kb of CCNE1 region (GRCh37, chr19: 30,102,901-30,515,215) among 3,532 bladder cancer cases and 5,120 controls of European ancestry. CCNE1 mRNA expression was evaluated by RNA-Sequencing in 6 pairs of tumor-normal bladder tissue samples and with custom-designed TaqMan assays in 42 muscle-invasive bladder tumor and 42 adjacent normal tissue samples. Logistic regression and multivariable linear models were used to test for the association between SNPs and bladder cancer risk, and between SNPs and mRNA expression. All models were assumed the log-additive SNP effects and adjusted for study sites, age, gender, and smoking habits when applicable. RESULTS: A total of 700 well-imputed and 76 GWAS-genotyped SNPs in the CCNE1 region were used for analyses. We identified an imputed SNP rs7257694, a coding synonymous variant (Ser390Ser) in CCNE1 gene, was in strong linkage disequilibrium (LD) with the GWAS marker rs8102137 (r^2=0.763, D'=0.995). With validation through actual genotyping, rs7257694 showed a per-allele odds ratio (OR) of 1.10 (95%CI=1.03-1.17) for bladder cancer risk, which was comparable to rs8102137 (OR=1.12, 95%CI=1.05-1.20) in the same samples. In addition to the main CCNE1 transcript, RNA-sequencing revealed 2 splicing forms of CCNE1 with deletion of exons 5 or 7. TaqMan assays showed significantly higher mRNA expression in bladder tumors than in adjacent normal bladder tissues of total CCNE1 (p=3.97E-12) and of a transcript without exon 5 (p= 6.15E-10). The mRNA expression of CCNE1 transcript without exon 7 was significantly associated with both rs8102137 and rs7257694 in normal and tumor bladder tissues (all $p<0.05$). CONCLUSION: In addition to the GWAS variant rs8102137, we identified a CCNE1 coding synonymous SNP rs7257694 as a variant associated with bladder cancer risk and with mRNA expression of the CCNE1 transcript without exon 7. Future studies are warranted to validate our findings in additional samples and investigate the possible functional mechanisms of this genetic association in bladder cancer.

3271T

Few mutations in known high-risk gastric cancer susceptibility genes in Chinese gastric cancer kindreds. *A.M. Goldstein[1], N. Hu[1], L.-J. He[2], X.-Y. Han[3], J. He[1,4], M. Rotunno[1], M. Malasky[1,4], H. Su[1], L. Wang[1], C. Wang[1], L. Burdett[1,4], B. Hicks[1,4], K. Jones[1,4], J. Boland[1,4], A. Hutchinson[1,4], M. Yeager[1,4], T. Ding[3], C. Giffen[5], M.A. Tucker[1], S.J. Chanock[1], M. Lee[6], P.R. Taylor[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, USA; 2) YangCheng Cancer Hospital, YangCheng, Shanxi, PR China; 3) Shanxi Cancer Hospital, Taiyuan, Shanxi, PR China; 4) Cancer Genetics Research Laboratory, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA; 5) Information Management Services, Inc., Silver Spring, MD, USA; 6) Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland, USA.

Gastric cancer (GC) causes more than 700,000 deaths each year. Familial GC results from the complex interplay of genetic and environmental factors. Several high-risk susceptibility genes for familial GC and related disorders have been identified, but most of these genes have not been examined in families from high-incidence regions such as the Taihang Mountains of North-Central China, a region with some of the highest rates reported for GC and esophageal cancer (EC). We searched the Human Gene Mutation Database (HGMD) and COSMIC to select germline disease-causing genes in familial GC. To allow for related phenotypes, we also included susceptibility genes for gastrointestinal polyps and colorectal cancer. Fourteen genes (*CDH1, MET, MUTYH, TP53, MLH1, MSH2, MSH6, PMS1, PMS2, APC, CTNNA1, SMAD4, BMPR1A, MADH4*) were selected. Families with 3 or more GC and/or EC cases (including ≥2 GC patients) from an ongoing family study were eligible for this study. Twelve eligible families were selected for exome sequencing; 8 families had at least 3 GC cases. The goal of our study was to determine whether high-risk GC-related susceptibility genes are important in GC families from the Taihang Mountains. Twenty-six patients with available DNA were exome sequenced. We interrogated the exome sequencing data for rare co-segregating mutations in the 14 GC-related susceptibility genes. Rare co-segregating variants were found in only 3 (*CDH1, MSH2, MUTYH*) of the 14 genes in two GC families. An *MSH2* substitution (rs63750716) was found in 2 patients (1 GC; 1 EC) with available DNA from a 4 case family (2 GC; 2 EC); it was seen at 0.3% frequency in Asian 1000 Genomes subjects and predicted to be neutral (MutationAssessor). Variants in *CDH1* and *MUTYH* were seen in a single family with 3 GC patients, but, each variant was seen in only 2/3 GC cases. The *CDH1* substitution was novel; the *MUTYH* substitution was reported in 0.5% of Asian subjects in 1000 Genomes; both variants were predicted to have medium functional impact (MutationAssessor). Further study is needed to determine whether these rare variants increase risk for GC in these two families. Overall, examination of 14 GC-related susceptibility genes revealed rare co-segregating variants in two GC kindreds from a high-risk region of China. No co-segregating mutations were observed in the remaining 10 familial GC kindreds suggesting that other genes and/or non-genetic factors are involved in familial GC risk in this region.

3272F

Transcriptome analysis reveals novel gene coding variants and fusion transcripts in infant acute lymphoblastic leukemia. *A.M. Gout[1,2], R.S. Kotecha[1,3,4], J. Ford[1], R.W. Francis[2], A.H. Beesley[1], M.N. Cruickshank[1], U.R. Kees[1].* 1) Division of Children's Leukaemia and Cancer Research, Telethon Institute for Child Health Research, University of Western Australia Centre for Child Health Research, Perth, Australia; 2) Division of Bioinformatics and Biostatistics, Telethon Institute for Child Health Research, University of Western Australia Centre for Child Health Research, Perth, Australia; 3) Department of Hematology and Oncology, Princess Margaret Hospital for Children, Perth, Australia; 4) School of Pediatrics and Child Health, University of Western Australia, Perth, WA, Australia.

Acute lymphoblastic leukemia (ALL) occurring in the first year of life is rare, accounting for 2-5% of pediatric ALL cases. Infant ALL is distinguished by unique clinical and biological characteristics, with an aggressive course following a short latency period. The mixed lineage leukemia (MLL) gene, located on chromosome 11q23, is involved in 80% of cases. Currently, 79 different MLL-fusion partner genes have been molecularly characterized with t(4;11), t(9;11) and t(11;19) the most frequent translocations in infant ALL. In this study we focused on MLL-rearranged infant ALL where diagnosis occurred at < 92 days. At present, the outcome for these infants remains poor with 26% five-year survival. Given the advent of next generation sequencing, further insight into the biology of the disease may identify potential targets for novel therapies and ultimately improve outcome. We performed RNA-sequencing (Illumina, 100bp paired end) on six primary patient infant ALL samples: three patients had the t(4;11) translocation, one patient had a t(11;19) translocation and a pair of monozygotic twins with a rare MLL-translocation partner gene, t(1;11). Upon alignment of sequence reads to reference sequences (including genome, splice junction and transcriptome sequences), a pipeline utilizing Genome Analysis ToolKit (GATK) functions, Annovar, SIFT, Polyphen2, dbSNP and COSMIC was used to process the sequence alignments and annotate single nucleotide variants (SNVs). Ingenuity Pathway Analysis was performed on gene lists associated with predicted damaging SNVs. This revealed an over-representation of cancer-associated genes harboring damaging SNVs that were shared among all six infant ALL patient samples. This gene set involves multiple genes previously reported to be involved with hematological neoplasia that may represent novel therapeutic targets for treatment. We also sought gene fusions in these datasets using FusionFinder that led to the identification of a number of novel putative gene fusions involving known oncogenes. Further studies are required to determine the role of these SNVs and gene fusions in infant leukemogenesis.

3273W

Chemotherapy-Induced Peripheral Neuropathy and Cognitive Dysfunction: Role of Genetic Variation. *K. Holohan[1,2], Y. Wang[3], B.C. McDonald[3,4], S.K. Conroy[2,3], D.J. Smith[3], J.D. West[3], K. Nho[3], S. Kim[3], A.J. Saykin[1,3,4].* 1) Medical and Molecular Genetics, Indiana University-Purdue University Indianapolis, Indianapolis, IN; 2) Training in Research for Behavioral Oncology and Cancer Control Program, Indiana University School of Nursing, Indianapolis, Indiana; 3) Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana; 4) Indiana University Melvin and Bren Simon Cancer Center, Indiana University School of Medicine, Indianapolis, Indiana.

Although the cerebral effects of pain have been studied via neuroimaging in many disease models, this topic has not yet been investigated with chemotherapy-associated peripheral neuropathy and cognitive dysfunction. Previous studies have reported cognitive dysfunction measured by decreased neuropsychological test performance, increased cognitive complaints, and alterations in neuroimaging measures, and treatment with paclitaxel and docetaxel has been strongly associated with increased incidence and severity of peripheral neuropathy symptoms (PNS). Significant genetic modifiers have been identified for both of these adverse events, but possible correlations between them have not been explored. We hypothesized that chemotherapy-induced peripheral neuropathy and cognitive dysfunction may be associated, and that these measures may be modified by genes *EPHA5*, *FGD4*, or *FZD3*, which have been previously associated with chemotherapy-induced peripheral neuropathy. We examined this hypothesis in a prospective cohort of breast cancer patients treated with (Ctx+, n=27) and without (Ctx-, n=26) chemotherapy and healthy controls (HC, n=26) studied at baseline (BL; post-surgery, before chemotherapy) and one month after treatment completion (1M) or yoked intervals. Neuroimaging included pulsed arterial spin labeling MRI to obtain cerebral blood flow (CBF), cognitive complaints were assessed using the Multiple Ability Self-Report Questionnaire (MASQ), PNS were assessed using the 9-item FACT/GOG-Ntx subscale, and all subjects were genotyped using the Illumina OmniExpress BeadChip. Neuroimaging analysis of Ctx+ patients using SPM8 indicated that PNS at 1M covarying for BL were positively associated with CBF change from BL to 1M in the left anterior cingulate region ($p<0.001$ uncorrected), which has been associated with pain processing. This measure of CBF change was set-tested in PLINK for association with *EPHA5*, *FGD4*, and *FZD3*. The *EPHA5* gene set was significantly associated with CBF change, as well as with PNS at 1M covarying for BL ($p<0.03$). Further analysis indicated that one of the most significant SNPs, rs13149946, was also associated with the language subdomain of the MASQ ($p=0.028$), suggesting that this gene may be involved in both pain processing and cognition, and could be a potential target to address both of these issues. Future investigation should test *EPHA5* for association with other measures of cognitive processing and replication in other cohorts.

3274T

Occult hepatitis B and carcinogenic markers in chronic hepatitis C infection. *R. issa.* molecular diagnostics, institute of genetic engineering, cairo, Egypt.

The presence of HBV-DNA in the patient's serum without detectable HBV surface antigen (HBsAg.) called occult infection. Detection of occult HBV infection in chronic hepatitis C virus patients was investigated by using qualitative PCR. Co-infection with occult HBV in chronic HCV patients increases the risk for progression to hepatocellular carcinoma (HCC).Detection of CD45-CD90+ as a biomarker in HCC patients by flow cytometry. We searched for serum HBV DNA in 30 patients with histologically verified HCV-related chronic liver disease, in addition to 10 healthy control subjects collected at National Liver Institute in Shebin El-Kom, Monofiya University, Egypt from January, 2010 to October, 2010. Off 40 patients, the sera of 9 (15.0%) were positive for HBV DNA by the different PCR assays, documenting an occult HBV infection. It found that 5 patient samples are positive for HBV DNA (Surface gene) (12.5%) of total 40 patient samples, also 3 patient samples are positive for HBV DNA (X gene) (7.5%) of total 40 patient samples, and only one patient sample was positive HBV DNA (core gene)(2.5 %) of total 40 patient samples. Only two samples from the nine positive samples were positive for both X-gene and Surface gene. In conclusion these data suggest that occult HBV infection may have clinical significance in chronic hepatitis C patients. Keywords: Biomarker, Flowcytometery, Hepatocellular carcinoma, Occult HBV. Introduction Occult hepatitis B infection (OBI) has been recognized for nearly 20 years. However, with the improvements in sensitivity of serological and genomic amplification assays, the frequency and attention given to this nearly silent form of the infection has been growing. Because low or very low viral load is part of the definition of OBI, molecular data are scarce, and, because most patients affected are asymptomatic, little clinical data, in particular histological, is available (Allain J.P. 2005). Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide ranging between 3% and 9% annually Velazquez et al., 2003. In Egypt, HCC reports to account for about 4.7% of chronic liver disease patients. HBV and HCV infections are strongly associated with liver cirrhosis and HCC Rahman et al., 2001. Africa ., 2008).

3275F

Fine mapping of the Finnish hereditary prostate cancer linked loci at 2q37 and 17q12-q22. *V.H. Laitinen[1], T. Rantapero[1], D. Fischer[2], E.M. Vuorinen[1], T.L.J. Tammela[3], T. Wahlfors[1], J. Schleutker[1, 4].* 1) Institute of Biomedical Technology/BioMediTech, University of Tampere and Fimlab Laboratories, Tampere, Finland; 2) School of Health Sciences, University of Tampere, Tampere, Finland; 3) Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland; 4) Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Turku, Finland.

Linkage studies of Finnish hereditary prostate cancer families have revealed a strong connection between prostate cancer and two chromosomal regions, 2q37 and 17q12-q22. The 17q12-q22 locus is the same previously reported in the US population from Michigan and recently, both of these genomic regions have been detected in multinational GWS analyses. Although the recently identified G84E variant in the *HOXB13* gene at 17q21-q22 has been detected at a high frequency in Finnish prostate cancer patients, this variant alone does not explain the observed linkage to 17q12-q22. Neither is the candidate gene at 2q37 known for the present. To explain the linkage, we therefore screened these two regions of interest by next-generation sequencing (NGS). The NGS produced over 100.000 unique sequence variants which were carefully filtered in a multistep prioritization process, yielding a subset of 58 putative prostate cancer associated variants co-segregating with the disease in the analyzed families. These variants were then validated in 1293 Finnish prostate cancer cases and in 923 controls with Sequenom MassARRAY system. Statistical analyses revealed 13 SNPs in 7 genes that were significantly associated with prostate cancer. Two novel susceptibility alleles were identified in the *ZNF652* gene (17q21.3), one allele in the *HDAC4* gene (2q37.2) and one in the *EFCAB13* gene (17q21.3). As a complementary approach, we investigated the impact of prostate cancer associated sequence variants on the regulation of genes located within the two linked regions. This targeted expression quantitative trait loci (eQTL) analysis was performed by combining the NGS data with transcriptome data obtained from RNA sequencing. In total, five candidate eQTLs were found at 2q37 and 1881 candidates at 17q12-q22 before multiple testing adjustment. The novel prostate cancer associated SNPs identified in this study may be exploited in cancer risk assessment, and they further support the suggested role of *ZNF652* as a prostate cancer candidate gene. Moreover, this is the first targeted resequencing project studying the locus on 2q. The regulatory elements discovered by eQTL mapping provide new insights into the complex genetic events contributing to prostate cancer predisposition.

3276W

Novel *EZR/ERBB4* fusion gene found in follicular variant of papillary thyroid cancer. *S. Lee[1], H.-G. Jee[2], K. Lee[2], J.-S. Seo[1].* 1) Biomedical Sci, Seoul National Univ, Seoul, South Korea; 2) Seoul National University Hospital, Seoul, South Korea.

Thyroid cancer is one of the most prevalent cancers worldwide, and papillary thyroid cancer accounts for ~80% of all thyroid cancers. Here we conducted RNA sequencing for 46 papillary thyroid cancer tissues in Korean patients and explored any fusion mutations found using the TopHat-Fusion program. As a result, we discovered 3 fusion mutations using our stringent criteria and these include *CCDC6/RET* and *PAX8/PPARG* fusions, which have been previously reported for this cancer type. The other *EZR/ERBB4* fusion is a novel fusion mutation found in papillary thyroid cancer, and the sample having this mutation was diagnosed as the follicular variant subtype with abnormal pathological findings. This is an in-frame mutation and the expression pattern verified this fusion at the exon level. Previous reports have shown that *ERBB4* is a driver gene for carcinogenesis in multiple types of cancers. In this study, we suggest the *EZR/ERBB4* fusion as a novel driver mutation in papillary thyroid cancer, showing the atypical follicular patterns of pathology.

3277T

Sequence variants in *BARD1* and breast cancer susceptibility: results from the Breast Cancer Family Registry study. *F. Lesueur[1,2], F. Damiola[1], N. Robinot[1], N. Forey[1], G. Durand[1], C. Voegele[1], M.P. Vallée[1], J.L. Hopper[3], M.C. Southey[4], I.L. Andrulis[5], E.M. John[6], S.V. Tavtigian[7], Breast Cancer Family Registry.* 1) Genetic Cancer Susceptibility group, International Agency for Research on Cancer, Lyon, France; 2) Inserm, U900, Institut Curie, Mines ParisTech, Paris, France; 3) Center for Molecular, Environmental, Genetic and Analytical Epidemiology, School of Population Health, EGA The University of Melbourne, Victoria, Australia; 4) Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria, Australia; 5) Department of Molecular Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 6) Cancer Prevention Institute of California, Fremont, CA, USA and Stanford University School of Medicine and Stanford Cancer Institute, Stanford, CA, USA; 7) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, USA.

The RING-finger mediated BARD1/BRCA1 heterodimer is essential for the tumor suppressor functions of BRCA1, and both BRCA1 and BARD1 possess a pair of tandem BRCT domains that interact in a phosphorylation-dependent manner with target proteins. Since missense mutations in *BRCA1* that disturb the BARD1/BRCA1 interaction lead to breast cancer predisposition, it has been suggested that comparable mutations in *BARD1* may be responsible for a proportion of familial breast cancer. Indeed, likely pathogenic germline variants in this gene have occasionally been described in breast and/or ovarian cancer families. With the aim to investigate the frequency and attributed risk of *BARD1* variants in breast cancer susceptibility and to clarify relevance of mutation screening of this gene to familial cancer clinics and counselors, we sought to estimate the frequencies and nature of rare *BARD1* variants in a sample of women with early onset breast cancer (N=1,311) and frequency matched controls (N=1,111) from three population-based centers of the Breast Cancer Family Registry. High throughput screening of *BARD1* coding and flanking intronic regions in this series identified 30 missense variants, 13 synonymous variants and three variants in intronic regions close to splice junctions. To distinguish functionally neutral variants from those that may contribute to breast cancer risk, we used a combination of phylogenetic information with current knowledge on protein structure. Predictions on functional relevance of detected variants were obtained from different prediction algorithms. Among variants likely to alter the key RING and BRCT domains, we report for the first time a cysteine substitution (c.221G>T; p.Cys74Phe) affecting the BARD1 C3HC4 RING motif. This substitution was found in one case. The only protein-truncating variant identified was the synonymous substitution c.1977 A>G (p.Arg659Arg) leading to aberrant transcript that removes the second BRCT domain of BARD1 (p.Cys53_Trp635delinsfsX12); it was found in 7 cases (0.5%) and 9 controls (0.8%). Overall, the number of variants predicted to be damaging or neutral were evenly represented among cases and controls. Hence, our findings do not confirm in a population-based study setting of women with early onset breast cancer a risk associated with rare truncating or missense variants in BARD1. However we cannot exclude the possibility that mutations in *BARD1* will explain some small proportion of hereditary breast cancer.

3278F

Field synopses of genetic variation in colorectal neoplasia. *J. Little[1], H. Campbel[2], G. Gresham[1], Z. Montazeri[1], S. Sivakumaran[2], E. Theodoratou[2].* 1) Epidemiology & Community Med, Univ Ottawa, Ottawa, ON, Canada; 2) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK.

Objectives: We present results from the CRC field synopsis, report on work in progress on colorectal polyps, and comment on differences in the nature of the evidence, such as volume and quality of evidence, and issues including manner of detection of polyps, investigation of initially detected vs. recurrent polyps, and subtype. Methods: We have recently completed a field synopsis for colorectal cancer (CRC) and are in progress of developing one on colorectal polyps. For the CRC synopsis, we reviewed over 10,000 titles, then collated and extracted data from >600 publications reporting on >400 polymorphisms in >100 different genes. We carried out meta-analyses to derive summary effect estimates for >90 polymorphisms in >60 genes, including unpublished data from GWAS. For the polyp synopsis, we collated and extracted data from more than 170 publications reporting on about 200 different genes. Odd ratios are estimated based on meta-analysis for 11 polymorphisms so far. We considered four genetic models (two additives, one dominant, and one recessive). To assess the credibility of associations, we applied the Venice criteria and added consideration of Bayesian False Discovery Probability. Results: Based on meta-analysis for CRC 16 independent variants at 13 loci represent the most highly credible findings and 23 variants at 22 loci have 'less credible' association. We haven't gene associated with polyps at this stage. In addition, we stimulate discussion about (a) updating of field synopses and (b) operationalization of the Venice criteria, for both of which there appear to be differences across field synopses. These issues have relevance in public health beyond the area of genetic susceptibility. Conclusions: Our data should help results of genetic associations studies to be placed in context and interpreted appropriately and should help direct future research effort.

3279W

Genomic analysis of inherited breast cancer among Palestinian women. *S. Lolas-Hamameh[1,2], D. Dweik[1], F. Fostira[4], T. Walsh[3], M.K. Lee[3], S. Casadei[3], MC. King[3], E. Levy-Lahad[2], M. Kanaan[1].* 1) Hereditary Research Laboratory, Bethlehem University, Bethlehem, Palestine; 2) Medical Genetics institute, Shaare Zedek Medical Center, Jerusalem, Israel; 3) Departments of Medicine and Genome Sciences, University of Washington, Seattle, Washington, USA; 4) National Centre for Scientific Research"Demokritos", Athens, Greece.

In the Middle East, breast cancer incidence among Palestinian women has historically been low, but with increased education and later initiation of child bearing among Palestinian young women, is now rapidly increasing. Furthermore, perhaps because of its historically low incidence, breast cancer among Palestinian women is strikingly familial. We explored the genetic bases of this familial risk in the context of providing culturally appropriate genetic counseling services to high-risk women. Participants were 274 Palestinian and Arab-Israeli women with breast cancer, most either with a positive family history of breast or ovarian cancer (116 subjects) or diagnosed at age 40 or younger (130 subjects). Genomic DNA was tested by BROCA, which enables the capture and simultaneous multiplex sequencing of all coding, regulatory, and intronic regions of 30 known breast and ovarian cancer genes. Considering only unambiguously damaging mutations (i.e. truncations, complete deletions, splice mutations leading to a mutant message, and missenses proven experimentally to be damaging), 29 of the 274 subjects (11%) carried a mutation responsible for their breast cancer. These included 22 of the 116 familial subjects (19%), 8 of the 130 young-onset-nonfamilial subjects (6%), and 0 of the 29 subjects not meeting either criterion. The damaging mutations included 7 in BRCA1, 10 in BRCA2, 2 in ATM, 2 in BARD1, and 1 each in TP53, CHEK2, CDH1, PALB2, ATR, BRIP1, and XRCC2. With two exceptions (BRCA2 p.E2229X and BRCA2 c.6462delTC), all mutations were different. Also, two subjects had two mutations each: in CHEK2 and BARD1; and in CHEK2 and ATM. Multiple variants potentially altering splicing and missenses potentially damaging to function are still in process of evaluation and remain good candidates. Based on the historical demography of the region, we anticipated that the spectrum of mutations predisposing to breast cancer in the Palestinian population would be broad, with multiple individually rare, highly localized mutations, rather than a small number of founder alleles; that is, a European pattern rather than an Ashkenazi Jewish pattern. This proved to be true. Breast cancer among Palestinian women is generally diagnosed at late stages, and consequently has poor prognosis. If Palestinian women at genetically high risk were provided the opportunity to undertake special scrutiny, staging and mortality could be substantially improved.

3280T

Prostate cancer risk regions in 8q24 and 17q24 are differentially associated with somatic *TMPRSS2:ERG* fusion status. *C. Maier[1,2], A.E. Rinckleb[1,2], M. Luedeke[1,2], J.L. Stanford[3,4], J. Schleutker[5,6], R.A. Eeles[7,8], M. Teixeira[9,10], S. Weikert[11], J. Hoegel[2], L.M. FitzGerald[3], T. Wahlfors[5], T. Visakorpi[12], K.A. Leinonen[12], T.L.J. Tammela[13], C.S. Cooper[7,14], Z. Kote-Jarai[7], S. Edwards[7], P. Paulo[9,10], C. Jéronimo[9,10], H. Krause[11], W. Vogel[2], S. Benlloch[15], A. Amin Al Olama[15], D.F. Easton[15], the PRACTICAL consortium.* 1) Department of Urology, University of Ulm, Ulm, Germany; 2) Institute of Human Genetics, University of Ulm, Ulm, Germany; 3) Fred Hutchinson Cancer Research Center, Division of Public Health Science, Seattle, Washington, USA; 4) Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA; 5) Institute of Biomedical Technology/BioMediTech and Fimlab Laboratories, University of Tampere, Tampere, Finland; 6) Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland; 7) The Institute of Cancer Research, Sutton, UK; 8) Royal Marsden National Health Service Foundation Trust, London and Sutton, UK; 9) Department of Genetics, Portuguese Oncology Institute, Porto, Portugal; 10) Abel Salazar Biomedical Sciences Institute, Porto University, Porto, Portugal; 11) Department of Urology, University Hospital Charité, Berlin, Germany; 12) Institute of Biomedical Technology/BioMediTech, University of Tampere, Tampere, Finland; 13) Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland; 14) Department of Biological Science, University of East Anglia, Norwich, UK; 15) Centre for Cancer for Cancer Genetics Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.

Deep sequencing of prostate cancer (PrCa) genomes has recently pointed at an early role of *ETS* gene fusions in tumorigenesis, and revealed a characteristic landscape of structural rearrangements accompanying the oncogene translocation. Since PrCa in general has been proven partially heritable, we considered fusion positive PrCa as a distinct tumor entity that may be traced back to specific sets of germline risk factors. In a collaborative setting involving 1,221 cases with somatic status of the oncogene *TMPRSS2:ERG* we have investigated 27 common PrCa risk variants, known from previous GWAS studies, for their particular contributions to fusion positive or negative subclasses. Mantel-Haenszel meta-analysis was used to compare frequencies of risk alleles between the two subtypes. In a first set of 552 cases the PrCa risk loci 10q11 (rs10993994: OR = 1.35; p = 0.015), 19q13 (rs2735839: OR = 1.73; p = 0.0035), 17q24 (rs1859962: OR = 1.29; p = 0.038) and 8q24 (rs16901979: OR = 0.53; p = 0.021) appeared associated with fusion status at nominal significance. A replication round comprising a further 669 cases verified a selective involvement of 17q24 in fusion positive PrCa (stage 1 and 2 combined: p = 0.0016), and of 8q24 in fusion negative PrCa (stage 1 and 2 combined: p = 0.0006). Both loci, 17q24 and 8q24, are gene deserts, where the causal process for PrCa susceptibility is not fully understood to date. Noteworthy, one additional locus (rs1447295) in 8q24, which represents a risk region independent from rs16901979, also showed association with fusion status in our study (stage 1 and 2 combined: OR = 0.70; p = 0.0024). The concurrence of both 8q24 variants being associated with fusion negative PrCa is not due to linkage disequilibrium, and indicates a common pathogenic mechanism that is unrelated to *ETS* activity. In contrast, *ETS* pathway involvement at the 17q24 locus is plausible, since long range interactions have previously been described between risk genotypes and the nearby *SOX9* gene, encoding a transcription factor which was observed to be co-expressed with *ERG* in the androgen driven profile of fusion positive tumors. In brief, our study further supports the distinctness of *ETS* positive vs. negative PrCa, and demonstrates in principle, that molecular subtypes could represent different entities also on the germline level. Larger sets of PrCa cases with somatic typing will be required in order to identify patterns of corresponding predisposing factors.

3281F

Identification of Susceptibility Loci in Hereditary Prostate Cancer Families Using Copy Number Variation and Linkage Analysis. *D. Mandal[1], E. Ledet[2], J.E. Bailey-Wilson[3], M. Li[4].* 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Tulane Cancer Center, Tulane University School of Medicine, New Orleans, LA; 3) National Human Genome Research Institute/National Institutes of Health, Baltimore, MD; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

In the United States, it is estimated that about 238,590 new cases of prostate cancer will be diagnosed and about 29,720 men will die from this disease in 2013. Family history is the most significant predictor of prostate cancer development and approximately 10% of prostate cancer cases are attributable to inheritable genetic factors. However, disease gene identification for prostate cancer has been extremely challenging due to both disease and genetic heterogeneity. To overcome the effects of ethnic disparity, genetic heterogeneity, incomplete penetrance, and missing heritability, our goal was to identify genetic components of prostate cancer by using a comprehensive approach combining both array comparative genomic hybridization (aCGH) analysis and linkage analysis. Eight prostate cancer cases were studied using aCGH to search for germ-line copy number variants (CNVs) associated with hereditary prostate cancer. The study subjects were from 8 large, high-risk, clinically homogenous families with European ancestry from Southern Louisiana. Three novel regions of CNVs were identified: 16q23, 11q22, and 2q22 in all 8 prostate cancer cases. Both model-based and model-free linkage analyses were performed on 4 of the most informative families (≥5 prostate cancer cases/family) from these 8 prostate cancer cases. Genotyping for linkage analyses was done using Illumina Infinium II SNP HumanLinkage-12 panel. Suggestive evidence of linkage was obtained at 2q14 (HLOD score of 1.94). Using aCGH, a 68 kb duplication was observed in this region in all 8 hereditary prostate cancer cases. These genetic findings, which were identified in multiple large, well-characterized families, provide new insight into hereditary prostate cancer. Future fine mapping of the 2q14 region in a larger cohort of patients, confirmation of novel CNVs in additional subjects and use of next-generation sequencing approaches are needed to provide further evidence of susceptibility loci for prostate cancer.

3282W

A systems biology approach reveals novel mechanisms associated with cancer metastasis in EMT-MET cell reprogramming. *R.C. McEachin[1], H. Roca[2].* 1) Computational Medicine and Bioinformatics, Univ Michigan, Ann Arbor, MI; 2) Urology, Univ Michigan, Ann Arbor, MI.

Cancer progression is characterized, in part, by altered or aberrant gene expression. Mesenchymal to Epithelial Transition (MET) is a reversible process (EMT) that is critical to metastasis in cancer progression. In this work, we established 'MET signatures' to characterize changes of gene expression in models of prostate cancer and breast cancer, where we induced MET by over-expression of OVOL1, OVOL2, or both OVOL transcription factors (TFs). We assessed differential expression by RNA-Seq (FDR ≤ 0.05 AND Fold Change ≥ ± 2.0) and established a set of genes representing the MET signature for each model (1,622 in BC and 2,692 in PC). The 740 genes at the intersection of these two sets represent the common MET signature for these two cancer models. We performed ConceptGen enrichment testing for this set and found significant enrichment for annotation consistent with cancer progression (e.g., biological adhesion, blood vessel development). Interestingly, we also found that annotation for regulation of expression by AP-1 TF is significantly enriched in this set (FDR 7.9E-5). Since we initially induced MET by over-expression of OVOL TFs, we hypothesized that AP-1 and OVOL TFs cooperate in regulation of the MET signature. Consistent with this model, using Genomatix Genome Analyzer we found that promoters for the set of MET signature genes are significantly enriched for AP-1 binding sites, OVOL sites, and the combined model including both sites. To help us understand the broader implications of the MET signature, we did two follow-on analyses. We first identified 129 genes showing at least +0.50 correlation with OVOL1 expression in the 178 cancer types from Barretina et al. in Oncomine. Of these 129 genes almost 60% (74 genes) are common to the MET signature, consistent with the MET signature's impact in a broadened range of cancer phenotypes. We then used GeneGo to model the network of interactions among the genes annotated as being AP-1 regulated in the ConceptGen analysis. We prioritized drug/target combinations in this network for follow-on testing, emphasizing the potential clinical/translational relevance of this work. In summary, we find significant evidence consistent with the OVOL TFs having important roles in MET in two modeled cancers. We also find evidence of cooperation between the OVOL TFs and AP-1 in MET, a broadened role of MET signature genes in a range of cancers, and immediate targets for follow-up in the clinic.

3283T

Copy number variation of the antimicrobial-gene, defensin beta 4, is associated with susceptibility to cervical cancer. *K. Miura[1], S. Abe[1], A. Kinoshita[2], H. Mishima[2], S. Miura[1], Kl. Yoshiura[2], H. Masuzaki[1].* 1) Dept OB/GYN, Nagasaki Univ Sch Med, Nagasaki, Japan; 2) Dept Human Genetics, Nagasaki Univ Sch Med, Nagasaki, Japan.

Background: Cervical carcinogenesis has two critical transition steps: persistent oncogenic human papillomavirus (HPV) infection and progression to cervical cancer. Oncogenic HPV infection alone is not sufficient to cause cervical cancer. Host genetic factors may also contribute to cervical cancer pathogenesis, but their roles remain to be determined. The defensin beta 4 gene, DEFB4, which has antimicrobial properties, is a candidate host genetic factor for susceptibility to cervical cancer. **Objectives:** The aim of this study was to investigate association between copy number variation of DEFB4 and susceptibility to cervical cancer in a population at high-risk of persistent oncogenic HPV infection. Methods: The study subjects comprised 204 women with cervical cancer, a population having a high-risk of persistent oncogenic HPV infection (cervical cancer group), and 200 healthy women from the general population (control group). Copy number variation of DEFB4 in each test sample was determined by relative quantitation using the comparative CT (ΔΔCT) method. Differences between the two groups were evaluated. **Results:** The median DEFB4 copy number in the cervical cancer group was four and in the control group was five (p=2.77e-4, t-test). The odds ratio of cervical cancer in individuals with 4 DEFB4 copies or less was higher (odds ratio 2.02; 95% confidence interval odds ratio 1.36-3.02), compared with that in individuals with 5 or more copies (odds ratio 0.49; 95% confidence interval odds ratio 0.33-0.74). Therefore, both groups indicated a two-tailed significant difference from an odds ratio of 1.00 at the 5% level; level. **Conclusions:** We found copy number variation of DEFB4 was a host genetic factor conferring susceptibility to cervical cancer. A lower DEFB4 copy number was associated with susceptibility to cervical cancer.

3284F

Effects of Waterpipe Smoking on gene expression. *Z. Montazeri, H. El katerji, J. Gomes, J. Little.* Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada.

Objectives: Recently, a sharp rise in waterpipe smoking has been observed in North America and Europe, especially among young adults. There is a belief that waterpipe is the healthiest way to smoke tobacco. We studied the effects of waterpipe smoking on gene expression among young waterpipe smokers. Methods: Adverse health effects of waterpipe smoking have been reported in different studies, but this evidence has been appraised as of poor quality. There is a gap in the evidence as to the potential health effects of waterpipe smoking. We are investigating the effects of waterpipe smoking on gene expression among young waterpipe smokers in Ottawa, Canada. We asked participants to respond to a questionnaire and to provide a saliva sample. DNA extracted from these samples is being analyzed using PCR arrays. Results: We selected 18 genes to study the potential carcinogenic effects of waterpipe smoking on cancer based on gene expression. They have been selected based on the fact that they are induced by cigarette smoking as well as they are in the pathways of cancer diseases; genes of xenobiotic metabolism are also included. The main inclusion criteria was that the individual was between the age of 18 and 25 and reported that they smoked waterpipe. The fold change between before and after one hour and a half of smoking waterpipe, effect size, ranged between 0.02 and 34.42. Conclusions: Results could be used to predict the health effects of waterpipe smoking. This research will be used in knowledge translation to enable public health professionals and policy makers to make informed decisions about the control of waterpipe smoking, including potential prevention strategies and cessation interventions.

3285W

Upregulation of TRF1 and TRF2 (telomere repeat binding factors) protein contributes to telomere shortening in renal cell carcinoma. *D. PAL[1], U. Sharma[1], R. Khajuria[1], S.K Singh[2], N. Kakkar[3], R. Prasad[1].* 1) Biochemistry, PGIMER, Chandigarh, Chandigarh, India; 2) Urology,PGIMER,Chandigarh, India; 3) Histopathology, PGIMER,Chandigarh,India.

Upregulation of TRF1 and TRF2 (telomere repeat binding factors) protein contributes to telomere shortening in renal cell carcinoma D Pal1* , U Sharma1, R Khajuria1, SK Singh2 , N Kakkar3 and R Prasad1. 1. Department of Biochemistry, PGIMER, Chandigarh 2. Department of Urology, PGIMER, Chandigarh 3. Department of Histopathology, PGIMER, Chandigarh *e.mail-: deekshapal64@gmail.com Background : TRF1 and TRF2 are telomere repeat binding proteins that are exclusively found at the telomeres. Both TRF proteins have a Myb-like helix-turn-helix domain in their carboxyterminus and a central conserved domain that includes sequences responsible for the formation of homodimers. The two proteins do not heterodimerize, and they differ substantially at the N terminus which is acidic in TRF1 but basic in TRF2. The protection of human telomeres crucially depends on these factors and it is reasonable to assume that the requirement for TTAGGG repeats at chromosome ends reflects the need for TRF1 and TRF2 binding. Telomere dysfunction is believed to be the significant factor in carcinogenesis. To elucidate the carcinogenesis mechanism, the expression of TRF1, TRF2 and change in telomere length were investigated in renal cell carcinoma (RCC). Materials and Methods: Total 80 cases of RCC treated by surgery under advanced Urology services of Nehru Hospital, at Postgraduate Institute of Medical Education and Research, Chandigarh were included in the present study. For comparison, normal renal cortex samples were taken in each case. Transcriptional expression of TRF1 and TRF2 were estimated by real time PCR. Whereas Protein levels were detected by using immunohistochemical and immunofluorscence method. The mean telomere length was determined by southern blotting followed by hybridization. Results: The expression of TRF1 and TRF2 were significantly higher in the tumor tissue in comparison with normal renal parenchyma. The mean telomere length in RCC tissue was significantly shorter than that in normal renal parenchyma. The mean telomere length in all tissue samples were inversely correlated with the level of TRF1 and TRF2 expression. Conclusion: Our result suggests that the upregulation of TRF 1 and TRF2 may work to reduce the telomere length in RCC and could contribute to the carcinogenesis of renal cell carcinoma.

3286T

Screening of *XRCC2* in breast cancer families. *L.M. Pelttari[1], S. Vilske[1], C. Blomqvist[2], K. Aittomäki[3], H. Nevanlinna[1].* 1) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 2) Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland.

XRCC2, one of the five RAD51 paralogs, is involved in the repair of DNA double-strand breaks through homologous recombination. Most of the breast cancer susceptibility genes, including the high risk genes *BRCA1* and *BRCA2*, have an important role in DNA damage repair. Deleterious mutations in the *XRCC2* gene were recently identified in an exome-sequencing study of familial breast cancer patients and a homozygous truncating mutation in the gene was identified in a Fanconi anemia patient. However, the association of *XRCC2* with breast cancer was not confirmed in a larger case-control study. To investigate the role of *XRCC2* mutations in breast cancer predisposition in the Finnish population, we sequenced the coding region, the 5'UTR, and the exon-intron boundaries of the gene in blood DNA samples of 344 familial breast cancer patients. This extensive set of non-BRCA1/2 breast cancer families included 342 families with at least three breast or ovarian cancer patients among first or second degree relatives and 2 families with two affected first degree relatives. We detected four known polymorphisms among the breast cancer families: one downstream, one 5'UTR, one intronic and one missense variant. No novel changes or truncating mutations were identified. None of the polymorphisms were predicted to be pathogenic and the minor allele frequencies were comparable to the population frequencies. Our results indicate that *XRCC2* does not contribute to familial breast cancer predisposition in the Finnish population. Taken together, it is unlikely that *XRCC2* has a significant contribution to breast cancer susceptibility.

3287F

Large-scale resequencing analysis of six melanoma susceptibility genes in the European Prospective Investigation into cancer and nutrition Cohort. *M. Pertesi[1], N. Forey[1], J. Oliver[1], N. Robinot[1], C. Voegele[1], EPIC. skin cancer working group[3], F. Le Calvez-Kelm[1], J. McKay[1], F. Lesueur[1,2].* 1) Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France; 2) Inserm U900, Institut Curie, Mines ParisTech, Paris, France; 3) International Agency for Research on Cancer, Lyon, France.

Cutaneous malignant melanoma (CMM) is an important health problem in fair-skinned populations worldwide, showing a dramatic increase in incidence over the past decades. Established risk factors include sun exposure, nevus propensity, pigmentary traits, and familial history. Inherited DNA sequence variants are involved in the development and prognosis of CMM and three classes of susceptibility genes have been described. Highly-penetrant mutations in the cell cycle related genes *CDKN2A* and *CDK4* account for about 2% of CCM cases across populations, while more moderately penetrant alleles (*MC1R*, *MITF*), and very common, low-penetrance SNPs in genes related to pigmentation, nevus count, immune response, DNA repair, and metabolism explain about 12% of the familial risk in European populations. Given that the genome-wide association studies conducted by International consortia were well powered to detect SNPs with OR>1.5, it seems unlikely that many additional common alleles with moderate effects on melanoma risk exist, indicating that other uncommon-to-rare susceptibility alleles remain to be identified. To test this hypothesis we designed a cost-effective multiplex-targeted resequencing assay specifically testing for 6 melanoma susceptibility genes identified in the pangenomic studies. About 17kB of DNA sequence corresponding to the exons and flanking intronic regions of *MC1R*, *MITF*, *ASIP*, *TYR*, *TYRP1* and *BAP1* genes is being screened in 1061 CMM cases, and 1061 matched controls from the European Prospective Investigation into cancer and nutrition Cohort. Our custom mutation screening workflow involves the creation of a barcoded sample library for each individual including amplified regions of interest and sequencing on the PGM Ion Torrent sequencer. Each step of the process is tracked by a Laboratory Information Management System. To assess the sensitivity and specificity of our custom workflow, we compared PGM data with High-Resolution Melt (HRM) and Sanger sequencing data in a subset of 429 subjects screened for variants in the highly polymorphic *MC1R* gene. We identified 462 and 461 sequence variations, with the two workflows, respectively. These correspond to 27 distinct common and rare variants, including 1 inframe deletion and 1 duplication. Single nucleotide variants were missed in 7 subjects screened by HRM/Sanger and in 8 subjects screened by PGM. Discrepancies concerning homozygous/heterozygous status were found in 12 subjects carrying a common SNP.

3288W

K939Q polymorphism in the XPC gene may not affect the risk of bladder cancer: a case-control study and a meta-analysis. *M. SANKHWAR[1], S.N. SANKHWAR[1], N. GUPTA[2], A. ABHISHEK[1], S. RAJENDER[2].* 1) Urology, King George Medical University, Lucknow, Uttar Pradesh, India; 2) Endocrinology Division, CSIR-CDRI, Lucknow, Uttar Pradesh, India.

Purpose: Lys939Gln (K939Q), an A>C transversion in exon 15 of the XPC gene, has been extensively studied in relation to cancer. Keeping in view its decisive polymorphic effects and collocations with bladder cancer, we have analysed this polymorphism in Indian bladder cancer patients, and conducted a meta-analysis to generate a pooled estimate about the effect of this polymorphism on bladder cancer. Material and methods: The polymorphism was analysed using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) in 234 case and 258 control samples, followed by chi square test for statistical analysis. Meta-analysis included a total of fourteen studies comprising of 5403 cases and 5688 controls. Forest plot and funnel plot were used to interpret statistical significance of association and publication bias, respectively. Egger's test was performed to confirm funnel plot symmetry statistically. Results: We observed no significant association between K939Q polymorphism and bladder cancer. Tobacco chewing was found to be a risk factor for bladder cancer (OR= 1.93; 95%CI=1.32-2.83, P=0.0005). Meta-analysis showed association of CC genotype with marginal significance in homozygous and recessive models (OR=1.28, 95%CI=1.03-1.59, P=0.021, and OR=1.27, 95% CI=1.05-1.54, P=0.014), which turned non-significant upon applying Bonferroni's correction. Meta-analysis on the basis of ethnicity showed no significant association in any of the ethnic groups. Conclusion: K939Q substitution does not affect bladder cancer risk in Indo-European populations of Uttar Pradesh and perhaps also in other populations across the world. .

3289T

COMPLEXO: Identifying the missing heritability of breast cancer via next generation collaboration. *M.C. Southey[1], D.E. Goldgar[2], COMPLEXO.* 1) Dept Pathology, Univ Melbourne, Melbourne, VICTORIA, Australia; 2) Huntsman Cancer Institute, The University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

A proportion of the remaining unexplained genetic susceptibility to breast cancer is likely to be explained by many rare mutations in a large number of genes. Coordinated international collaboration offers great potential to advance the discovery of additional breast cancer susceptibility genes by increasing the likelihood of identifying functionally relevant genetic variants in the same genes in multiple families. A new consortium, COMPLEXO (a name chosen to reflect the complexity of the exome), has been formed to facilitate collaborations between researchers actively applying massively parallel sequencing to understand the genetics of breast and ovarian cancer. The aim of COMPLEXO is to bring to massively parallel sequencing the same power of large sample sets that have proven so successful in examining the role of common variants in cancer populations via the consortium model such as the Breast Cancer Association Consortium (BCAC), the Ovarian Cancer Association Consortium (OCAC) and the Collaborative Oncology Gene-environment Study (COGS). However, sequencing studies provide additional challenges in terms of defining specific modes of collaboration given differences in sequencing and targeted capture platforms, bioinformatics platforms, the need to integrate ongoing studies in many centers and socio-ethical-legal issues that are not as relevant to initiatives that are genotyping common genetic variation. These issues are relevant to research in all complex human diseases. We will describe our working and governance structures, our early experiences and ongoing activities aimed at identifying more of the missing heritability of breast cancer.

3290F

Common genetic polymorphisms in the coagulation factors 5 and 10 genes are associated with risk of breast cancer and correlate with increased coagulation activity. *M. Tinholt[1,2,3], MK. Viken[1], S. Nyberg[4,5], AL. Børresen-Dale[4,5], V. Kristensen[4,5,6], KK. Sahlberg[4,5], R. Kåresen[3,7], E. Schlichting[7], O. Garred[8], G. Skretting[2], AE. Dahm[2], BA. Lie[1], PM. Sandset[2,3], N. Iversen[1].* 1) Dept. of Medical Genetics, Oslo University hospital and University of Oslo; 2) Dept. of Haematology and Research Institute of Internal Medicine, Oslo University Hospital; 3) Institute of Clinical Medicine, University of Oslo; 4) Dept. of Genetics, Institute for Cancer Research, Oslo University Hospital; 5) The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo; 6) Dept. of Clinical Molecular Biology (EpiGen), Akerhus University Hospital, University of Oslo; 7) Dept. of Breast and Endocrine Surgery, Oslo University Hospital; 8) Dept. of Pathology, Oslo University Hospital.

Tumor cells express coagulation factors, thereby creating a pro-thrombotic state in cancer. Besides increasing the risk of thrombosis, pro-coagulants may promote pro-cancer signaling independent of coagulation. We aimed to gain a better understanding of the molecular mechanisms controlling the interplay between cancer and hemostasis through a Norwegian case-control study of 390 non-treated breast cancer patients and 350 healthy controls. SNP associations in the hemostatic genes *F2, F3, F5, F7, F10* and *TFPI* (n=42 from HapMap) were explored, in addition to the F5 Leiden (rs6025) and the prothrombin (rs1799963) variants. Genotype-phenotype correlations were conducted between the SNPs that were associated with breast cancer and mRNA expression in tumor and plasma levels of twelve hemostatic parameters. Increased levels of the coagulation factors F8, F9 and VWF, and decreased levels of antithrombin were detected in the patients. Global activation of coagulation was demonstrated in an endogenous thrombin generation assay, by increased D-dimer levels, and increased activated protein C resistance (APCR). The genetic risk factors of thrombosis (F5 Leiden and the prothrombin variants) showed no association with breast cancer. Four common SNPs in *F5* and one in *F10* were significantly associated with risk of breast cancer. The minor alleles of *F5* rs12120605 and rs6427202 were more frequent in patients (OR=1.52, [95%CI,1.07-2.17] and OR=1.40, [95%CI,1.02-1.91]). They were associated with increased tissue factor and factor 5 mRNA expression in tumors, respectively. In contrast, the minor alleles of *F5* rs6427199 and rs9332542 were more frequent in controls (OR=0.71, [95%CI,0.52-0.96] and OR=0.65, [95%CI,0.49-0.87]), and they correlated with antithrombin in plasma and factor 5 mRNA levels in tumor, respectively. This implies that these SNPs may be located in trans- and cis-acting DNA elements that contribute to regulate the pro-coagulant expression in plasma and tumors. The *F5* rs6427199 also correlated with lowered APCR, partly explaining the case-control variation in APCR not explained by the F5 Leiden variant. *F10* rs3093261 also associated with breast cancer (OR=1.60, [95%CI, 1.17-2.17]), but with no clear effect on hemostatic phenotypes. In conclusion, this study demonstrates novel associations of common SNPs in hemostatic genes (*F5* and *F10*) influencing the risk of breast cancer, and that the pro-coagulant phenotype verified in patients may be due to genetic effects.

3291W

Estrogen and Progesterone receptor gene polymorphisms as risk factors for Cervix Cancer in South Indian Population. *P. Upendram[1], V. Kiran Kumar[1], S. Poornima[1], A. Shah[2], Q. Hasan[1].* 1) Department of Genetics & Molecular Biology, Kamineni Hospitals, Hyderabad, India; 2) Department of Oncology, Kamineni Hospitals, Hyderabad, India.

Carcinoma of the uterine cervix is the most frequent gynaecological malignancy affecting women in developing countries despite being a potentially preventable disease. Globally, it is the second most frequent cancer affecting women with a steady rise in incidence among younger women. It is evident from current literature that specific polymorphisms in hormone receptor genes may enhance the risk for certain gynaecological malignancies. Altered responses to estrogen and progesterone lead to steroid hormone dependent disorders like endometriosis, endometrial cancer, breast cancer and ovarian cancer. Despite evidence that estrogen may play an important role in the carcinogenesis of cervical cancer, its action and mechanism in cervical cancer invasion is not well defined. To the best of our knowledge there are no studies that have focused on the estrogen and progesterone receptor (ER and PR) gene variants in cervical cancer especially in the Asian Indian population. The aim of this case control study from South India was to establish the association between the T/C SNP of estrogen receptor (ER) alpha gene recognized by PvuII enzyme (rs2234693) and a 306bp Alu insertion in the PGR gene (rs1042838) with cervical cancer. DNA was isolated from a total of 200 women comprising 100 cases with cervical cancer and 100 age-matched healthy controls using the salting out technique. Polymorphisms were evaluated by Polymerase Chain Reaction followed by restriction enzyme digestion / gel electrophoresis. Our results showed a statistically significant association of the C allele of ER gene (OR= 5.8072; 95% CI=3.7671 to 8.9521; p<0.0001) with cervix cancer. The homozygous CC genotype (OR= 26.000; 95% CI=8.8791 to 76.1338; p<0.0001) showed a 26 fold increased risk for this cancer. However, the T2 allele of PR gene did not show any significant association with cervix cancer (OR=1.3700; 95% CI=0.7193 to 2.6093; p= 0.33). Estrogen and its receptors are implicated in the promotion and prevention of various cancers. The uterine cervix is highly responsive to estrogen, however, the role of estrogen receptor polymorphisms in cervical cancer has not been well established. Our results suggest that the ER PvuII polymorphism but not PROGINS polymorphism can be used as a bio-marker to identify women with a high risk of developing cervix cancer. A larger study in different ethnic groups is warranted for establishing the association of cervix cancer with the ER polymorphism.

3292T

Fine-mapping of Genome-wide Association Study-identified Risk Loci for Colorectal Cancer in African Americans. *H. Wang[1], C.A. Haiman[2], T. Burnett[1], B.K. Fortini[2], L.K. Kolonel[1], B.E. Henderson[2], L.B. Signorello[3,4], W.J. Blot[3,4], T.O. Keku[5], S.I. Berndt[6], P.A. Newcomb[7], M. Pande[8], C.I. Amos[9], D.W. West[10], G. Casey[2], R.S. Sandler[5], R. Haile[2], D.O. Stram[2], L. Le Marchand[1].* 1) University of Hawaii Cancer Center, Honolulu, HI; 2) Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 3) International Epidemiology Institute, Rockville, MD; 4) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN; 5) Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, NC; 6) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 7) Fred Hutchinson Cancer Research Center, Seattle, WA; 8) Department of Epidemiology, the University of Texas M. D. Anderson Cancer Center, Houston, TX; 9) Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH; 10) Cancer Prevention Institute of California, Fremont, CA.

Genome-wide association studies of colorectal cancer (CRC) in Europeans and Asians have identified 21 risk susceptibility regions (29 index SNPs). Characterizing these risk regions in diverse racial groups with different LD structure can help localize causal variants. We examined associations between CRC and all 29 index SNPs in 6,597 African Americans (1,894 cases and 4,703 controls). Nine SNPs in 8 regions (5q31.1, 6q26-q27, 8q23.3, 8q24.21, 11q13.4, 15q13.3, 18q21.1, 20p12.3) formally replicated in our data with one-sided p-values < 0.05 and the same risk directions as reported previously. We performed fine-mapping of the 21 risk regions (including 250kb on both sides of the index SNPs) using genotyped and imputed markers at the density of the 1000 Genomes Project to search for additional or more predictive risk markers. Among the SNPs correlated with the index variants, two markers, rs12759486 (or rs7547751, a putative functional variant in perfect LD with it) in 1q41 and rs7252505 in 19q13.1, were more strongly and statistically significantly associated with CRC, after conditioning on the index signals (P<0.0006). The average per allele risk was improved using the replicated index variants and the two new markers (OR = 1.14, P = 6.5×10^{-16}) in African Americans, compared to using all index SNPs (OR = 1.07, P = 3.4×10^{-10}). The contribution of the two new risk SNPs to CRC heritability was estimated to be 1.5%. This study highlights the importance of fine-mapping in diverse populations.

3293F

Zinc Transporter Gene Variants and Susceptibility to Urological Cancers. *L. Wu[1], K. Rabe[1], A. Parker[2], G. Petersen[1].* 1) Mayo Clinic, Rochester, MN; 2) Mayo Clinic, Jacksonville, FL.

Introduction: Laboratory studies have demonstrated that zinc and zinc transporters play important roles in the development of urological cancers, including prostate, bladder, and renal cell cancer. To date however, there have been no efforts to evaluate whether germline variation in zinc transporter genes alter the risk of developing these cancers. Motivated by this, we evaluated associations of known variants in zinc transporter genes with risk of prostate, bladder, and renal cell cancer using three NCI Genome Wide Association Study (GWAS) data sets downloaded from dbGaP. Methods: Data consisted of GWAS data from the PLCO Screening Trial (prostate; 1172 cases, 1157 controls); Bladder Cancer study (3527 cases, 5119 controls); and Renal Cancer study (1453 cases, 3531 controls). All genotyped variants in 10 ZNT family genes (ZNT1-ZNT10) and 14 ZIP family genes (ZIP1-ZIP14) in each data set were analyzed by logistic regression models to evaluate the association of each variant with risk of developing prostate, bladder, and renal cancer, respectively. We evaluated these associations in a univariable setting, after adjusting for available covariates, such as age and gender. Correction for multiple comparisons used Max T permutation tests. Results: One variant in ZIP11, rs8081059, was associated with increased risk of renal cancer (OR=1.28, 95% CI= 1.13-1.45; corrected p= 0.049). Another ZIP11 variant, rs11871756, was associated with an increased risk of bladder cancer risk (OR=1.43, 95% CI 1.24-1.63; corrected p=0.0002). We also found three variants in ZIP11 associated with a decreased risk of bladder cancer: rs11077654 (OR=0.76, 95% CI 0.68-0.85; corrected p=0.001), rs9913017 (OR=0.76, 95% CI= 0.68-0.85; corrected p= 0.002), and rs4969054 (OR=0.78, 95% CI 0.69-0.88; corrected p=0.02). Interestingly, these three protective variants are within the same LD block. There were no associations between zinc transporter variants and prostate cancer risk. Conclusion: We report for the first time that specific genetic variants of a zinc transporter gene, ZIP11, are associated with risk of renal and bladder cancer. If validated in larger, more focused studies, further investigations are warranted to explore the functional significance of ZIP11 in these two cancers.

3294W

The Prevalence of the *HOXB13* G84E Prostate Cancer Risk Allele in Men Treated with Radical Prostatectomy. *K.A. Zuhlke[1], J. Beebe-Dimmer[2,3], C. Yee[2,3], A.M. Johnson[1], P.C. Walsh[4,5], S.D. Isaacs[4,5], W.B. Isaacs[4,5], K.A. Cooney[1,6].* 1) University of Michigan Medical School, Department of Internal Medicine, Ann Arbor, MI; 2) Wayne State University, Department of Oncology, Detroit, MI; 3) Karmanos Cancer Institute, Detroit, MI; 4) Johns Hopkins University, Baltimore, MD; 5) James Brady Urologic Institute, Baltimore, MD; 6) University of Michigan Medical School, Department of Urology, Ann Arbor, MI.

Purpose: A rare, non-conservative substitution (G84E) in the homeobox transcription factor *HOXB13* gene, located on chromosome 17q21, was recently reported to confer an increased risk of prostate cancer. As a follow-up to this investigation, DNA samples from 9,559 men undergoing radical prostatectomy were genotyped to determine the frequency and clinical correlates of this mutation in men with prostate cancer.

Materials and Methods: DNA samples from men treated with radical prostatectomy at the University of Michigan and John Hopkins University were genotyped for G84E and confirmed by Sanger sequencing. The frequency and distribution of this allele was determined according to specific patient characteristics (family history, age at diagnosis, pathologic Gleason grade and stage).

Results: 128 of 9,559 patients were heterozygous carriers of G84E (1.3%). Patients who possessed the variant were more likely to have a family history of prostate cancer (46.0% vs. 35.4% p=0.006). G84E carriers were also more likely diagnosed at a younger age compared to non-carriers (55.2 years vs. 58.1 years; p<0.0001). However, there was no difference in the proportion of patients diagnosed with high-grade or advanced stage tumors by carrier status.

Conclusions: In our study, carriers of the rare G84E variant in *HOXB13* were both younger at the time of diagnosis and more likely to have a family history of prostate cancer compared to homozygotes for the wild-type allele. No significant differences in allele frequency were detected according to select clinical characteristics of prostate cancer. Further investigation is required to evaluate the role of *HOXB13* in prostate carcinogenesis.

3295T

Prevalence of germline mutation p.R337H in the *TP53* gene in families with multiple cases of cancer. *K.C. Andrade[1,2], K.M. Santiago[1], A. Nobrega[1], F.P. Fortes[1], M.I. Achatz[1].* 1) A.C. Camargo Cancer Center, São Paulo, Brazil; 2) UNESP, São Paulo State University - Biosciences Institute, Botucatu - São Paulo - Brazil.

BACKGROUND: Germline mutations in the *TP53* gene are associated to the Li-Fraumeni Syndrome (LFS) and its variants Li-Fraumeni-Like (LFL). The diseases predispose carriers for early onset tumors, including soft tissue sarcoma, pre-menopausal breast cancer, central nervous system tumors and adrenocortical carcinoma. In Brazil, there is a high frequency of a germline *TP53* mutation in South and Southeastern population, due to a founder effect. It is estimated to be present in 0.3% of the local inhabitants. However, only a few families have been diagnosed with such alteration. Our hypothesis was that the current criteria applied for LFS are not enough to detect them. Therefore, the development of suitable criteria to select them will enable not only the identification of at-risk families, but also provide adequate screening and early detection of associated tumors. OBJECTIVE: To determine the efficacy of wide criteria for detection of p.R337H carriers. METHODS: 31 patients from Oncogenetics Department, A.C. Camargo Cancer Center, SP, were selected and tested for the p.R337H mutation. Criteria for inclusion were: (1) more than three family members with cancer AND (2) at least one of them under age 50 AND (3) two of them being first or second degree relatives. Germline DNA was extracted from peripheral blood and analyzed by Restriction Fragment Length Polymorphism. The confirmation of positive finding was done through direct sequencing of exon 10. RESULTS: One out of 31 patients (3,22%) was found to carry the p.R337H mutation. The patient developed ductal invasive breast cancer at age 47 and invasive adenocarcinoma of the lung at age 48. In addition, an extensive cancer family history was reported. DISCUSSION: These results show that the proposed criteria may detect carriers, but further studies including a larger group of families will be useful to define its effectiveness. Also, in order to improve sensibility, the inclusion of more stringent criteria will be needed. Moreover, one of the tumors manifested in family history is of great interest: the development of an Ewing's Sarcoma (ES) in her second degree relative at age 9. Current data indicate that the annual incidence of ES has had an average of 2.93 cases per million. Overall, 27% of them occur in the first decade of life and lesions in the *TP53* gene are only observed in 10% of this kind of cancer. Therefore, it is one of the few reports of a very early onset childhood ES in a LFS family.

3296F

Is *IGSF1* involved in human pituitary tumor formation? *F.R. Faucz[1], A.D. Horvath[2], P. Xekouki[1], E. Szarek[1], G. Evgenia[1], A.D. Manning[1], I. Levy[1], E. Saloustros[1], R.B. de Alexandre[1], M. Nesterova[1], C.A. Stratakis[1].* 1) Section on Endocrinology & Genetics, Program on Developmental Endocrinology & Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD; 2) McCormic Genomic and Proteomic Center, Department of Biochemistry and Molecular Biology, The George Washington University, Washington, DC.

IGSF1 is a membrane glycoprotein highly expressed in the anterior pituitary gland. Mice with a deficiency in *Igsf1* show central hypothyroidism and increased body size. Recently, mutations in this gene were found in humans with central hypothyroidism and testicular enlargement [Sun et al, Nat. Genet. 44(12):1375-81, 2012]. We looked for *IGSF1* germline variations in patients with giantism from the NIH data registry and in healthy controls. We also looked for the expression of IGSF1 in growth hormone (GH)-producing adenomas. We sequenced 11 patients with giantism and 92 health individuals (100% White Americans) for *IGSF1* germline mutations. Immunohistochemistry for IGSF1 was performed in sections from three GH-producing adenomas and in normal pituitary. In 1 out of 11 patients we identified the sequence variant p.Asn604Tre (c.1811A>C), which by *in silico* analysis is potentially destructive. The same variation, however, was found in 2 of our controls. We also identified 29 more variations in a total of 18 healthy individuals (16.85%) (3 nonsense, 3 frameshifts and 25 missense), but only 4 of them have been described previously. Immunohistochemistry showed variable IGSF1 staining in the GH-producing tumor from the patient with the *IGSF1* mutation compared to a GH-producing tumor from a patient negative for *IGSF1* mutations and to a normal control. *IGSF1* gene appears quite polymorphic. The mutation identified in one of our patients with giantism was also detected in healthy controls. The increased expression of IGSF1 in patients with an *IGSF1* germline mutation may indicate that *IGSF1* does not have a causative role in pituitary tumor development but may work as a modifier during oncogenesis.

3297W

Mutational Landscape of Candidate Genes in Familial Prostate Cancer. *A. Johnson[1], K. Zuhlke[1], C. Plotts[2], J. Douglas[2], S. Thibodeau[3], K. Cooney[1,4].* 1) Internal Medicine - Hem/Onc, Univ Michigan, Ann Arbor, MI; 2) Human Genetics, Univ Michigan, Ann Arbor, MI; 3) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Urology, Univ Michigan, Ann Arbor, MI.

Family history is a major risk factor for prostate cancer (PCa), suggesting a genetic component to the disease. However, traditional linkage and association studies have failed to fully elucidate the underlying genetic basis of familial PCa. Here we use a candidate gene approach to identify potential PCa susceptibility loci in whole-exome sequencing data from familial PCa cases. 500 candidate genes were identified based on function and/or previous association with prostate or other cancers. SNPs in these candidate genes were identified in whole-exome sequence data from 33 PCa cases from 11 multiplex PCa families. Overall, 3513 candidate gene SNPs were identified, including 805 missense and 9 nonsense variants. Fifty-six missense variants were shared by all 3 family members in each family in which they were observed; 43 of these were observed in all 33 cases and were presumed to be rare variants in the reference sequence and excluded from further analysis. The remaining 13 shared, missense variants as well as the 9 nonsense variants were re-sequenced for verification. Confirmed variants were then tested in all additional family members with available DNA to assess co-segregation with PCa status. All 9 nonsense variants and 12 of 13 shared, missense variants were confirmed. One of the confirmed missense variants was *HOXB13* Gly84Glu which was recently identified as a PCa susceptibility allele. The Gly84Glu minor allele was observed in 1 family where it showed partial co-segregation with disease. Two other missense variants, *BLM* Gln123Arg and *PARP2* Arg283Gln, and 1 nonsense variant, *CYP3A43* Arg441Ter, showed complete co-segregation with PCa status. Mutations in *BLM* are known to cause Bloom Syndrome, a recessive disorder with multiple phenotypic effects including predisposition to various cancers. Additionally, a *BLM* truncating mutation was recently identified as a possible breast cancer susceptibility allele. *PARP2* plays a role in DNA damage repair and has been studied as a therapeutic target in multiple cancers. *CYP3A43* is a member of the cytochrome P450 superfamily which is involved in drug and steroid hormone metabolism, and *CYP3A43* variants have previously been associated with PCa risk and mortality. Four additional variants, *RAD18* Glu453Ter, *IGHMBP2* Val293Ile, *RAD51D* Asn138Ser and *POLE* Ala31Ser, displayed partial co-segregation with PCa. Further research is needed to determine the exact role of these variants in PCa susceptibility.

3298T

Development of a Next Generation Sequencing panel to assess hereditary cancer risk that includes clinical diagnostic analysis of the *BRCA1* and *BRCA2* genes. *B. Roa[1], K. Bowles[1], S. Bhatnagar[2], N. Gutin[1], A. Murray[1], B. Wardell[2], M. Bastian[2], J. Mitchell[2], J. Chen[2], T. Tran[2], D. Williams[2], J. Potter[2], S. Jammulapati[2], M. Perry[2], B. Morris[2], K. Timms[2].* 1) Myriad Genetic Laboratories, Inc., Salt Lake City, UT; 2) Myriad Genetics, Inc., Salt Lake City, UT.

Approximately 7% of breast and 11-15% of ovarian cancers are estimated to be due to germline DNA mutations, the majority of which occur in the *BRCA1* and *BRCA2* genes. However, a comprehensive hereditary breast and ovarian cancer risk assessment should include germline sequencing of the *BRCA1* and *BRCA2* genes, as well as additional genes with a known association in breast/ovarian cancer patients. Sanger DNA sequencing has been the gold standard for molecular genetic analysis. However, it is labor-intensive and costly for the analysis of large gene panels. Next Generation Sequencing (NGS) platforms allow for efficient analysis of larger gene panels, but lack of standardization of sample preparation, NGS platforms, and data analysis presents challenges to diagnostic laboratories. Optimized assay design and validation are critical to maximize the analytical sensitivity and specificity of NGS assays, and to ensure high quality interpretation for clinical decision making. We developed a 25-gene NGS hereditary cancer panel that uses RainDance PCR technology for high-throughput sample preparation, Illumina HiSeq and MiSeq NGS technologies, and commercially available and lab-developed informatic tools. The panel and assays were custom-designed to ensure high-quality, clinically actionable results. Initial assessment of analytical sensitivity and specificity was performed by comparing *BRCA1* and *BRCA2*. NGS was performed on 1864 anonymized patient samples, which had previously undergone Sanger sequencing of *BRCA1* and *BRCA2*. Sanger sequencing identified 15,878 variants, of which 681 were unique and 482 were classified as disease-associated mutations. We identified 15,877 variants at an initial sensitivity of >99.99% for *BRCA1* and *BRCA2*. One polymorphic variant was missed due to a variant under the primer. Sensitivity was subsequently optimized through process improvements. No additional variants were found by NGS, yielding a specificity of 100%. This preliminary analysis facilitated assay optimization, and a comprehensive validation of all 25 genes in the NGS panel. This analysis indicates that a NGS gene panel designed to meet rigorous quality standards can be used to provide clinical sequencing results equivalent to those obtained from Sanger DNA sequencing analysis. Utilization of appropriately designed and implemented NGS technology will allow for analysis of larger gene panels, resulting in greater sensitivity without loss of specificity.

3299F

Development of 3C-based target sequencing technology for candidate gene discovery in prostate cancer.
MJ. Du[1], TZ. Yuan[1], RL. Dittmar[1], XY. Huang[1], SN. Thibodeau[2], L. Wang[1]. 1) Pathology, Medical College of Wisconsin, Milwaukee, WI; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

So far, genomewide association studies have identified over 70 chromosome loci (represented by single nucleotide polymorphisms, SNPs) with an increased risk of prostate cancer. However, the biological roles of these loci in prostate cancer development are unclear. As many of these risk-loci have been found in non-coding regions of the genome, with many residing at some distance from any nearby annotated genes, many of these causal variants are believed to regulate target genes through long range interactions. Although chromosome conformation capture (3C) is used to analyze the long range interactions between different fragments, the traditional 3C assay and its variations are not sufficient to identify unknown genomic regions that interact with multiple risk loci in a cost-efficient manner and in a single enrichment assay. Here, we developed a 3C-based multiple targets enrichment assay and applied this assay to the analysis of 3 prostate cancer risk regions at 8q24. The novelty of this assay is to use short oligos as primers and label target fragments by multiple primers extension. Briefly, a standard approach was used to generate a 3C library by using LnCap cell line, followed by sonication and size selection of 3C DNA into 300-500bp fragments. Nine short oligos (in three risk regions at 8q24) were then hybridized to the fragmented 3C library. Primers extension at high annealing temperature (65°C) was performed with the presence of biotin-labeled dCTP. After strepavidin beads purification, a nano scale library construction method was applied to build a sequencing library. qPCR analysis showed up to 1000 fold target enrichment under our current condition. Finally, a perl-based computer algorithm was written to analyze the long range interactions from large scale sequencing data. The sequencing data showed that the risk regions interacted with multiple chromosome loci. However, interactions with MYC gene, 250kb downstream from nearest prostate cancer risk SNP, was modest. This result indicates that the 8q24 risk regions may physically interact and regulate multiple genes across the whole genome and the new multiple primers extension enrichment assay has great potential to detect the long range genetic interactions. Further development of this multiple targets enrichment 3C method will facilitate candidate genes identification and help to define the biological roles of prostate cancer risk SNPs.

3300W

From GWAS to therapy: Fatty acid synthase in uterine leiomyomata.
Z. Ordulu[1], M. Hayden[1], S. Eggert[2], B.J. Quade[2], C.C. Morton[1,2]. 1) Ob/Gyn, Brigham and Women's Hospital, Boston, MA; 2) Pathology, Brigham and Women's Hospital, Boston, MA.

Uterine leiomyomata (UL) pose a major public health problem given high prevalence (>70%) in women of reproductive age and their indication for >200,000 hysterectomies in the U.S. annually. A genome-wide SNP linkage panel was analyzed in 261 white UL sister pair families from the Finding Genes for Fibroids study (FGFF). Genome-wide association studies performed in two independent cohorts of white women identified a SNP with a p-value that reached genome-wide significance (rs4247357, P = 3.05E-08, odds ratio (OR) = 1.299). The candidate SNP was under a linkage peak and in a block of linkage disequilibrium in 17q25.3 which spans the genes fatty acid synthase (FASN), coiled-coil domain containing 57 (CCDC57) and solute carrier family 16, member 3 (SLC16A3). It has been reported previously that FAS transcripts and/or protein levels are upregulated in various neoplasms and implicated in tumor cell survival. By immunohistochemistry (IHC) we found FAS protein levels are elevated (3-fold) in UL in comparison to matched myometrial tissue. We assessed effects of FAS inhibitor C75 on cell counts in primary cell cultures of matched UL and myometrium by Coulter counting and by cell proliferation assays at 24, 72, and 120 hours after treatment with different concentrations of the inhibitor. Cell counts are reduced in a dose and time dependent manner with a half maximal inhibitory concentration (IC50) of approximately 50 uM. Additional samples are being obtained to stratify the results by rs4247357 genotype and FAS expression status. FASN represents the first potential UL predisposition gene identified in white women and the responsiveness of UL to inhibition of FAS can be a potential model for targeted therapy to mitigate the risks and complications of surgical interventions.

3301T

Coding Variants at the Hexa-allelic Amino Acid 13 of HLA-DRB1 Explain Independent SNP Associations with Follicular Lymphoma Risk.
J.N. Foo[1], K.E. Smedby[2], N.K. Akers[3], M. Berglund[4], I.D. Irwan[1], X. Jia[5], Y. Li[1], D.E.K. Tan[1], J.X. Bei[1,11], J. Chang[12], L. Conde[6], H. Darabi[7], P.M. Bracci[8], M. Melbye[9], H.O. Adami[7], B. Glimelius[4], D. Lin[12], W.H. Jia[11], A.L.H. Seow[13], C.C. Khor[1,13], H. Hjalgrim[9], L. Padyukov[10], K. Humphreys[7], G. Enblad[4], C.F. Skibola[6], P.I.W. de Bakker[5], J. Liu[1,13]. 1) Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore; 2) Department of Medicine, Clinical Epidemiology Unit, Karolinska Institutet, Stockholm, Sweden; 3) School of Public Health, Division of Environmental Health Sciences, University of California - Berkeley, Berkeley, CA; 4) Department of Radiology, Oncology and Radiation Sciences, Uppsala University, Uppsala, Sweden; 5) Broad Institute of Harvard and MIT, Cambridge, MA; 6) School of Public Health, Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 8) University of California, San Francisco, Department of Epidemiology and Biostatistics, San Francisco , CA; 9) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 10) Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; 11) State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China; 12) Department of Etiology and Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China; 13) Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, Singapore.

Non-Hodgkin lymphoma represents a diverse group of blood malignancies, of which follicular lymphoma (FL) is a common subtype. Previous genome-wide association studies have identified independent SNPs, in particular rs2647012 and rs10484561, within the HLA class II region significantly associated with FL risk. To determine if coding variants in HLA genes are responsible for these associations, we imputed classical HLA alleles and coding variants in our discovery GWAS samples of 379 FL cases and 791 controls using a phased reference panel consisting of 2,767 unrelated founders of European descent, with genotypes for 2,537 SNPs in the HLA region, 263 classical HLA types determined to 4-digit resolution and 372 variable amino acid positions encoded in the coding regions of the HLA genes for each founder. We conducted global multiallelic tests and/or biallelic trend tests across all the imputed variants and validated the top result in two independent datasets, for a total of 689 cases and 2,446 controls. The strongest association was observed at a single hexa-allelic amino acid 13 position within the HLA-DR peptide binding groove encoded by the HLA-DRB1 gene ($P=2.3\times10^{-15}$). This variant may fully explain the associations observed at this locus, including rs2647012 ($P=4.72 \times 10^{-11}$; after adjustment P= 0.804) and rs10484561 ($P=2.61 \times 10^{-11}$; after adjustment P = 0.356). Conversely, conditioning on genotypes at rs2647012 ($P=1.68\times10^{-7}$), rs10484561 ($P=7.58\times10^{-7}$) or both SNPs (P=0.005) did not fully eliminate the association observed at amino acid 13, suggesting that the genotypes at amino acid 13 are well- but not fully- tagged by these two SNPs. Amino acid 13 is tightly linked to the nearby amino acid 11, and both are previously reported to be associated with rheumatoid arthritis (Raychaudhuri et. al. 2012 Nat Genet 44:291-296). We provide preliminary evidence that the same coding variants may influence FL risk in the Chinese population ($P= 1.70 \times10^{-4}$) and are now expanding our efforts to confirm the association in the Asian population. This coding variant may explain the complex SNP associations identified by GWAS and suggests a common HLA-DR antigen-driven mechanism for the pathogenesis of FL and rheumatoid arthritis. Further trans-ethnic comparisons across European and Asian sample collections is expected to provide further insights into similarities and differences in the genetic factors underlying these complex and heterogeneous malignancies in each population.

3302F

Expression-based classification of melanoma cell sensitivity to temozolomide. *T. Zhang[1], M. Xu[1], C. Sereduk[2], D. Chow[2], J. Trent[2], H. Yin[2], K. Brown[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Cancer and Cell Biology Division, The Translational Genomics Research Institute (TGen), Scottsdale.

While there are multiple promising new immunotherapeutic and targeted approaches for treating melanoma, not all patients are eligible for these therapies and most of those that are ultimately relapse, highlighting a critical need for improved chemotherapeutics. Temozolomide (TMZ) is an orally administered DNA-methylating agent used in the treatment of advanced cancers including melanoma. TMZ has shown some advantage in the clinic in terms of preventing central nervous system relapse and patient quality of life. In order to find novel vulnerabilities that may be exploited to sensitize melanoma cells to TMZ, we utilized pre-existing expression profiling data and known TMZ-sensitivity from 18 melanoma cell lines to develop a classifier of TMZ-sensitivity. We used the support vector machine (SVM) method to build the model, and applied this classifier to predict the TMZ-sensitivity in 25 independent melanoma cell lines. We subsequently validated these predictions by experimentally determining the IC50 of these cell lines to TMZ using a high-throughput cell viability assay. The result showed that more than 80% of melanoma cell lines were accurately predicted. We analyzed the genes differentiating TMZ-sensitive and -resistant melanoma cells using both microarray data, and most recently RNA-sequencing data using Ingenuity Pathways Analysis. We identified two key genes (MITF and EGFR) each known to regulate a large number of the differentially expressed genes. EGFR and a large network of EGFR targets were significantly underexpressed in TMZ resistant lines. At the same time, MITF and a large group of known MITF targets were significantly overexpressed in resistant lines. We also identified E-cadherin (CDH1), which is known to play a key role in invasion and transition from radial to vertical growth in melanoma, as highly upregulated in resistant cell lines. We functionally characterized the role of EGFR in TMZ sensitivity via EGF treatment or overexpression in resistant cell lines, as well as antibody, small molecule, and shRNA inhibition of EGFR in sensitive cell lines. Small-molecule inhibition of EGFR decreases downstream ERK and AKT pathway signaling in both TMZ-sensitive cell lines tested. Our expression profiling data suggest that TMZ sensitivity may be mediated by multiple large gene regulatory networks, and that EGFR may play a role in therapeutic response of melanoma to TMZ.

3303W

Mining of TCGA data links AGR3 dysregulation to surfactant homeostasis in lung cancer. *A. Umesh, J. Delaney, J. Shima, J. Park, R. Wisotzkey, M. Shekar, I. Kupershmidt.* NextBio, Santa Clara, CA.

We previously identified AGR3, a protein disulfide isomerase with unknown physiological function, as an RNA expression marker for triple negative breast cancer based on ESR1/PGR/ERBB2 RNA levels using NextBio (NB) Clinical, a patient-centric analytical platform. Here we investigated AGR3's role in 19 cancers by mining genomic data of 11,000 patients currently integrated into NB Clinical and performed correlations against publically available curated genomic data using NB Research. Analysis of RNA expression data from patients of 19 cancers revealed AGR3 dysregulation in ovarian, kidney, lung, head/neck, breast, uterine, and brain cancers, using a 2-fold cutoff in either direction as compared to normal tissue. These patients were further stratified based on AGR3 over vs underexpression, using a 2-fold cutoff against disease median. Fifty top-ranked dysregulated genes from each cancer totaling 291 genes were identified and imported to NB Research for further insights. The NB Body Atlas application showed our 291 gene signature to most highly correlate with lung amongst all normal tissues, and highest with bronchial and airway epithelial cells amongst cell types. AGR3 itself also had the highest correlation with pulmonary epithelial cells, and was highly correlated with lung cancer when queried with the NB Disease Atlas application. We followed up by assessing the top 50 genes obtained with NB Clinical from lung cancer patients alone. Stratification by AGR3 expression showed concomitant dysregulation of genes involved in pulmonary surfactant homeostasis and secretion, including surfactant proteins (SFTPA-D), surfactant associated-proteins (SFTA1-2), and secretoglobins (SCGB1A1,SCGB3A1-2). Analysis of patient clinical parameters showed decreased AGR3 expression in 66% of current smokers and in 71% of patients with the primitive subtype of lung squamous cell carcinoma (LSCC). AGR3 was increased in 65% of patients with secretory LSCC. Through analysis of individual patients and correlations with public genomic data, we identify dysregulated pulmonary epithelial AGR3 expression as a biomarker in lung cancer with an effect on surfactant homeostasis. This is functionally akin to its paralog, AGR2, which partakes in intestinal mucus production and is reported to be overexpressed and secreted in prostate and pancreatic cancers (Park et al 2008; Zhang et al 2005). These data highlight epithelial secretion-related mechanisms and the AGR genes in cancers.

3304T

Down-regulation of *SIDT1* gene in leukemia. *J. Mazzeu[1], S.G. Magalhães[2], L.H. Sakamoto[2,3,4], R. Camargo[2,3,4], M.B. Daldegan[4,5], B.D. Lima[2].* 1) Ciências genômicas e Biotecnologia, Universidade Católica de Brasilia, Brasilia, DF, Brazil; 2) Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF, Brazil; 3) Hospital da Criança de Brasília José Alencar, Setor de Oncologia e Hematologia Pediátrica Núcleo de Oncologia e Hematologia Pediátrica, Secretaria de Saúde do DF, Brasília, DF, Brazil; 4) Núcleo de Genética, Secretaria de Saúde do DF, Brasília, DF, Brazil; 5) Núcleo de Hematologia e Hemoterapia, Secretaria de Saúde do DF, Brasília, DF, Brazil.

The use of small interfering RNAs (siRNAs) has been a promising therapeutic approach for the treatment of various diseases including leukemias. These small non-coding siRNAs can trigger the sequence-specific degradation of homologous mRNAs. However, cell-specific delivery of small interfering RNA (siRNAs) is the main obstacle to in vivo applications of siRNAs.It has been shown that SID-1 and its human homologue SIDT1 that localize to the plasma membrane facilitate siRNA uptake in human systems.Knowing that leukemic cells are difficult to be transfected and based on the hypothesis that SIDT1 could facilitate siRNAs delivery into these cells to perform post-transcriptional regulation we examined the human *SIDT1* expression levels in myeloid and lymphoid leukemia patients. We have tested 80 samples from patients with ALL, AML and CML, seven leukemic cell lines (697, RS4;11, Nalm-6, Jurkat, REH, HL-60 and K562) and samples of non-leukemic individuals as controls.We investigated *SIDT1* mRNA levels using real-time quantitative PCR (RT-qPCR) . The results of the relative quantification analysis of *SIDT1* gene expression show that the gene was down-regulated in all seven human leukemic cell lines analyzed, particularly for 697, REH and HL-60 cell lineages when compared to the non-leukemic individuals. The K562 cell lineage does not present any expression of *SIDT1* gene, since no mRNA was detected in RT-qPCR ($p<0,05$). Low expression levels of *SIDT1* mRNA were also observed in all samples positive at diagnosis for ALL, AML and CML ($p<0,05$).Several miRNAs (miR-31, miR-155 and miR-564) have been shown to be abnormally down-regulated in CML cell lines and in patients with CML as compared to non-leukemic individuals. Based on this evidence we suggest that down-regulation miRNAs could be result of the absence or lower expression of the gene *SIDT1* that decrease the siRNA transfer between cells. Our results demonstrate that the presence of cellular changes that lead to leukemia causes changes in the expression level of mRNA of a gene that is essential in the activation pathway by interfering RNA transport of exogenous dsRNA. The mechanisms that lead to this decrease are still unknown, as there are only few studies on gene *SIDT1*. However, these studies have great importance for scientific research and for clinical application as *SIDT1* can be exploited as a marker to establish the presence, persistence or reappearance of leukemic hematopoiesis.

3305F

Genetic polymorphisms in the 9p21 region associated with multiple cancers. *W. Li[1], R. Pfeiffer[1], F. Gu[1], Z. Wang[1,2], M. Yeager[1,2], X. Deng[1,2], L. Amundadottir[1], J. Luo[3], S. Chanock[1], N. Chatterjee[1], M. Tucker[1], A. Goldstein[1], X. Yang[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institute of Health, Rockville, MD; 2) Cancer Genomics Research Laboratory, NCI-Frederick, SAIC-Frederick Inc. Frederick, MD; 3) Information Management Services, Inc., Calverton, MD.

Background and aims The chromosome 9p21 region encompasses several tumor suppressor genes and germline genetic polymorphisms in this region have been associated with the risk of several cancers. The goal of this study was to identify common genetic variants in the 9p21 region that were associated with multiple cancers using publicly accessible genome-wide association studies (GWAS) data. **Methods** We analyzed 9p21 SNPs (21,067,104-32,440,834, hg18) from eight GWAS studies with data deposited in the database of Genotype and Phenotype (dbGAP), including esophageal squamous cell carcinoma (ESCC), gastric cancer (GC), pancreatic cancer, renal cell carcinoma (RCC), lung cancer (LC), breast cancer, bladder cancer (BC), and prostate cancer (PC). The number of subjects ranged from 2252 in PC to 7619 in LC. We first performed single SNP analyses for each study separately using logistic regression. For SNPs with $P<0.01$ associated with at least one cancer, we conducted the meta-analyses through a subset-based statistical approach (ASSET) to combine *P* values across studies. We also calculated gene-level *P*-values using a resampling-based adaptive rank truncated product method. Bonferroni correction was used to adjust for the number of SNPs/genes and studies examined. **Results** We identified five SNPs that showed significant associations for LC [rs12686452 (C9orf72), rs12683422 (LINGO2), rs7849984 (LINGO2), rs2492813 (MOB3B), and rs10511838 (intergenic region of LINGO2-ACO1)], two SNPs for ESCC [rs1063192 (CDKN2B), rs2157719 (CDKN2BAS)], two for BC [rs2764736 (intergenic region of ELAV2-TUSC1) and rs1502895 (intergenic region of LINGO2-ACO1)], and one for RCC (rs7033375, intergenic region of C9orf72-LINGO2) (all $P\leq 2.59\times10^{-6}$). In the meta-analysis by ASSET, rs1502895 (C-T) was positively associated with BC and LC ($P=8.16\times10^{-11}$); rs12683422 (C-T) showed positive effect for LC ($P=1.10\times10^{-24}$) but negative for the subset of RCC and BC ($P=1.48\times10^{-4}$); rs10813205 (C-T, intergenic region of LINGO2 and ACO1) was positively associated with the subset of GC, LC, PC and RCC ($P=7.75\times10^{-5}$). At gene level, MOB3B, LINGO2, and C9orf11 were significantly associated with the risk of LC, while CDKN2B, CDKN2A, and CDKN2BAS were associated with the risk of ESCC (all $P\leq 4.70\times10^{-5}$). **Conclusions** Several genetic variants in the 9p21 region were associated with the risk of multiple cancers, suggesting that this region may contribute to a shared susceptibility across different cancer types.

3306W

Cryptic mutations in introns and exons leading to splice alterations in breast cancer genes. *S. Casadei[1], T. Walsh[1], C.H. Spurrel[2], A.M. Thornton[1], J.B. Mandell[1], S.M. Stray[1], M.K. Lee[1], M.C. King[1, 2].* 1) Medical Genetics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA.

Despite the widespread availability of genomic sequence, we still know very little about the frequency and spectrum of deeply intronic variants that alter splicing so as to introduce a stop and thereby increase risk of disease. Such mutations could lead either to extension of an exon (a bleeding exon) or to creation of a new exon from intronic sequence (exonification). We are exploring this phenomenon in 594 families with multiple cases of breast and/or ovarian cancer, for whom all affected relatives are wildtype for all known breast cancer genes, based on evaluation by BROCA (see poster by T. Walsh et al.). Our first experiment has been to mine the complete BRCA1 intronic data that was generated by BROCA. The target region was chr17:41,190,000-41,280,000, which includes the 3'UTR, all intronic sequence, 5'UTR, and bidirectional promoter of BRCA1, and the first exon of NBR2, but not the off-target duplicated pseudogene of BRCA1. We identified every variant in this region that was rare (<0.01 carrier frequency among 3000 persons tested by BROCA; and minor allele frequency <0.005 on dbSNP) and familial (present in at least two affected relatives in its host family). This screen yielded 195 variants in BRCA1 introns. We next used NNSPLICE0.9 to predict the activation of cryptic donor (5'ss) or acceptor (3'ss) splice sites at each intronic variant. Of the 195 rare familial variants, 29 (15%) were predicted to lead to altered splicing; 10 by activating a cryptic 5'ss and 19 by activating a cryptic 3'ss. For example, a variant deep in BRCA1 intron 5 was shared by three cousins with breast cancer in CF232. The variant was predicted to activate a cryptic donor splice site (taaagtag > TAAAlgtaa; 5'ss score 0.09 < 0.98), leading either to an extension of 249 additional exonic nucleotides, or to introduction of a pseudoexon by pairing with an acceptor splice site located 63bp upstream (3'ss: 0.85). In either scenario, an early stop is included in the newly exonic sequence. Of the 29 variants predicted to alter splicing, 15 are located near sites of nucleosome occupancy, 12 are embedded in Alu sequences, and 2 are located <300bp from a constitutive exon. We will test the consequence of each of these variants by RT-PCR molecular analysis of patient RNA. It is important to understand how intronic events affect splicing. Bioinformatics tools can predict novel cryptic exons, which can then be tested by molecular analysis.

3307T

***MEIS1* silencing influences the expression of cancer/testis antigen *SSX2* in myeloid leukemia cell line K562.** *J. Torres-Flores[1,2], A. Aguilar-Lemarroy[1], A. Bravo-Cuellar[1], L.F. Jave-Suarez[1].* 1) Division of Immunology; Centro de Investigacion Biomedica de Occidente; Instituto Mexicano del Seguro Social; Guadalajara, Jalisco, 44340, Mexico; 2) Doctoral Program of Human Genetics; University of Guadalajara, Guadalajara, Jalisco, 44340, Mexico.

Introduction: *MEIS1* is a transcription factor involved in proliferation and differentiation of hematopoietic cells. *MEIS1* deregulation is associated with malignancies, including leukemias. Recent studies indicate that the presence of *MEIS1* in leukemic cells promotes their proliferation. *SSX2* is a cancer/testis antigen normally expressed in human germ cells and several solid tumors. Conversely, *SSX2* has low expression in hematopoietic malignancies. SSX2 is a nuclear protein with transcriptional repressor activity due to possible interaction with *Polycomb* group members; these genes play a role in regulating the expression of the *HOX* genes by epigenetics modulation. **Aim:** To show evidence that *SSX2* can be regulated by *MEIS1* gene expression. **Methods:** As part of a larger research, we modified a cell culture of K562 through shRNA for silencing the expression of *MEIS1* mRNA. Three cultures of K562 cells were transformed: one with a shRNA sequence to interfere with the messenger of *MEIS1* by targeting exon 9, other for exon 13, and another with lentiviral vector empty. RT-PCR expression analysis was performed using three housekeeping genes to normalize results: *ACTB*, *RPL32* and *GAPDH*. Computer analysis was performed using Matinspector v8.0 software. **Results:** K562/LVX+E9 and +E13 knockdown-cell lines showed *MEIS1* subexpression by qPCR by comparing with K562-LVX-. Interestingly, *MEIS1* silencing increased *SSX2* expression. *In silico* analysis revealed three target regions with *MEIS1* binding potential located 3' to the transcription start (positions 1218, 2230 and 3402), these possible binding sites are contained in the intronic regions 1, 3 and 4, respectively. Additionally, a possible binding site was also identified 3023 basepairs upstream of *SSX2* initiation site, this binding site is located in the *LOC791098* pseudogene. **Discussion and conclusions:** *MEIS1* is involved in cell proliferation and differentiation. However, its role in cancer development is poorly understood. Our observations indicate that *MEIS1* down-modulation increase expression of *SSX2*, a cancer-testis antigen found in solid tumors but not reported in hematopoietic malignancies and leukemia-derived cell lines. We can speculate that overexpression of *MEIS1* represses *SSX2* gene, which could compromise the repressor activity of the *PolyComb* group complex. Additional studies at the protein level are needed to confirm our findings.

3308F

Monoallelic epigenetic silencing of the Apc tumor suppressor gene and parent of origin effects in a rat model of colon cancer. *J. Amos-Landgraf[1], A. Irving[2], M. Ford[2], W. Dove[2].* 1) Veterinary Pathobiology, University of Missouri, Columbia, MO; 2) McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, WI.

We have developed a rat genetic model of Familial Adenomatous Polyposis (FAP) that recapitulates many aspects of the human disease and has certain advantages over existing mouse models of colon cancer. The Polyposis in rat colon (Pirc) rat harbors a nonsense mutation at amino acid 1137 of Apc, a gatekeeper of the Wnt pathway involved in over 80% of human colon cancers. The Pirc rat develops the majority of its tumors in the colon, in contradistinction with existing mouse genetic models where tumors form preferentially in the small intestine. The Pirc rat also exhibits a sex bias with males having an increased susceptibility, as is seen in the human disease. We have shown that tumors initiate through both somatic genetic and epigenetic mechanisms, and that genetic background greatly alters the colonic phenotype. We utilized qRT-PCR, aCGH, Sanger sequencing, and quantitative allele-specific pyrosequencing of DNA and cDNA in heterozygous animals to monitor genetic and epigenetic events in tumors. Remarkably, we found 30% of Pirc rat tumors maintained heterozygosity of Apc in their genomic DNA. We speculate that either additional genetic or epigenetic events occur to initiate these tumors. To score allelic ratios in DNA and cDNA from F1 tumors provides a general assay for allele-specific changes in expression, irrespective of mechanism. Half of the tumors in the Pirc rat that maintained heterozygosity of DNA had a strong bias of expression for the mutant allele in cDNA.

To explore the strong influence of genetic background we created two congenic derivatives of the Pirc strain on the BN and ACI genetic backgrounds. The BN-Pirc line is highly resistant to tumor development (male average of 0.75 tumors), while the ACI-Pirc congenic line shows a dramatic susceptibility with males developing over twice as many tumors as the F344N/Tac co-isogenic line (average 30.3 tumors). The ACI-Pirc congenic line crossed with BN develops an average of 3.7 ± 2.1 tumors per animal at 6 months of age. Interestingly, the parent of origin of the ACI allele has a significant impact on tumor number. When the ACI-Pirc allele is inherited maternally the male offspring develop an average of 2.9 ± 1.5 tumors, in contrast to an average of 4.9 ± 2.3 tumors when the Pirc allele is inherited paternally (Wilcoxon rank sum P=0.01). We hypothesize that genetic variation in imprinted genes in the ACI genetic background influence tumor initiation in the progeny of these crosses.

3309W

Evaluation of a cancer gene sequencing panel in a hereditary risk assessment clinic. *Y. Kobayashi[1], A.W. Kurian[2,3], E. Hare[1], M.A. Mills[2], K.E. Kingham[2], A.S. Whittemore[3], V. McGuire[3], G. Gong[3], U. Ladabaum[2], M. Cargill[1], J.M. Ford[2,4].* 1) InVitae, San Francisco, CA; 2) Department of Medicine, Stanford University, Stanford, CA; 3) Department of Health Research and Policy, Stanford University, Stanford, CA; 4) Department of Genetics, Stanford University, Stanford, CA.

Sequencing panels of multiple cancer-associated genes are entering clinical practice, but little is known about the performance and yield of such testing among relevant patient populations. 199 adult female patients were referred to the Stanford Clinical Cancer Genetics Program for clinical assessment of hereditary breast and ovarian cancer risk; 141 had breast cancer and 57 carried known mutations in *BRCA1* or *BRCA2* (*BRCA1/2*). Germline DNA samples from these patients were sequenced on a custom multigene sequencing panel. This panel covered the entire coding region, exon-intron boundaries (+/- 10bp) and any other known pathogenic variants for 43 genes that have published associations with risk of breast, ovarian and other cancers, and an additional 32 cancer-associated variants elsewhere in the genome. A validation set of 200-300 samples is currently being processed. Analytic results for *BRCA1/2* sequencing and pathogenicity interpretations were concordant with prior clinical testing for all patients. Twenty variants that were designated as pathogenic, either based on published literature or due to a novel truncating or splice donor/acceptor effect, were observed in genes other than *BRCA1/2*, including *ATM*, *BLM*, *CDH1*, *CDKN2A*, *MUTYH*, *MLH1*, *NBN*, *PRSS1* and *SLX4*. Many patients carried more than one variant (including variants of uncertain clinical significance). Thirteen patients had pathogenic variants which warranted a change in cancer screening or preventive interventions based on practice guidelines: these patients were invited for confirmatory clinical testing, genetic counseling and screening recommendations. One 53-year old patient with a personal history of breast and endometrial cancers was found to carry a pathogenic *MLH1* mutation; she underwent risk-reducing salpingo-oophorectomy and colonoscopy, with removal of a tubular adenoma.

Among patients referred for clinical evaluation of hereditary breast and/or ovarian cancer risk, a comprehensive sequencing assay for 43-cancer associated genes and 32 cancer-associated variants identified 20 patients [10.1%, 95% confidence interval (CI) 6.5%-15.1%] with pathogenic mutations in genes other than *BRCA1/2*, of which 13 (6.5%, CI 3.8%-11%) mutations prompted a change in cancer screening or prevention. Disclosure of research testing results to participants who donated specimens several years previously appears feasible and well-tolerated.

3310T

Novel Integrative Genomics Approach to Biomarker Discovery in Prostate Cancer. *C. Hicks[1,2,3], L. Miele[1], T. Koganti[1], J. Sitthi-Amorn[1], S. Vijayakumar[3].* 1) Cancer Institute, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216; 2) Department of Medicine, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216; 3) Department of Radiation Oncology, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216.

Recent advances in high-throughput genotyping and reduction in genotyping costs have made possible identification of common genetic variants associated with increased risk of developing prostate cancer using genome-wide association studies (GWAS). However, despite this remarkable success, several challenges remain in translating GWAS discoveries into clinically actionable biomarkers. First, genetic loci associated with prostate cancer generally explain very little of the disease risk. Second, the trait-SNP associations alone do not necessary lead directly to the identification of causal genes and do not provide insights about the broader biological context in which the identified variants and associated genes operate. Here we report the results of a novel integrative genomics approach that combines GWAS information with gene expression data to elucidate the molecular mechanisms underpinning GWAS findings and to identify gene regulatory networks and biological pathways enriched for genetic variants. Our working hypothesis is that genes containing SNPs associated with increased risk of developing prostate cancer are functionally related and interact with one another and their downstream targets in gene networks and pathways. We tested this hypothesis using publicly available GWAS information on 300 SNPs mapped to 175 genes derived from over 350,000 cases and over 350,000 controls and gene expression data derived from 493 samples (347 samples with prostate cancer, 25 samples with metastatic cancer and 121 control samples). We performed both supervised and unsupervised analysis followed by network and pathway analysis using Ingenuity IPA. Supervised analysis revealed 70 SNP-containing genes associated with prostate cancer and 65 genes associated with metastatic disease. Unsupervised analysis using hierarchical clustering revealed that the SNP-containing genes are co-expressed and have similar patterns of expression profiles. In addition, we identified novel genes which are functionally related with SNP-containing genes. Network and pathway analysis revealed gene regulatory networks and biological pathways enriched for SNPs. Among the identified pathways included the Thrombopoietin, androgen, prolactin and the IGF-1 signaling pathways, all of which have been implicated in prostate cancer. Integration of GWAS information with gene expression data provides insights about the broader context in which genetic variants operate.

3311F

Exomic Sequencing to Identify Germline Variants in Familial Melanoma. *H.N. Shabbir[1], M.T Landi[2].* 1) Miami Dade College, Miami, FL; 2) National Cancer Institute, Department of Cancer Epidemiology and Genetics.

Germlines are the source of DNA in all cells. A mutation at the germline level is the first step to developing cancer, and the vast majority of cancer is genetic. Melanoma, the leading cause of skin cancer death, is known to be highly heritable and rare. Using a family model, high risk variants related to melanoma can be identified. The goal of the study is to integrate information from sequencing, epigenetics, and expression to identify functional and regulatory genes that are associated with melanoma. Families with two or more 1st degree relatives with melanoma were considered at high risk and were investigated in this study. Initially, sequencing data of families with 3 or more relatives with the disease were examined and shared DNA variants were selected for further examination. Genetic databases and annotation tools were used to identify genes based on their known gene function and regulation, pathways, and variant conservation. Gene browsers were also used to identify any histone markers, DNA methylation sites, and other epigenetic indicators. Based on our candidate genes, there is a possibility of genetic heterogeneity, in which multiple genes may be responsible for disease susceptibility. Selected candidate genes will undergo fine mapping to further investigate the region and replication in additional families and population studies of melanoma.

3312W

Molecular Classification of Myeloproliferative Neoplasms (MPNs) Patients in the State of Qatar According to World Health Organization (WHO) 2008 Criteria. *N. Al-Dewik[1,2], Q. Fernandes[2], YK. Naidu[2], N. Bonnin[3], E. Verger[3], B. Cassinat[3], JJ. Kiladjian[4], H. El Ayoubi[2], M. Yassin[2].* 1) QMGC, HMC, Doha, Qatar; 2) Molecular Genetics of Cancer Research Lab, NCCCR-HMC, Doha, Qatar; 3) Unité de Biologie Cellulaire, France; 4) Centre d'Investigations Cliniques, Hôpital Saint-Louis, Paris, France.

Background:MPNs are clonal haemopoietic disorders that are characterized by excessive proliferation of one or more of blood lineages. MPNs include PV, ET and PMF which are associated by the presence of JAK2 V617F mutation in about 90% of PV and 50% of ET and PMF. The molecular workup of JAK2 & related gene mutations were included in WHO 2008 as one major criterion for the diagnosis of Ph- MPNs. Aims: To genetically characterize MPN patients (pts) in Qatar according to the latest (WHO 2008) criteria using molecular studies for JAK2 V617F mutation, JAK2 exons 12-15 & MPL (S505N & W515 L/K) mutations. Methods: Blood samples were collected from suspected MPN cases & DNA was extracted. Allelic discrimination assays were used to evaluate point mutations causing JAK2 V617F & MPL (W515 L/K) mutations. JAK2 Exon 12 was analyzed using High Resolution Melting Curve (HRM) assay & Sanger Sequencing. In some cases the entire MPL exon 10 & exons 12-15 was studied by RNA extraction followed by cDNA synthesis, amplification & sequencing. Results: 300 patients were classified into PV, ET and PM. Out of 119 PV, 97% of cases were positive for the JAK2 V617F mutation and 3% of cases were negative for other mutations. Out of 165 ET, 48% of cases were positive for JAK2 V617F, one had MPL S505N mutation and 50% of cases were negative for other mutations. Out of 15 PMF, 33% of cases were positive for JAK2 V617F and one unclassified case was characterized by DVT had JAK2 exon 13 mutation (R564L). Conclusion: This study used novel molecular approaches to confirm the diagnosis of MPNs cases in Qatar. The observed patterns of mutations were found to be similar to the international data. In our cohorts of patients, JAK2 V617F mutation was found to be present in almost every patient with PV, nearly 50% of ET patients and less than 50% of PMF patients due to the low number of PMF patients in this study. Our original findings show the presence of MPL S505N mutation in one ET patient which was reported both as an inherited or acquired mutation in very rare cases of ET and R564L mutation in one unclassified MPNs case. An ET patient with MPL S505N progressed to AML. The impact of R564L mutation found in 1 atypical case is still unknown & needs further investigation. We report for the first time the presence of most frequent mutations found in MPNs patients in Qatar. This study is preliminary before further molecular investigations to explore & identify mutations in other candidate genes.

3313T

Novel association between tanning addiction and *PTCHD2* alleles using a whole-exome variant array. *A.E. Bale[1], A.T. Dewan[1], L.M. Ferrucci[1], J. Gelernter[1], D.J. Leffell[1], S.T. Mayne[1], J. Stapleton[2], B. Cartmel[1].* 1) Yale Univ, New Haven, CT; 2) Cancer Institute of New Jersey, New Brunswick, NJ.

Skin cancer is the most common malignancy worldwide; and its rapidly increasing incidence in young adults is related to a preventable exposure, excess ultraviolet (UV) radiation from indoor and outdoor tanning. Several studies have provided evidence for biological dependence on the physiologic effects of UV, analogous to substance dependence. We hypothesized that tanning addiction, like substance addiction, may have a genetic underpinning. European Americans under age 40 in Connecticut who participated in a case-control study of basal cell carcinoma (BCC) were recontacted to complete an online survey (response rate=82%) that included 3 tanning addiction questionnaires to assess lifetime symptoms of tanning dependence. DNA collected as part of the BCC case-control study was genotyped for 319,000 variants~mainly exonic~on the Affymetrix Axiom 319 Array. Of 476 subjects who passed QC, 79 exhibited symptoms of tanning dependence on all 3 scales and 213 were negative for tanning dependence on all scales; individuals with an intermediate phenotype (tanning dependent on one or two scales) were excluded. Including missense, nonsense and splice site changes, there were 157,955 variants across 14,903 genes. A gene-based burden analysis was performed in which each sample was dichotomously scored for each gene as either having one or more minor alleles or no minor alleles. For each gene, differences in these frequencies by tanning dependence underwent a Fisher's exact test. For one gene, *PTCHD2*, we observed significantly more individuals without symptoms of tanning dependence with a minor allele, compared to individuals who exhibited symptoms of tanning dependence (OR = 0.27, $p=2.5\times10^{-6}$, p value for significance with Bonferroni correction=3.36×10^{-6}). To validate this method of analysis, we evaluated Axiom Array data on BCC cases (n=354) and controls (n=363) from the parent study. The top gene in this BCC risk analysis was *MC1R* ($p=1.67\times10^{-6}$; OR=2.07), as expected based on our previous analysis of this same sample in which we directly sequenced *MC1R*. *PTCHD2* is a gene of unknown function whose product is expressed mainly in the brain. While requiring replication, the association between variants in this gene and reduced risk for tanning addiction supports a biological basis for this behavioral trait. Delineating molecular mechanisms that drive tanning dependence may lead to novel medical interventions to stem the epidemic of tanning and skin cancer in young people.

3314F

The PROFILE Feasibility Study: Genetic prostate cancer risk stratification for targeted screening. *R. Eeles[1,2], E. Bancroft[1,2], N. Taylor[2], T. Dadaev[1], E. Page[1], D. Keating[1], N. Borley[3], N. Desouza[4], C. Goh[1,2], E. Saunders[1], A. Lee[5], D. Easton[5], A. Antoniou[5], Z. Kote-Jarai[1], E. Castro[1,2].* 1) Oncogenetics Team, Div of Genetics & Epidemiology, Inst Cancer Res, Sutton, Surrey, United Kingdom; 2) Cancer Genetics Unit. The Royal Marsden NHS Foundation Trust, London, UK; 3) Academic Urology Unit. The Royal Marsden NHS Foundation Trust, London, UK; 4) Radiology Department. The Royal Marsden NHS Foundation Trust, London, UK; 5) Center for Cancer Genetic Epidemiology, University of Cambridge, UK.

Background: Prostate cancer (PC) screening is controversial and a better assessment of individualized PC risk is needed. Several single nucleotide polymorphisms (SNPs) conferring a cumulative risk of PC have been identified. We have explored the potential role of genetic markers for targeted screening in a population with increased risk of PC due to family history (FH) of the disease. **Methods:** PROFILE was developed as a pilot study to determine the feasibility of targeted PC screening using prostatic biopsy (PB) and its association with specific genetic profiles in men with FH. We also evaluated the role of PSA and Diffusion Weighted MRI (DW-MRI) as screening tools in this population. 100 men aged 40-69 with FH of PC were enrolled. Cumulative SNP risk scores were calculated by summing 59 risk alleles for each locus using the weighted effect (log-additive model). DW-MRI was performed in 50 patients. Participants were asked to undergo a 10 core PB regardless of baseline PSA. **Results:** 35% of invited men entered the study. Median age 53 yrs (40-69) and median PSA was 1.15. Ninety men accepted to undergo a PB as primary PC screening. Twenty-two tumours were found and 45% of them were clinically significant [Median age 64yrs (47-69), median PSA 5.4 (0.91-9.3)] which compares with 24% in general population screening. The predictive performance of DW-MRI, PSA, genetic model and genetic model plus PSA measured by AUC were: 0.85, 0.73, 0.57 and 0.74, respectively. The genetic model performed best in men with a normal PSA of <3(AUC 0.63). Analyses of a 78 SNP profile from the recent COGS results are underway. **Conclusions:** Our results indicate that PB is acceptable for PC screening in men with FH of PC. The numbers of men undergoing DW-MRI were small but the AUC would warrant a larger study. The SNP risk score was more predictive in men with PSA<3 where PB would not normally be undertaken, therefore an expanded study to investigate the role of genetic profiling in directing PB in PC screening is indicated.

3315W

Association between colorectal cancer risk and variants in the human exome. *T.A. Harrison[1], M. Lemire[2], F.R. Schumacher[3], P.L. Auer[4], C. Qu[1], S.I. Berndt[5], S. Bézieau[6], H. Brenner[7], P.T. Campbell[8], A.T Chan[9], J. Chang-Claude[10], D.V. Conti[3], D. Duggan[11], S. Gallinger[12], R.C. Green[13], S.B. Gruber[3], E. Jacobs[8], S. Jiao[1], L. Le Marchand[14], L. Li[15], P.A. Newcomb[1, 20], R.E. Schoen[16], D. Seminara[17], M.L. Slattery[18], J.D. Potter[1, 19], E. White[1, 20], U. Peters[1, 20], G. Casey[3], L. Hsu[1], T.J. Hudson[2, 21].* 1) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Ontario Institute for Cancer Research, Toronto, ON, Canada; 3) Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI; 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 6) Service de Génétique Médicale, CHU Nantes, Nantes, France; 7) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; 8) American Cancer Society, Atlanta, GA; 9) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 10) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, German; 11) Translational Genomics Research Institute, Phoenix, AZ; 12) Department of Surgery, Toronto General Hospital, Toronto, ON, Canada; 13) Memorial University of Newfoundland, St. John's, NL, Canada; 14) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 15) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; 16) Department of Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA; 17) Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD; 18) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT; 19) Centre for Public Health Research, Massey University, Wellington, New Zealand; 20) School of Public Health, University of Washington, Seattle, WA; 21) Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Toronto, ON, Canada.

Although the use of genome-wide association studies (GWAS) has resulted in notable successes in identifying common single nucleotide polymorphisms (SNPs) associated with colorectal cancer (CRC), genotyping platforms to date have primarily focused on common alleles with frequencies greater than 5%. The Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) investigated the spectrum from common to rare genetic variation in the human exome by successfully genotyping 10,105 cases and 10,009 controls of European ancestry on the Illumina Human Exome array with over 240,000 markers targeting rare non-synonymous, splice site, and stop variants. We examined associations between CRC and 61,229 single variants with minor allele frequencies (MAF) $\geq$ 0.1% using logistic regression analysis and a Bonferroni-corrected p-value threshold of $0.05/61{,}229 = 8.2 \times 10^{-7}$. Additionally, we assessed CRC risk in relation to aggregation of variants with MAF < 5% in 14,570 genes using MiST, which combines burden and variance component tests (Bonferroni-corrected p-value threshold of $0.05/14{,}570 = 3.43 \times 10^{-6}$). In addition to statistically significant associations with known GWAS loci for colorectal cancer, we identified potentially novel variants with p-values $< 1 \times 10^{-4}$ for follow up. After conducting a combined meta-analysis with an additional 5,825 cases and 7,814 controls of European descent from the Colorectal Cancer Family Registry (CCFR) study and a Kentucky population-based case-control study with Exome-array data, as well as additional GECCO studies with 1000 Genome Project imputed data, we identified a novel marginally significant association for a low frequency, missense variant (MAF = 0.046) in *HIVEP1* [discovery p-value = 1.54×10^{-5}, OR = 1.23; combined p-value = 5.5×10^{-7}, OR = 1.21]. None of the gene-based tests remained statistically significant after accounting for multiple comparisons. In summary, our study provides support for an association between a novel low-frequency variant in *HIVEP1* and CRC risk, and suggests that even larger sample sizes are needed to identify less frequent and rare variants in the human exome that contribute to the development of CRC.

3316T

Pathway analysis shows different leukemia subtypes are involved with distinct biological mechanisms. *Ll. Hsu[1], FB. Briggs[2], XR. Shao[2], AP. Chokkalingam[1], KM. Walsh[3], C. Metayer[1], LF. Barcellos[2], P. Buffler[1].* 1) School of Public Health, University of California, Berkeley, Berkeley, California, CA, USA; 2) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA.

The incidence of acute lymphoblastic leukemia (ALL) has been found to be nearly 20%; higher among Hispanics than non-Hispanic Whites. However, only a few studies have been conducted in the Hispanic population. Genome-wide association studies (GWAS) have showed evidence for association in IZKF1, ARID5B, CEBPE, CDKN2A, and BM1-PIP4K2A. However, these loci account for <10%; of leukemia genetic risk, indicating additional susceptibility loci are yet to be discovered and advanced bioinformatics tools may further guide future directions. We applied pathway-based analyses in Hispanic GWAS data of the California Childhood Leukemia Study (CCLS), testing for biological functions that are significantly enriched in childhood leukemia. Furthermore, we compared whether different biological pathways were overrepresented in major disease subtypes, including B-cell ALL, hyperdiploid B-ALL and TEL-AML1 ALL. The study population is comprised of 323 Hispanic ALL cases and 454 controls from the CCLS, using Illumina OmniExpress v1 genotyping platform. For pathway analyses, we selected genes that had at least one associated significantly SNP (P<0.001), when adjusted for age, gender, and genetic ancestry. Of the 187 genes identified, 185 genes were incorporated for analysis using a hypergeometric test to compare the submitted list to a reference of all human genes using WebGestalt v.2. The top five overrepresented KEGG pathways in ALL include axon guidance (FDR=5.1e-06), protein digestion and absorption (FDR=0.0007), melanogenesis (FDR=0.0014), leukocyte transendothelial migration (FDR=0.0021), and focal adhesion (FDR=0.0021). Interestingly, 90% of the identified pathways are associated with cancer development, such as Wnt signaling and MAPK pathway. Between different disease subtypes, pathway analyses results indicate that hyperdiploidy B-ALL and TEL-AML1 ALL involve distinct biological mechanisms compared to ALL. This is the first study to show distinct biological pathways are overrepresented in different leukemia disease subtypes using pathway-analyses approaches. We further used LASSO and targeted maximum likelihood estimation method to identify causal SNPs within each pathway. The results demonstrate that newly developed bioinformatics tools and causal inference methods can help identify novel loci to further understand leukemia pathogenesis.

3317F

A Genome-Wide Association Study Identifies New Susceptibility Loci for Esophageal Adenocarcinoma and Barrett's Esophagus. *D.M. Levine[1], W.E. Ek[2], R. Fitzgerald[3], S. MacGregor[2], D.C. Whiteman[4], T.L. Vaughan[5].* 1) Department of Biostatistics, University of Washington School of Public Health, Seattle, WA USA; 2) Statistical Genetics, Queensland Institute of Medical Research, Queensland, Australia; 3) MRC Cancer Cell Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, UK; 4) Cancer Control, Queensland Institute of Medical Research, Queensland, Australia; 5) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA USA.

Esophageal adenocarcinoma (EA) is a cancer with rising incidence and poor survival. Most patients survive less than a year, and five-year survival is about 20%. For reasons that are not clear, incidence of EA is substantially higher in men and persons of European ancestry. Most EAs arise in a specialized intestinal metaplastic epithelium, which is diagnostic of EA's precancerous lesion, Barrett's esophagus (BE). We report the results of the first genome-wide association study of EA, and the first to examine EA together with its precancerous lesion BE. We utilized fourteen, largely population-based, epidemiologic studies from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON) collected over the past two decades. We analyzed 5,564 cases (2,393 EA and 3,175 BE) and 10,120 controls in two phases. For the combined case group we identified three novel genome-wide associations. The first is on 19p13 ($P=3.6\times10^{-10}$) in CRTC1, a putative cancer gene. A second is on 9q22 ($P=1.0\times10^{-9}$) in BARX1, which encodes a transcription factor important in esophageal specification. A third is on 3p14 ($P=5.5\times10^{-9}$) near the transcription factor gene, FOXP1, which regulates esophageal development. Consistent with our previous findings showing extensive polygenic overlap between EA and BE, our most significant results were for EA and BE combined, and, together, suggest that much of the genetic basis for EA lies in the development of BE, rather than progression from BE to EA. Further dissection of these loci is likely to lead to insights into the etiology of this rapidly fatal cancer.

3318W

GENE EXPRESSION NETWORKS FOR NON-SMALL CELL LUNG CANCER MAJOR SUBTYPES. *L. Liang[1, 2], M.B. Freidin[3, 4], E. Lim[3, 4], A.G. Nicholson[5], S. Popat[3, 6], M.F. Moffatt[3], W.O. Cookson[3].* 1) Epidemiology, Harvard Sch Public Health, Boston, MA; 2) Biostatistics, Harvard Sch Public Health, Boston, MA; 3) National Heart and Lung Institute, Imperial College London, London, UK; 4) Department of Thoracic Surgery, Royal Brompton Hospital, London, UK; 5) Department of Histhopathology, Royal Brompton Hospital, London, UK; 6) Lung Unit, Royal Marsden Hospital, London, UK.

Lung cancer (LC) remains the most common cancers worldwide for decades, and causes more than a million deaths annually. LC was in focus of numerous genetic and genomic studies including global transcriptome profiling for the purpose of identification of robust expression signatures for early diagnostics, molecular classification, disease prognosis, and response to treatment regimes. In this study we used systems biology approaches to identify transcription networks as well as individual loci that provide the way to explain transcriptome complexity of LC. Lung tumour and matched normal specimens from 69 patients with squamous cell carcinoma (SQCC, 25 patients) and adenocarcinoma (ADC, 44 patients) were analyzed. We found that gene expression profile can remarkably distinguish between tumour subtypes. Among the 36 co-expression network modules identified, three modules are highly significantly (pvalue<1e-8) associated with both subtypes, one module with ADC only and two modules with SQCC only. These networks contain very interesting patterns of genes. Controlling for false discovery rate < 0.001, we identified 2869 transcripts were differentially expressed between tumour subtypes and 69 transcripts were associated with smoking status. The hub genes of identified network modules that were consensus or specific to tumour subtypes reveal important biological pathway and molecular signature. To the best of our knowledge, this is the first study utilising co-expression network algorithm for the analysis of co-expression networks in non-small cell lung cancer. Our results might reveal new biomarkers to distinguish between ADC and SQCC, and identify new potential therapeutic targets.

3319T

HPV and Cervical Cancer in Guatemala and Venezuela-Low-cost screening with NextGen sequencing. *H. Lou[1], G. Villagran[2], U. Odey[3], J. Sawitzke[4], D. Wells[4], J. Troyer[4], M. Dyba[5], A. Ruch[6], R. Orozco[7], V. Arguata[7], E. Gharzouzi[2], E. Alvirez[3], M. Dean[5].* 1) SAIC Genetics Core, CIP, CCR, NCI, Frederick MD; 2) Instituto Cancerologia, Guatemala City, Guatemala; 3) Hospital Central Universitario 'Dr. Antonio M Pineda', Barquisimeto, Lara State. Venezuela; 4) SAIC-Frederick, Frederick MD; 5) LEI, CCR, NCI, Frederick MD; 6) Toledo Hospital, Toledo, Ohio; 7) Hospital San Juan de Dios, Guatemala City, Guatemala.

High-risk human papillomavirus infections are causally related to cervical cancer development. Cervical cancer represents the most dramatic cancer health disparity of women in the world, with over 200,000 deaths annually, concentrated in poor, rural, and indigenous populations. Guatemala and Venezuela are illustrative of this disparity as cervical cancer is the predominant cause of cancer cases and deaths in women. To determine the HPV viral strains in invasive tumors, we initiated a prospective collection of invasive cervical cancer cases with tissue collected in RNAlater. DNA-based testing has been shown to be useful in identifying persistent carriers at risk for cervical cancer. To determine the prevalence of HPV in women over 30, cervical swabs were collected at the time of a Pap exam and preserved in 3ml of Scope mouthwash. A total of 351 cancer cases with an average age of 53 (22-81 years), and 395 swabs of non-cancer cases over age 30 were obtained. We amplified the HPV L1 gene from DNA purified from cancer tissues and cell lysates from cervical swab samples using BS GP5+ and GP6+ primers and nested PCR was used to confirm the undetermined samples by primer MY09/11 (1st run PCR) and BS GP5+ and GP6+ (2nd) run PCR. DNA sequencing of HPV from tumor tissue revealed detectable in HPV in 96% of samples with HPV16 in 53%, HPV18 in 12%, HPV45 in 10% and HPV 26, 31, 33, 35, 39, 52, 56, 58, 67, 68, 69, and 73 in 0.8-4.9% of samples. The prevalence of HPV infection in women without cancer is approximately 12%. To employ NextGen sequencing for HPV determination we bar-coded the same BS GP5+ and GP6+ primers and sequenced on the Ion Torrent PGM. An average of 8000 reads were obtained for each sample. Highly comparable data was obtained as from the conventional sequencing and mixtures of HPV types as well as polymorphisms within the isolates was observed. From the HPV16 and HPV18 samples, oncoprotein E6 and E7 RNA was quantitated to estimate viral load. In conclusion, vaccination against HPV16 and HPV18 can protect for up to 70% of invasive cervical cancers, but active screening will be required for many years. A sensitive and low cost approach to cervical sampling, including Next-Gen sequencing has been developed that could be employed to reduce the burden of cervical cancer worldwide.

3320F

Individualized prostate cancer risk prediction using a polygenic risk score. *R. Szulkin, T. Whitington, F. Wiklund, PRACTICAL consortium.* Department of Medical Epidemiology and Biostatistics,Karolinska Institutet, Stockholm, Sweden.

Introduction: Multiple common prostate cancer susceptibility single nucleotide polymorphisms (SNPs) have ben established. Their ability to individually discriminate between prostate cancer patients and healthy controls are limited due to small effect sizes; however, combinations of multiple SNPs have shown predictive capacity. The aim of this study was to develop an optimal polygenic risk score that distinguish between prostate cancer cases and controls, as well as between aggressive and non-aggressive disease.

Materials and methods: Utilizing a large international prostate cancer genetics consortium (PRACTICAL) approximately 83,000 SNPs, enriched for prostate cancer association, genotyped in 22,924 cases and 22,889 controls were available for analysis. Prostate cancer cases were classified as having aggressive (Gleason score>7) or non-aggressive (Gleason score <7) disease. To discriminate between cases/controls and aggressive/non-aggressive disease we constructed polygenic SNP risk scores that were assessed in logistic regression models. Increasing number of SNPs were added to the score, ranked according to strength of association with prostate cancer, and evidence of biological function extracted from public resources including the ENCODE project. To select the optimal number of SNPs in the score and to avoid over fitting, a five-fold cross validation procedure was employed in a 'training dataset' (85% of data) and then evaluated in a 'test dataset' (15% of data).

Results and conclusion: Initial analysis show that prostate cancer prediction based on a risk score including 54 established prostate cancer susceptibility SNPs result in an area under the curve (AUC) of 0.66. By adding an additional 80 SNPs into the risk score, selected by high association with prostate cancer, the predictive performance increased notably (AUC=0.70). We anticipate that this result can be further improved by incorporating functional information to rank SNPs. We show that a polygenetic risk score may be a clinically useful tool to identify individuals who would benefit from targeted screening.

3321W

BRCA1 and BRCA2 (BRCA) gene analyses on an economic platform: A global consortium to demonstrate the feasibility of a shared, dedicated workflow for non-optical next generation sequencing (NGS) with a custom BRCA AmpliSeq kit on the Ion Torrent PGM™. *J. Weitzel[1], J. Costa[2], A. Mensenkamp[3], A. Ekici[4], J. Herzog[1], M. Ligtenberg[3], H. Feilotter[5], P. Park[5], A. Hidalgo-Miranda[6], N. Williams[7], R. Ellis[7], J. Carlos-Machado[2].* 1) Clinical Cancer Gen, City of Hope, Duarte, CA; 2) Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; 3) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Netherlands; 4) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany; 5) Department of Pathology and Molecular Medicine Queen's University, and Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 6) Cancer Genomics Laboratory National Institute of Genomic Medicine, Distrito Federal, Mexico City, Mexico; 7) West of Scotland Genetic Services, Laboratory Medicine, Southern General Hospital, Glasgow, Scotland, UK.

While the 'Jolie effect' has refocused attention on the central role of BRCA gene analyses in the diagnosis and prevention of hereditary breast and ovarian cancer, there is a global disparity in access to affordable testing. The development of bench top NGS technologies holds promise for faster, more comprehensive and cost-effective methodologies than Sanger sequencing. We describe here a global consortium developing and demonstrating the feasibility of a shared, dedicated workflow for clinical grade BRCA gene analyses using Ion AmpliSeq™ multiplex PCR technology combined with Ion PGM™ System. The non-overlapping primers were designed to provide 1) 100% coverage of all coding exons and exon-intron boundaries; 2) overlapping amplicons covering exons; 3) no SNPs in the last five nucleotides of primer; and 4) maximum of three non-validated SNPs per primer. The technical protocol was developed and piloted by centers in the Netherlands and Portugal, and the current phase III testing includes 5 additional centers (California, USA; D.F., Mexico; Glasgow, Scotland; Erlangen, Germany; Ontario, Canada). More than 200 different known germline BRCA mutation positive cases were selected (~30 per center), many representative of a given region, with the aim to assess the ability to detect and call the full spectrum of mutation types, with additional cases in or within close proximity to homopolymer regions. Data analyses include independent and blind evaluation and power estimation of the new methodology. All samples were studied using both Sanger sequencing and the Ion PGM™ System. For data-analysis various software packages were evaluated, with an aim to harmonize the analyses as a custom workflow on Ion Reporter to facilitate use in widely disparate settings. The technical platform has been established in all 7 centers, with excellent performance characteristics: 100% coverage of the targeted regions; >100x coverage across all but 1 amplicon in BRCA2 (>40x), while using barcoding to analyze 8 cases simultaneously on a 316 chip. There was high sensitivity and specificity; technical challenges in the bioinformatic component and emerging copy number variation detection methods will be discussed. This work demonstrates the potential for mutational screening of BRCA1 and BRCA2 using the Ion AmpliSeq™ technology combined with the Ion Torrent PGM, and the feasibility of deploying a common protocol to diverse geographic settings, with a close collaboration among peers.

3322T

Inherited predisposition to cancer in Mexican women. *C.M. Laukaitis[1,2], A. Chaudhury[1], C. Mauss[2], T. Walsh[3], P.A. Thompson[2], A.M. Lopez[1,2], A. Daneri Navarro[4], M-C. King[3].* 1) Department of Medicine, University of Arizona, Tucson, AZ., USA; 2) University of Arizona Cancer Center, Tucson, AZ, USA; 3) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA; 4) Centro Universitario De Ciencias De La Salud Universidad De Guadalajara Sierra Mojada No 950, Edificio P Primer Niviel, Colonia Independencia Guadalajara, Jalisco, Mexico.

With the emergence of comprehensive sequencing approaches for evaluation of breast and ovarian cancer genes, it is our hope to offer risk assessment for these cancers as part of an integrated program of cancer prevention for Latina women in the U.S. Southwest and Mexico. However, our experience suggests that genomic analysis of cancer risk in this population presents special challenges, because Hispanic control populations have been less thoroughly evaluated than others for naturally occurring variation in critical genes. The goal of the present pilot project is to apply BROCA capture and sequencing to identify and characterize inherited damaging mutations of all classes in all known breast and ovarian cancer genes. Participants in the study are 92 women of Mexican ancestry (49 from southern Arizona and 43 from northern Mexico) who were diagnosed with invasive breast cancer and enrolled in the ELLA Binational Breast Cancer Study. Of the participants, 15% (14/92) carry unambiguously damaging mutations (truncations, complete gene deletions, and missense mutations shown experimentally to lead to loss of protein function) in a known breast cancer gene: 5 in BRCA1, 5 in BRCA2, 2 in CHEK2, 1 in PALB2 and 1 in RAD51C. Several of these clearly damaging mutations are population specific.The challenge arises because an additional 9% of participants (8/92) carry other rare, and hence likely population-specific, mutations of unknown functional consequence in the same genes. Four participants carry variants in BRCA1 or BRCA2 at sites predicted by in silico tools to alter splice enhancers. Four other participants carry missense mutations in CHEK2 that are predicted by in silico tools to be damaging to kinase function. None of these variants appear in public databases of the 1000 Genomes Project, in dbSNP, or in the NHLBI exome sequencing project; none are characterized functionally in gene-specific databases. The problem of variants of unknown significance arises in all populations, certainly, but in our experience it is disproportionately frequent in populations with little previous evaluation for naturally occurring benign variation. Both functional and population genetics approaches will be important in resolving the meaning of this variation.

3323F

Association of Genetic variants of Cancer stem cell genes (CSC) in Gallbladder Cancer Susceptibility in North Indian population. *K. Sharma[1], B. Mittal[1], A. Yadav[1], A. Kumar[1], S. Misra[2], V. Kumar[2].* 1) Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences (SGPGIMS), Lucknow, UP, India; 2) Surgical Oncology. King George Medical University, Lucknow,India.

Introduction: Many properties of malignant cells are suggestive of those in normal stem cells. Current evidence suggests that cancer stem cells (CSC) are accountable for cancer progression and recurrence. Germline variants in CSC genes may bring about distorted gene function and activity, thus causing interindividual variations in a patient's chemoresistance and tumor recurrence capability. Still, the associations of genetic variations in these genes with GBC have not yet been done. Methods: The present study included 598 subjects including 398 GBC patients and 200 healthy controls from North India. This study examined association of ALCAM rs1157 G>A, CD44 rs187116 G>A, CD44 rs187115 T>C, ALDH1A1 rs13959 A>G. Genotypes were determined by self designed PCR-RFLP and ARMS-PCR. Statistical analysis was done by SPSSver16. In-silco analysis was performed using Bioinformatics tools (F-SNP, FAST-SNP). Results: Statistical analysis by logistic regression showed marginal association of ALCAM rs1157 G>A polymorphism with GBC risk (p=0.05). However CD44 rs187116 G>A, CD44 rs187116 T>C, ALDH1A1 rs13959 A>G were not associated with GBC risk. CD44 rs187116 G>A, CD44 rs187116 T>C haplotypes did not show significant association with GBC. On stratification based on gender, ALCAM rs1157 [AA] genotype showed increased risk of GBC in females [p=0.017] and CD44 rs187116 [GA] genotype with males [p=0.017]. Subdividing the GBC patients on the basis of gallstone status, ALCAM rs1157 G>A polymorphism imparted higher risk in patients with stones when compared to controls [p=0.04]. In case-only analysis, risk was not modified with tobacco usage and age of onset. Insilico analysis of ALCAM rs1157 G>A and CD44 rs187116 T>C revealed variable change in transcriptional and splicing regulation respectably. Conclusion: This is the first report of association of cancer stem cell genetic variants with GBC susceptibility. The study found important role of ALCAM rs1157 G>A with GBC risk. Grant support: DBT, DST and ICMR Government of India.

3324W

Semiconductor-based next-gen sequencing reveals cancer risk-associated genetic variations in Chiba J-MICC cohort study. *J. Katayama[1], Y. Hayashi[1], H. Mikami[2], H. Kageyama[2], M. Ohira[2], S. Yokoi[2], H. Nagase[2].* 1) Life Technologies, Tokyo, Japan; 2) Chiba Cancer Center research Institute, Chiba, Japan.

Ion AmpliSeq technology is a simple, efficient and rapid process for enriching hundreds to thousands of genomic targets for next generation sequencing from 10 ng of FFPE or whole genomic DNA in a single tube. We conducted this next-gen sequencing approach to identify risk-associated genetic variants primarily in cancer risk population in a population-based cohort study following 7,900 of healthy people in a local area of Chiba since 2005. To focus the screening for cancer related gene, we used comprehensive cancer panel with Ion personal genome machine sequencer for this screening. We evaluated sufficient data amount for detection of germ-line SNVs by comprehensive cancer panel on 318chip. A preliminary study using 20 risk for stomach cancer and 8 healthy individuals by 4 barcoded run reveal over 80 percent region covered by X20 coverage. (Mean: 88.03, Range: 81.17-93.46). So, we concluded 4 samples parallel run has sufficient data amount for germ-line SNVs screening. The comparison of germ-line SNVs between risk and healthy control, we identify more frequent unknown variants and non-synonymous germ-line SNVs. These variants need the further evaluation by conventional method to avoid false positive. And also these several common variants in risk group but not in healthy group are further validating in the expanded risk and healthy population. The ability to examine 409 high-profile genes at once empowers that researchers may realize a more complete picture of the genetic variation that associated with cancer risk in a restricted population.

3325T

A support vector machine classifier for estrogen receptor positive and estrogen receptor negative early-onset breast cancer. *R. Upstill-Goddard[1], D. Eccles[1], S. Rafiq[1], W. Tapper[1], J. Fliege[2], A. Collins[1].* 1) Human Genetics and Cancer Sciences, Faculty of Medicine, University of Southampton, Duthie Building (Mailpoint 808), Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, United Kingdom; 2) Centre for Operational Research, Management Science and Information Systems, University of Southampton, Mathematics Building (54), Highfield Campus, Southampton, SO17 2JL, United Kingdom.

A major breast cancer subtype distinction depends upon the expression of receptors for the hormone estrogen on tumour cells. Tumours that express large numbers of estrogen receptors are termed estrogen receptor-positive while cancers expressing few receptors are estrogen receptor-negative. The influence of germline genetic variants on driving the development of either of these cancer subtypes in an individual is poorly understood. A novel approach was used to explore genome-wide single nucleotide polymorphism (SNP) data from breast cancer samples; a machine learning algorithm, called a support vector machine (SVM), was used to build a classification model capable of accurately distinguishing between estrogen receptor-positive and -negative breast cancer cases based on this SNP data. Machine learning methods have seldom been applied to genome-wide data in this way. A sample of 542 breast cancer patients, all of whom were diagnosed before the age of 40, were genotyped for ~500 000 genome-wide SNPs. A subset of 200 SNP variants was selected from the full set of SNPs based on a strong association with the receptor-negative phenotype ($p<0.0005$), determined using a chi-squared association test. Five SVM classification models, each using a different choice of kernel mapping function, were produced from the genotype data for the 200 SNPs in all 542 samples. In all cases classification accuracy exceeded 90%. A set of 139 genes are associated with the subset of 200 SNP variants and are thus implicated in the estrogen receptor phenotype in early-onset breast cancer. Functional classification of these genes identified enrichment for functions of the immune system. This finding is consistent with the role of inflammation in cancer and observations of a particular subtype of estrogen receptor-negative breast cancer that exhibits a good prognosis alongside extensive immune cell infiltration into the tumour site. Further exploration of the influence of SNPs on the estrogen receptor subtype distinction is on going, with particular focus on SNPs in immune system pathways. Understanding the relationship between the underlying patient genome and the process of oncogenesis is critical for the development of appropriate targeted treatment and disease prevention.

3326F

Inversion-mediated gene fusions involving NAB2-STAT6 in malignant meningioma. *C. Ling[1], F. Gao[1], L. Shi[1], G. Zada[1,2], WJ. Mack[1,2], K. Wang[1,3,4].* 1) Zilkha Neurogenetic Institute, Keck School of Medicine,University of Southern California, Los Angeles, USA; 2) Department of Neurosurgery, Keck School of Medicine, University of Southern California, Los Angeles, USA; 3) Department of Psychiatry, Keck School of Medicine, University of Southern California, Los Angeles, USA; 4) Division of Bioinformatics, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, USA.

Meningiomas are the most common primary intracranial tumors, with ~3% meeting criteria for anaplastic/malignant (WHO grade 3) classification. Despite previous gene expression studies using microarray techniques, gene fusion events in meningiomas have not been reported. In this study, we explored transcriptional gene fusion in three malignant and four benign meningiomas, using the high-throughput RNA sequencing (RNA-Seq) technique. Candidate gene fusions were detected using both TopHat-Fusion and Chimerascan. In one malignant meningioma, we identified an inversion-mediated fusion between two adjacent genes (NAB2 and STAT6) on chromosome 12q13.3, which created two novel in-frame transcripts. Sanger sequencing on both cDNA and genomic DNA was used to validate fusion transcripts and map the exact chromosomal inversion junctions. The fusion of NAB2-STAT6 was localized between the end of exon 6 of NAB2 and the beginning of exon 18 of STAT6, with 98 junction reads identified from RNA-Seq data. The fusion of STAT-NAB2 was supported by 126 junction reads, which connects the last nucleotide of exon17 of STAT6 with the first nucleotide of exon7 of NAB2. Interestingly, the fused exons of STAT6 encoded the transcription activation domain of STAT6, and gene expression of NAB1, NAB2, TGFB1 and CASP9 was up-regulated in the malignant subject compared to other subjects, suggesting a functional role of this gene fusion product in activating transcription of these genes. On the other hand, gene fusion of STAT6-NAB2 is likely to produce a protein product incapable of transcription activation. Expression of the STAT6 downstream target gene BCL2L1 is down-regulated, implying a potential functional effect of STAT6-NAB2 fusion. The NAB2-STAT6 fusion has been recently reported as a recurrent fusion event in solitary fibrous tumors. Thus, the gene fusions detected in our study may play a role in malignant transformation of meningiomas and further investigation with more samples is necessary to understand the cause of genome instability observed.

3327W

Copy number gains on chromosome 21 contribute to risk of acute lymphoblastic leukemia in non-Down syndrome children. *A.J. de Smith[1], A.M. Yiorkas[2], R.G. Walters[3], K.M. Walsh[1,4], C. Metayer[5], L.F. Barcellos[5], P.A. Buffler[5], A.I. Blakemore[2], A.P. Chokkalingam[5], J.L. Wiemels[1].* 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Section of Investigative Medicine, Imperial College London, London, United Kingdom; 3) Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, United Kingdom; 4) Division of Neuroepidemiology, Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94143, USA; 5) School of Public Health, University of California Berkeley, Berkeley, CA 94704, USA.

The causes of childhood acute lymphoblastic leukemia (ALL), the most prevalent cancer in children, are largely unknown. Children with Down syndrome have an approximately 20-fold increased risk of developing ALL, suggesting a role for genes on chromosome 21 (Chr21). In this study, we investigated germline copy number variation on Chr21, with the hypothesis that copy number gains in non-Down syndrome individuals may be associated with increased ALL risk. Genome-wide SNP data (Illumina Human OmniExpress array) were available for 305 Hispanic children with ALL and 451 controls matched on age, gender, ethnicity and maternal race, from the California Childhood Leukemia Study (CCLS). Copy number variants (CNVs) were predicted using PennCNV, with default parameters. For quality control, samples with excess log R ratio variance and GC waviness, and CNVs with a PennCNV confidence score <20 were excluded. Cases with Down syndrome were also excluded. Association analysis was carried out in PLINK, with minimum CNV segment length at the default 20kb. We assessed the burden of large copy number gains according to the average number per sample, with a cut-off of 100kb for large and 500kb for very large gains. Case-control association analysis of specific Chr21 copy number gains was also carried out. A total of 1264 copy number gains were detected on Chr21 across all subjects. There was a highly significantly increased rate of large copy number gains (>100kb) in cases (0.744) compared to controls (0.588), with $p = 5.5 \times 10^{-6}$ (Fisher's exact test, one-tailed). The rate of very large copy number gains (>500kb) was also significantly higher in cases (0.049) than controls (0.007) ($p = 2.1 \times 10^{-4}$). A common copy number gain at the *SLC19A1* locus was nominally associated with increased ALL risk, with 75.1% cases carrying this CNV compared to 67.6% controls (odds ratio = 1.44, 95% CI: 1.04-2.00, p = 0.028). *SLC19A1* is involved in regulation of intracellular levels of folate, which is required for DNA repair and methylation, and previous studies suggest folate pathway genes may play a role in childhood ALL susceptibility. Results of this study suggest that large copy number gains on Chr21 contribute to risk of ALL in children without Down syndrome. Identifying Chr21 genes associated with childhood ALL should provide insight into the etiology of this disease, and may help to explain the increased risk of ALL in children with trisomy 21.

3328T

A Case of Hyperdiploidy in CLL/SLL: A Rare Cytogenetic Event Associated with Poor Prognosis. *M. DeNicola, S. Pullakat, S. Yea, P. Delgado, L. Yang, N. Rao, C.A. Tirado.* Pathology & Laboratory Medicine, UCLA, Los Angeles, CA., USA.

Hyperdiploidy has been described in a variety of malignancies including acute lymphoblastic leukemia and plasma cell myeloma, in which the abnormality is associated with a very good prognosis. Herein we describe a 61 year-old female who presented with an eleven month history of enlarging lymph nodes in her neck. A CT scan revealed bulky confluent lymphadenopathy in the left neck involving stations 2-5 with extension into the superior left mediastinum, enlarged lymph nodes in the right neck involving stations 2B and 5, and a large soft tissue density overlying the left occipital bone measuring 9 cm in maximum dimension. She denied fevers, chills, or night sweats, but did endorse mild fatigue. Her WBC count was 4,900/uL with a differential of 55% neutrophils, 33.7% lymphocytes, and 9.2% monocytes. The hemoglobin was 8.5gm/dL, platelets 227,000/mcL, ESR 38, and LDH 212. SPEP showed two IgM kappa monoclonal proteins. A core biopsy of the left neck mass showed infiltration of the lymph node by atypical lymphoid cells with slightly irregular nuclei, condensed chromatin, and medium amounts of cytoplasm. The atypical lymphoid cells expressed CD19, CD20, PAX5, CD5, BCL2, and IgM with a Ki67 of 10%; the cells were negative for CD10, CD23, BCL1, BCL6, and IgD. Flow cytometry revealed a monotypic B cell population with kappa light chain restriction and co-expression of CD20 and CD5. Molecular pathology studies were positive for clonal B-cell gene rearrangement. While consideration was given to nodal marginal zone lymphoma, a final diagnosis of atypical chronic lymphocytic lymphoma (CLL) was made. The initial chromosome analysis showed an abnormal karyotype described as 46-48,XX,add(3)(q12),+16,+mar[cp3]/46,XX[1]. A second karyotype a week later showed a hyperdiploid clone described as : 46,X,-X,-3,-6,+7,+9,-14,-15,+16,+17,+17,+20,-22[1]/46,XX[19]. Concurrent FISH studies using the CLL FISH panel showed with an extra copy of chomosome13 and an extra copy of the short arm of chromosome 17. FISH for t(11;14) was negative. These results suggest the presence of an underlying complex hyperdiploid karyotype. Hyperdiploidy is a rare event in SLL/CLL and it is usually associated with a poor prognosis.

3329F

The cytogenetic landscape and downstream consequences in 16,172 cancer samples. *L. Franke[1], R.S.N. Fehrmann[2], J. Karjalainen[1], L. de Vries[2], M. van Vugt[2], C. Wijmenga[1].* 1) Department of Genetics, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands; 2) Department of Oncology, University Medical Centre Groningen, University of Groningen, Groningen, Netherlands.

With the growing public availability of gene expression data, integration and re-analysis of thousands of samples has the potential to discover new patterns that yield important novel biological insight. We integrated gene expression profiles from 77,840 samples (human, mouse and rat) to identify 2,206 'transcriptional components' (TCs) that explain the major regulators of gene expression. These TCs in combination with a 'guilt-by-association' approach allowed us to predict biological functions of individual genes, identifying and experimentally proving FEN1 as an essential homologous recombination repair gene. Many of these TCs describe metabolic, physiological and cell-type specific differences. Surprisingly, upon correcting expression data of 16,172 primary tumor samples for these TCs, we oberved that the residual expression signals accurately reflected somatic copy number aberrations (SCNAs), thereby revealing the cytogenetic landscape in these tumors. A genome-wide association study at an unprecedented large scale was performed that resulted in a robust genomic signature that associated individual SCNAs to genomic instability. In addition, analysis of this cytogenetic landscape revealed clear commonalities and differences between different types of cancer. Subsequent expression quantitative trait locus (eQTL) analysis enabled us to link SCNAs to their downstream effects on biological pathways. In summary, our method provides a tool to re-analyze and re-interpret the vast amount of expression data currently available to explore cytogenetic alterations in relation to phenotypic data and to discover new biological functions of genes.

3330W

Genomic copy number signatures uncovered genetically distinct group which is different from Adenocarcinoma and Squamous cell carcinoma of the lung cancer. *E. LEE[1], J. Lee[2], H. Kim[1], B. Shin[1], J. Kim[1], A. Kim[1], B. Kim[1].* 1) Korea university Guro hospital, Seoul, South Korea; 2) Korea university College of medicine.

Adenocarcinoma (AC) and squamous cell carcinoma (SCC) consisting of the most of non-small cell lung carcinoma (NSCLC) have different clinical presentation, morphology, treatment and prognosis. Recent studies suggested that fundamental genetic alterations related to carcinogenesis of each tumor type may be different. In this study, we evaluated the difference of genetic alteration of ACs and SCCs that may be applied for diagnosis of NSCLC. Genomic copy number alterations (CNAS) of 50 primary NSCLC samples (25 ACs and 25 SCCs) as well as paired normal tissue were investigated by using array comparative genomic hybridization (CGH). Common CNAs were evaluated in each subtype and compared each other to establish CNA signature. Total CNAs were more frequently identified in SCCs than ACs. The most common CNAs were gain of 3q in SCCs and 7q in ACs. 44 genes from SCCs and 28 genes from ACs were located in the commonly (> 68%) gained and lost region. The results of gene clustering revealed three clustered groups - Group I: AC, Group II: SCC, and Group III: Genetically distinct another group consisting AC and SCC. The present study characterized genetic differences of the subtype of NSCLC including AC and SCC, which grouped genetically unique and distinct three groups by gene clustering including group III, which has different genetic signature from AC and SCC, but similar histopathology with AC or SCC. The further study of significance of this finding and the possibility of diagnostic application of this genetic marker is necessary.

3331T

The link between the losses of TP53 and ETV6 and del(5q) and/or -7/del(7q) in the progression of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). *J. Lee[1], R. Zhang[2], Y. Kim[3], X. Wang[3], Y. Li[2], S. Li[3].* 1) Pathology, Korea University, Seoul, South Korea; 2) Department of Hematology, China Medical School, Sheyang, Liaoning, P.R. China; 3) Department of Pediatrics, The University of Oklahoma health Sceinces Center, OK, USA.

-5/del(5q) and -7/del(7q) are the most common chromosomal abnormalities in myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) and are often accompanied each other with poor prognosis. To evaluate relationship between -5/del(5q) and -7/del(7q) in the different background of acquired genomic alteration, integrated analyses of 26 cases with MDS/AML carrying del(5) and/or -7/del(7) were performed by using karyotype, FISH and array CGH and characterized by grouping; Group I: cases only with del(5q) (6 cases), group II: cases only with -7/del(7q) (11 cases) and group III: cases with concurrent del(5q) and del(7q) (9 cases). The overlapped common deleted region (CDR) of chromosome 5 from group I and III was 5q31.1-33.1 (130,562,020-150,625,216 bp: hg18) with ~29 Mb in size and chromosome 7 from group II and III was 7q31.31-q36.1 (119,547,309-149,033,790 bp: hg 18) with ~29 Mb in size. Total 249 other copy number alterations (CNAs) from the genome except chromosomes 5 and 7, which were ~ 78.3 % of total CNAs (318), were observed. The group III [concurrent del(5q) and -7/del(7q)] is a distinctive entity carrying the most numerous other CNAs, recurrent other CNAs, cryptic CNAs (<5Mb) and complex CNAs. Gains of PDGFD, MLL, ETS1 and ADAMTS8 and/or losses of MANF, TP53 and WRAP53 were highly associated with del(5q) since they are shared in both groups I and III. Whereas loss of ETV6 shown to be specially associated with concurrent del(5q) and -7/del(7q) (group III). Theses CNAs or genes associated with particular group or particular chromosomal abnormality may play secondary role of group specific disease progression and should be further evaluated for their clinical significance and influence on therapeutic approaches in the patient with MDS/AML carrying del(5q) and/or -7/del(7q).

3332F

Targeted genomic DNA and cDNA next generation sequencing identifies a high frequency of kinase gene fusions in Spitz tumors Involving *ROS1*, *ALK*, *RET*, *NTRK1*, and *BRAF*. *K. Wang[1], J. He[1], T. Wiesner[2], G. Otto[1], R. Yelensky[1], MF. Berger[2], R. Esteve-Puig[3], D. Lipson[1], K. Brennan[1], VA. Miller[1], MT. Cronin[1], BC. Bastian[3], PJ. Stephens[1].* 1) Foundation Medicine, Cambridge, MA; 2) Memorial Sloan-Kettering Cancer Center, New York; 3) University of California San Francisco, San Francisco.

Background: Gene fusions that generate chimeric oncoproteins are becoming increasingly important diagnostic markers and drug targets in both hematologic malignancies and solid tumors. However, the development of high throughput methodologies is required to systematically interrogate cancer genomes to characterize gene fusions. We applied targeted genomic DNA and cDNA sequencing to a series of Spitz tumors, a family of melanocytic neoplasms in which biologic behavior which can range from an indolent behavior to widespread metastatic disease, to gain insights into the genetic underpinnings of the poorly understood disease. **Methods:** Genomic DNA and total RNA was isolated from formalin fixed paraffin embedded (FFPE) tissue samples from 20 benign Spitz nevi and 8 atypical Spitz tumors (with morphological features inconsistent with *HRAS* or *BRAF/BAP1* mutation). DNA sequencing was performed for 3,230 exons of 182 cancer-related genes plus 37 introns of 14 genes commonly fused on indexed hybridization-captured libraries to an average unique coverage of 997x, with 99.96% of exons being sequenced at ≥100x coverage. RNA sequencing was performed on indexed libraries captured using the cDNA Kinome hybridization kit (Agilent) generating >50,000,000 unique pairs per specimen. **Results:** Genomic rearrangements were observed in 19/28 (68%) Spitz tumors with only a single known somatic point mutation in *HRAS* detected. The rearrangements fused the intact tyrosine kinase domains of *ROS1*(36%), *ALK*(14%), *RET*(7%), *NTRK1*(7%), and *BRAF*(4%) to a wide range of predominantly novel 5'partners including *PWWP2A, PPFIBP1, ERC1, MYO5A, CLIP1, HLA-A, ZCCHC8, DCNT1, LMNA* and *CEP89*. These gene rearrangements formed constitutively activated chimeric oncogenes that activated oncogenic signaling pathways. All fusions occurred in a mutually exclusive pattern and were more common in younger patients compared to patients whose tumors did not harbor fusions (median age 14 versus 24 years, p=0.02). **Conclusions:** Targeted genomic DNA and cDNA sequencing can be applied to cancer genomes to identify both known and novel gene fusions which may be useful as diagnostic markers and therapeutic targets but thorough analytic validation to determine accuracy is required prior to routine use in the clinic.

3333W

A rare case of de novo AML with basophilic differentiation and abnormal cytogenetics with tetraploidy and structural abnormalities in a 50-year-old male. *A. Yenamandra[1], F.C. Wheeler[1], D. Zalepa[1], A. Gardner[1], M. Guanchez-Lopez[1], M.A. Thompson[1], J. Douds[1], S. Mohan[2], A. Shaver[1].* 1) Dept. of Pathology, Immunology and Microbiology, Vanderbilt University Medical Center, Nashville, TN 37203; 2) Div. of Hematology and Oncology, Dept. of Medicine, Vanderbilt University Medical Center, Nashville, TN 37203.

Massive hyperdiploid and tetraploid karyotype with structural abnormalities has been reported to be rare in de novo AML and MDS especially in male patients. These patients have been reported to have a low remission rate and short survival and define a prognostically unfavorable cytogenetic group in de novo AML (Iyer, RV et al., Cancer Genet and Cytogenet. 1; 148(1):29-34, 2004 and Bene, MC et al., Leukemia 20-725, 2006). We report a 50-year-old male with a history of diabetes who presented with pancytopenia, neutropenia, and worsening hypoxemia, ultimately requiring intubation and transfer to the MICU in March 2013. He was diagnosed with AML on the basis of increased blasts in his peripheral blood smear (sample collected 03/09/13). His bone marrow biopsy (sample collected 03/11/13) was markedly hypercellular with 80-90% cellularity and was involved by AML by morphology (48.5% blasts) and immunophenotype. The blasts were remarkable for large, basophilic granules, and a range of immature basophils were seen in the background. Blasts had an immunophenotype suggestive of basophilic differentiation (CD13+, CD33 moderate, CD34+, CD117-, CD123+). Molecular genetic studies were negative for KIT Asp816Val mutation, FLT3 ITD, or mutations in CEBPA or NPM1. Both peripheral blood and bone marrow samples revealed an abnormal karyotype with tetraploid chromosomal complement and structural anomalies of chromosome 1 and 12. His karyotype was 92,XXYY,dup(1)(q21q32)x2,del(12)(q24.1)[10]/46,XY[10]. Fluorescence in situ hybridization (FISH) for AML and MDS (5,7,8,20,t(8;21), t(15;17), inv(16) and MLL) probes on the bone marrow sample (03/11/13) revealed 4 copies for all the target regions tested. His bone marrow on 03/27/13 on day 14 of induction with HiDAC and etoposide revealed residual acute myeloid leukemia, with persistent blasts that demonstrated more prominent basophilic differentiation and increased background basophilia, as well as persistence of the abnormal tetraploid clone. There was a decrease in the percentage of cells with the abnormal karyotype, as well as a decreased percentage of blasts, compared to the previous bone marrow biopsy. Reinduction chemotherapy was administered and most recent follow up bone marrows were negative for leukemia by morphology and cytogenetics. Acute basophilic leukemia is an extremely rare subset of AML, and tetraploid karyotype is a rare cytogenetic finding; the combination of these two findings is unique in the literature.

3334T

Complex karyotype in a patient with AML-M1. *L. Bobadilla-Morales[1,2], C. Ortega-de la Torre[1,2], HJ. Pimentel-Gutiérrez[1], C. Barba-Barba[2], M.C. Almodóvar-Cuevas[2], G. Serafín-Saucedo[2], G. Macías-Salcedo[2], M.E. Velázquez-RiverA[2], A. Márquez-Mora[2], S. Gallegos-Castorena[3], F. Sánchez-Zubieta[3], A. Corona-Rivera[1,2].* 1) Laboratorio de Citogenética Genotoxicidad y Biomonitoreo, Instituto de Genética Humana 'Dr. Enrique Corona Rivera' Biol Molecular Genomica, Univ Guadalajara, CUCS, Guadalajara, Jalisco, Mexico; 2) Unidad de Citogenética, Servicio de Hematología Oncología Pediátrica, División de Pediatría, Nuevo Hospital Civil "Dr. Juan I. Menchaca', Guadalajara, Jalisco, México; 3) Servicio de Hematología Oncología Pediátrica, División de Pediatría, Nuevo Hospital Civil "Dr. Juan I. Menchaca', Guadalajara, Jalisco, México.

Purpose: We present the cytogenetic findings in a patient with acute myeloblastic leukemia-M1 (AML-M1) at diagnosis. Clinic Presentation: We present a case of a 1 year old male, who presented a tumor in the right side of the neck of approximately 8cm by 10cm size; it was multilobed, without pain and no erythema or heat. In the left side he had nodes smaller than 1 cm size. In inguinal area there was multiple nodal growth of 1 to 3cm diameter. TAC study reported neck solid lesion on the right side. Bone marrow had decreased cellularity, no megakaryocytes, 92% blasts, which were large with abundant chromatin, visible nucleoli and cytoplasm with abundant inclusions. The diagnostic was AML-M1. Method: In cytogenetic studies were performed cultures with RPMI1640 and additives, and marrow max culture media. Conventional cytogenetic study in bone marrow showed a trisomy 8, a rearrangement involving chromosomes 10 and 12, and a marker chromosome that seemed that had chromosome 11 material. For that reason Fluorescent in situ hybridization (FISH) was performed for MLL (Vysis LSI MLL Dual Color, Break Apart Rearrangement Probe), AML1/ETO (Vysis RUNX1/RUNX1T1 DF FISH Probe Kit), ETV6/RUNX1T1 (Vysis LSI ETV6(-TEL)/RUNX1(AML1) ES Dual Color Translocation Probe Set), and finally performed whole chromosome painting (WCP) of 11 and 12 chromosomes (Kreatech). Results: The preliminary karyotype result was: 47,XY, +8,t(10;12)(p11.2;21),-11,+mar. The FISH results were disruption of MLL gene (nuc ish(MLLx2)(5'MLL sep 3'MLLx1)[207/300]) and an extra signal of ETO gene (nuc ish(ETOx3,AML1x2)[59/200]). And Finally the WCP results were: ish t(10;12)(p11.2;q21)(wcp12+),inv(11)(pter->p11.2::q23->p11.2::qter)[cp30]. Conclusions: With all previous data we concluded that there was a trisomy 8, a pericentric inversion of chromosome 11 involving MLL gene and a translocation between chromosomes 10 and 12. All this showed the final result as 47,XY,+8,t(10;12)(p11.2;21), inv(11)(pter->p11.2::q23->p11.2::qter).ish t(10;12)(p11.2;q21)(wcp12+),;),inv(11)(pter->p11.2::q23->p11.2::qter)(wcp 11+)[cp30]. This results stress the importance of an integrated analysis between different methods especially in complex karyotypes. Unfortunately the patient died after influenza disease, and it was impossible to make an association between prognosis and karyotype.

3335F

Single chromosome contribution to karyotypic evolution in tumor formation and metastases. *J. Brown[1,2], B. Flynn[1], C. Obergfell[1], M. Longo[1], M. O'Neill[1], R. O'Neill[1], Peromyscus Genetic Stock Center.* 1) Molecular and Cell Biology, University of CT, Storrs, CT; 2) Allied Health Sciences, University of CT, Storrs, CT.

Genetic lesions that contribute to the progression of cancer from a primary tumor to metastases often involve a complex process of chromosome evolution, either progressive or catastrophic; however, the primary factor(s) that precede complex rearrangements remain largely unknown. As a model to investigate the mechanisms that drive metastatic potential, an inbred line of *Peromyscus leucopus* was isolated that carries a high frequency of malignant Harderian gland tumor formation, which immediately and predictably metastasizes to the lymphatic system, lungs and liver. Employing fluorescence *in situ* hybridization we have identified a recurrent abnormality, intrachromosomal amplification of Chromosome 5, among three individuals. Deep-sequencing of the microdissected Chromosome 5 has been employed to identify a candidate for the initiation of the primary lesion that contributes to both tumor susceptibility and karyotypic evolution in the metastases. Our *Peromyscus* strain is a highly suitable *a priori* model system in which to further study the mechanistic link between the derivation of distinct chromosomal aberrations in a primary tumor and those in secondary metastases.

3336W

The Impact of Microarray in Diagnosing Pediatric Acute Lymphocytic Leukemia. *C.M. Higgins, D.L. Pickering, M.L. Wiggins, W.G. Sanger, B.J. Dave.* Human Genetics Laboratory, Munroe Meyer Institute, University of Nebraska Medical Center, Omaha, NE.

Acute lymphocytic leukemia (ALL) is the most common form of childhood cancer representing 23% of cancer diagnoses in children. With current therapeutic regimens, long-term survival can be expected in approximately 80% of pediatric ALL cases; however, up to 20% will remain treatment refractory. Cytogenetics and FISH have been the gold standard for defining genetic abnormalities and facilitating therapeutic stratification of pediatric ALL cases. More recently, microarray studies, which provide higher resolution and genome wide analyses, have identified novel abnormalities that are significant and recurrent with prognostic value; however, the test is yet to be included in the routine clinical diagnostic armamentarium. To better refine the diagnosis and improve prognosis/risk stratification we included microarray in our pilot studies utilizing all three techniques in six newly diagnosed pediatric B-ALL cases. We found abnormalities in all six cases with all three methodologies: cytogenetics, FISH and microarray. Ploidy changes were noted by all three techniques; known ALL-associated and prognostically significant translocations were observed by cytogenetics and FISH; the subtle translocation involving *TEL/AML* (*ETV6/RUNX1*) was detected only by FISH (2 cases); and microarray identified multiple genomic alterations that were either too subtle to be determined by cytogenetics and/or were not included in the standard ALL-FISH panel. Abnormalities determined by microarray alone included *IGH* deletions in 66% (4 cases), *IKZF1* deletions in 50% (3 cases), deletions in *EVI1* in 33% (2 cases), and a deletion of *TP53* in one case. There were overlaps in detection of abnormalities between FISH, cytogenetics and microarray studies. A translocation involving the *EVI1* gene region was determined by cytogenetics but could not be confirmed by FISH. Only microarray detected the microdeletion within the *EVI1* region. One hyperdiploid case exhibited a deletion of *TP53* which was deciphered by microarray and confirmed by FISH. These abnormalities that are known to confer poor prognosis, including the *IKZF1* deletions, would not have been apparent without using a combination of all three techniques. Our studies suggest that consistent use of combined testing may play a critical role in the state of the art diagnosis, and subsequently assist in therapeutic stratification to determine treatment regimens that are most appropriate for the disease state of these young patients.

3337T

Molecular characterization of der(1)t(1;19) in a patient with myelodysplastic syndrome. *J. Kim[1], J. Choi[2], J. Choi[3], Y. Kim[3], K. Lee[3,4].* 1) Lab Med, Yonsei University Wonju College of Medicine, Wonju, South Korea; 2) Pathology, Yonsei University Wonju College of Medicine, Wonju, South Korea; 3) Lab Med, Yonsei University College of Medicine, Seoul, South Korea; 4) Corresponding.

Translocation (1;19)(q23;p13), creating a fusion between *E2A* and *PBX1* genes, is a chromosomal rearrangement associated with childhood pre-B cell acute lymphoblastic leukemia (ALL), which is detected in about 5% of pediatric patients diagnosed with ALL. In contrast, rare case of der(1)t(1;19)(p13;p13.1) has been reported in mostly myeloid neoplasia or myelodysplastic syndrome (MDS). Only about 9 cases of such translocation have been reported in the literature and the specific break point has not been studied. Here, we report a case of t(1;19)(p13;p13.1) in a MDS patient and the characterization of the breakpoint by using conventional karyotyping, multicolor FISH and SNP array. The karyotyping reavealed 47,XX, +der(1)t(1;19)(p13;p13.1) in 20 metaphase cells, and involvement of chromosome 1 and 19 and the breakpoint was confirmed by multicolor FISH. In order to refine the breakpoint region, Affymetrix Cytoscan 750K array was carried out, and gain in the region of 1q21.2q44 (143,932,349-248,660,805) and 19p12p13.3 (260,911-24,177,726) was found, indicating a whole arm translocation. Because it is a rare chromosomal rearrangement but found recurrently in a subset of myeloid neoplasms, the possible role of such unbalanced whole arm translocation in the leukemogenesis and the possibility of gene involved in such chromosomal rearrangement need to be determined.

3338F
Characterization of Uterine Leiomyomas by Whole-genome Sequencing. *M. Mehine[1], E. Kaasinen[1], N. Mäkinen[1], R. Katainen[1], K. Kämpjärvi[1], E. Pitkänen[1], H. Heinonen[1], R. Bützow[2], O. Kilpivaara[1], A. Kousmanen[1], H. Ristolainen[1], M. Gentile[4], J. Sjöberg[3], P. Vahteristo[1], L. Aaltonen[1].* 1) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Department of Pathology, Haartman Institute, University of Helsinki, Helsinki, Finland; 3) Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland; 4) CSC-IT Center for Science, Espoo, Finland.

Uterine leiomyomas are benign but impact the health of millions of women. We performed whole genome sequencing and expression profiling of 38 leiomyomas and corresponding myometrium from 30 patients. We detected complex chromosomal rearrangements (CCRs) resembling chromothripsis in our leiomyoma samples. CCRs appears to be an unexpectedly frequent phenomenon in myometrium and had occurred in the presence of normal TP53 alleles. These rearrangements are best explained by a single event of multiple chromosomal breaks and random reassembly, and had created tissue-specific driver changes such as rearrangements between HMGA2 and RAD51B loci, aberrations at the COL4A5/COL4A6 locus and interstitial deletions on chromosome 7. In some cases we found that CCRs had occurred more than once in single tumor cell lineages. We also found multiple separate tumor nodules to be clonally related. This could in part explain the common occurrence of synchronous lesions in affected uteri. Further work should examine whether the mechanisms underlying CCRs are identical to those causing chromothripsis.

3339W
Fluorescence in situ hybridization (FISH) on abnormal metaphases helps unmask the location of hidden genes/genomic segments in cases with abnormal karyotypes and abnormal FISH panels. *N. Mitter.* Cytogenetics and FISH Labs, Dianon Pathology (Lab Corp), Shelton, CT.

Some complex rearrangements result in an ‘apparent’ loss of some chromosome segment or a chromosome rearrangement not detectable by classic cytogenetics analysis; but a limited FISH analysis shows the presence of two or more copies of the apparently missing chromosome segment or the splitting of a gene at a breakpoint involved in a ‘masked’ chromosome rearrangement. Microarray analysis can confirm the copy number variations for all genomic regions, but it cannot point the location of these masked/ apparently missing regions. In such cases, a FISH analysis with appropriate probes on the cytogenetically abnormal looking metaphases comes to the rescue. We describe here two such cases, where in each case it was cytogenetically a real puzzle compared to the FISH panel studies, but metaphase FISH analyses solved these puzzles. In the first case, fifteen of the twenty cells examined had only one copy of a normal chromosome 5 but also a copy of chromosome 14 with apparent genetic material from the long arm of chromosome 5 attached to its short arm, and presence of extra genetic material of indeterminate origin on a few other chromosomes. FISH results for MDS panel, however, revealed not only a loss of EGR1 gene, but also two normal signals for genes on 5p (D5S23 and D5S721). The cytogenetic analysis by itself had failed to identify the location of the second 5p. Follow-up FISH analysis on the karyotyped metaphases revealed the presence of chromosome 5 short arm material (band 5p15.2) inserted in the long arm region of chromosome 5 with a deletion of EGR1 gene at band 5q31. This implies a two-break event for chromosome 5 (most likely a pericentric inversion followed by an unbalanced translocation to the short arm of chromosome 14, resulting in presence of two copies of the probed short arm region, as well as del(5q). This analysis, therefore, also helped in understanding the series of events that led to the unique FISH panel results. In second case, there was only one ‘normal’ copy of chromosome 3, along with presence of a long segment of extra genetic material on chromosome 11, besides other abnormalities. Add(11q) material had a complex banding pattern. FISH analysis for BCL6 gene revealed presence of three copies of the gene but the location of two copies was not apparent. Follow-up metaphase FISH analysis revealed presence of two copies of BCL6 gene next to each other on add(11q) material, indicating a duplication of 3q segment.

3340T
SNP Chromosomal Microarray Analysis with DNA extracted from Carnoy's fixed cell pellets of Myelodysplastic Syndrome (MDS)/Acute Myeloid leukemia (AML). *N. Rao, D. Cherukuri, I. Amarillo, F. Quintero-Rivera.* Department of Pathology & Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA.

SNP chromosomal microarray analysis (SNP-CMA) is an important tool in detecting copy number changes and loss of heterozygosity (LOH) in patients with inherited genetic disorders and in various cancers. The common protocol for SNP-CMA testing requires DNA extraction from fresh tissue samples. Carnoy’s fixed cell pellets are typically available after routine karyotype and FISH studies for any additional studies. In this pilot study, we wanted to determine the concordant/ discordant results from karyotype/ FISH and with SNP-CMA obtained using the Affymetrix Cytoscan® HD and Agilent CytoGenomics array platforms with DNA extracted from Carnoy’s fixed cells. For this study twelve specimens from patients diagnosed with MDS or secondary AML and with abnormal karyotype and/or FISH results, were selected. Our observations indicate that SNP-CMA is superior to karyotype/FISH in detecting LOH regions. However, routine FISH/karyotype is more sensitive than SNP-CMA for the detection of small clones, such as in cases with 2 abnormal metaphases or with 8% abnormal interphases, both of which are clinically relevant for monitoring of residual/recurrence disease or its progression. Copy number changes were accurately detected with either of the array platforms when the chromosomal abnormalities are present in 30% of the cells analyzed. SNP-CMA analysis detected additional chromosomal aberrations in 4 of the 12 cases. However, these aberrations did not change the cytogenetics prognostic score. In the second part of our study, we performed SNP-CMA in 8 MDS cases (< 19% blast) but with normal karyotype/ FISH results. There were no cases with copy-neutral loss of heterozygosity (CN-LOH) of >5 Mb, but one showed a pathogenic deletion. Our data suggest that DNA extracted from Carnoy’s fixed pellets is a reliable source for identification of CNVs and CN-LOH by SNP-CMA. The diagnostic yield in MDS patients with normal karyotype/FISH is increased by 12.5% with SNP-CMA testing. In conclusion, our study shows the utility of SNP-CMA with DNA extracted from Carnoy’s fixed pellets as a reflex test in patients diagnosed with MDS with a normal karyotype/FISH. This is especially useful in a clinical setting where it can prevent the need for a repeat or fresh bone marrow biopsy and the patient’s re-evaluation. Such studies will also aid in the identification of additional genomic aberrations that can have an impact on the molecular sub-classification, treatment or prognosis.

3341F
Genomic aberrations in myeloid sarcoma (MS): characterization of formalin-fixed paraffin-embedded (FFPE) samples by whole-genome SNP arrays and next-generation sequencing. *M. Sukhanova[1], K.M. Mirza[2], Z. Li[1], F. Stolzel[4], K. Onel[3], M.M. Sasaki[3], P. Reddy[2], L. Joseph[2], G. Raca[1].* 1) Department of Medicine, The University of Chicago Medical Center, Chicago, IL; 2) Department of Pathology, The University of Chicago Medical Center, Chicago, IL; 3) Department of Pediatrics, The University of Chicago Medical Center, Chicago, IL; 4) Medizinische Klinik und Poliklinik I, Universitatsklinikum Carl Gustav Carus, Dresden, Germany.

Genetic abnormalities in tumor cells are the strongest prognostic indicator for response to therapy and survival in patients with acute myeloid leukemia (AML). However, when AML presents as a tumor of malignant granulocytic precursor cells localized outside of the bone marrow (so-called myeloid sarcoma), it initially does not get recognized as a hematologic malignancy, and fresh tumor samples do not get obtained for cytogenetic and molecular testing. FFPE tissue is often the only available specimen type for analysis of MS. The purpose of this study was to determine whether the analysis of FFPE samples using whole-genome SNP arrays and next-generation sequencing (NGS) allows detection of genetic abnormalities of prognostic significance in MS. We analyzed five cases of MS without BM involvement, where cytogenetic and molecular studies were either not available, or did not identify prognostically significant genetic abnormalities. Array testing was performed using CytoScan HD array (Affymetrix Inc., Santa Clara, CA). Targeted resequencing of AML associated genes was performed on the Personal Genome Machine (PGM) from Ion Torrent (Life Technologies, Carlsbad, CA), with two custom primer sets (AML and MDS panel) designed to amplify coding regions of 19 genes associated with myeloid malignancies. Somatic, disease-associated genetic abnormalities were identified in all tested samples. Additionally, aberrations of clear prognostic significance were observed in four out of five cases. In two cases array analysis detected multiple unrelated copy number abnormalities and LOH events; the presence of numerous aberrations is analogous to a complex karyotype, which is associated with a very poor prognosis in AML. Additional two cases showed an NPM1 mutation without a concomitant FLT3-ITD, a combination associated with a favorable prognosis. In conclusion, array analysis and NGS using FFPE samples revealed previously unidentified disease-associated genetic abnormalities in all five samples of MS, and in four cases provided information that would have been of major prognostic importance if available at the time of diagnosis. Whole-genome arrays and NGS are becoming increasingly available in the clinical setting, and should be used for detection of prognostic genomic aberrations in MS in the absence of fresh material for cytogenetic studies.

3342W

A case of pediatric acute lymphoblastic leukemia presenting with a (9;12) translocation involving *JAK2* and rearrangement of *MLL* at 11q23 with an apparent insertion at 6q27. *C. Tirado, M. DeNicola, D. Rao, P. Delgado, A. Jarrin, L. Yang, N. Rao.* Pathology & laboratory Medicine, UCLA, Los Angeles, CA.

Acute lymphoblastic leukemia (B-ALL) is the most common malignancy in pediatric patients and the leading cause of cancer-related death in children and young adults. Herein we present a 13-year-old boy who presented with abdominal pain for three months and transient fevers over the last three weeks. On presentation, he was found to have leukocytosis (WBC 76.5 × 10^3/uL), anemia (hemoglobin 5.3 g/dL), and thrombocytopenia (platelet count 15.3 × 103/uL). A differential count showed 94% blasts which expressed CD34, bright CD10, CD19, partial CD20, partial CD38, partial TdT, CD79a, and HLA-DR by flow cytometry. A bone marrow biopsy was performed, and showed a hypercellular marrow extensively involved (~95%) by sheets of lymphoblasts with irregularly shaped nuclei, immature chromatin, and occasional cells with vacuolated cytoplasm. The immunophenotype of the blasts was the same as that found in the peripheral blood. A diagnosis of B-lymphoblastic leukemia was rendered. Chromosome analysis of the bone marrow showed 5/20 cells examined with an MLL insertion on 6q27 and a balanced translocation between 9p24 and 12p11.2. This findings were confirmed by metaphase FISH. The karyotype was described as : 46,XY,ins(6;11)(q27;q23q23),t(9;12)(p24;p11.2)[5]/46,XY[15]. FISH analysis using interphase nuclei also showed MLL rearrangement: *MLL* (11q23) split signals in 23.6% (71/300) of the nuclei examined, suggestive of an *MLL* (11q23) gene rearrangement . Additionally, deletion of the 5' *IGH@* region, corresponding to the variable segment in 88.3% (265/300) of the nuclei analyzed, was observed, suggestive of a deletion or an unbalanced rearrangement involving chromosome 14q32. These findings were described as: nuc ish(MLLx2)(5'MLL sep 3'MLLx1)[71/300] nuc ish(3'IGH@x2,5'IGHx1)(3'IGH@ con 5'IGH@x1)[265/300] . Rearrangements of 9p24 usually involving *JAK2* (confirmed by FISH) induce dimerization or oligomerization of *JAK2* without ligand binding, resulting in constitutive activation of *JAK2*, and are seen in B- Acute lymphoblastic leukemia (B-ALL) [1]. Rearrangements of the short arm of chromosome 12 are also seen in B-ALL. Correlation with other clinical and hematological data is required in cases such as this.

3343T

Utility of CD-138 negative fraction for chromosome analysis in Plasma cell dyscrasias (PCD): A novel approach. *G. Velagaleti, G. Mohamed, W. Ehman Jr., C. Mendiola, V. Noronha, V. Ortega.* Dept Pathology, Univ Texas Hlth Sci Ctr, San Antonio, TX.

FISH analysis is superior to chromosome analysis in detecting important prognostic genetic abnormalities in PCD. However, its sensitivity is hampered due to paucity of plasma cells in whole bone marrow and often shows false-negative results when frequency of abnormal cells is below the laboratory's cut-off values. Studies have shown that the abnormality detection rate in enriched plasma cells is greater than unselected plasma cells, but purification techniques are limiting to only FISH when bone marrow volumes are inadequate. The inability to perform chromosome analysis may compromise patient care since chromosome analysis is equally important for detecting non-plasma cell related abnormalities, such as secondary myelodysplastic syndrome or when diagnosis is undefined. To resolve this critical issue and optimize limited quantity received, we designed a study where an immnuno-magnetic CD138 enriched positive selection was used for FISH while the negative fraction was used to retrieve the remaining cellular components (RCC) for chromosome analysis. After validating this approach in a pilot study, we implemented this strategy in 2012 for routine clinical diagnosis in patients with PCD. When there was adequate sample volume available, both whole bone marrow and RCC were used for chromosome analysis. Results showed 100% (96/96) success rate for chromosome analysis using RCC. PCD related genetic abnormalities had concordant results on both FISH from isolated plasma cells and chromosome analysis in 8.3% (8/96) of cases while 4.2% (4/96) of cases showed discordant results. Population variants such as loss of Y chromosome were observed in 6.3% (6/96) of cases. Abnormal karyotypes of myeloid origin or population variants were found both in whole bone marrow and RCC cultures. Karyotypes with PCD related aberrations were seen only in whole bone marrow cultures in 37.5% of the cases (3/8) while the remaining 62.5% (5/8) of cases showed them in both whole bone marrow and RCC cultures. Further analysis of the discordant cases revealed non-myeloma related abnormalities by CA [+8, del(5q), der(9) and del(13q)] while FISH showed either normal or myeloma related abnormalities. Our results confirm the feasibility of retrieving the RCC from the CD138 negative fraction for chromosome analysis and prove to be an innovative strategy for karyotyping in PCD patients with insufficient sample volumes.

3344F

A Unique Rearrangement of PDGFRA and ETV6 in a Three-way Translocation t(4;12;6) in a Patient with Acute Myeloid Leukemia Progressed from Chronic Myelomonocytic Leukemia. *N.M. Guruju, R. García, K. Wison, S. Monaghan, P. Koduru.* Det of Pathology, UT Southwestern medical center, Dallas, TX., USA.

Cryptic deletion at 4q12 leading to fusion between FIP1L1 and PDGFRA is a genetic lesion underlying myeloid neoplasms associated with hypereosinophilia. Reciprocal translocation(s) involving 4q12 leading to fusion of PDGFRA with other partners have also been infrequently reported. Here we report a unique rearrangement of PDGFRA with ETV6 due to a three-way translocation t(4;12;6) in a 70-year-old man. The patient had an 8-month history of chronic myelomonocytic leukemia that was treated with high dose hydroxyurea and azacitidine. He then presented with abdominal swelling and pain and was found to have WBC of 119×109L. Imaging revealed splenomegaly and widespread lymphadenopahty. A needle core biopsy of an abdominal mass revealed a low-grade follicular lymphoma. A bone marrow evaluation revealed acute myeloid leukemia with monocytoic differentiation and low-level involvement by follicular lymphoma. The myeloid blasts expressed CD13, CD33, CD34, CD36, CD45, CD64, CD71, CD117 and HLA-DR. Cytogenetic analysis of bone marrow cells by G-banding demonstrated a 46,XY,t(4;12;6)(q12;p13;q21) karyotype in eighteen of the 20 analyzed cells. By FISH, there was evidence of FIP1LI/PDGFRA (4q12) gene rearrangement by both interphase and metaphase analysis which showed the PDGFRA portion of the probe translocated to 12p13. Further FISH evaluation with the ETV6 break-apart probe revealed the translocation of the 5'-end of the ETV6 gene at 12p13 to chromosome band 6p21. Thus, results from both cytogenetics and FISH studies were consistent with the presence of a PDGFRA and ETV6 rearrangement. Additional studies are necessary to identify the partner at 6p21 that may have fused with ETV6. The PDGFRA/ETV6 fusion has previously been reported in one patient with chronic eosinophilic leukemia, who was responsive to imatinib. In AML, t(4;12)(q12;p13) is a rare recurrent abnormality with CHIC2/ETV6 fusion. One prior reported case of AML with t(4;12) refractory to chemotherapy was unresponsive to imatinib. Similarly, our patient did not respond to imatinib. He also failed subsequently induction chemotherapy with idarubicin and cytarabine. To the best of our knowledge, this is the first reported case showing a three-way translocation involving PDGFRA and ETV6 genes that developed into an acute myeloid leukemia without eosinophilia.

3345W

Detection of the EML4-ALK fusion gene in non-small cell lung cancer (NSCLC): our FISH service to date. *C. Durajczyk[1], D. Massie[1], K. Kerr[2], D. Stevenson[1].* 1) Cytogenetics Department, Aberdeen Royal Infirmary, National Health Service, Aberdeen, Scotland, United Kingdom; 2) Pathology Department, Aberdeen Royal Infirmary, National Health Service, Aberdeen, Scotland, United Kingdom.

The anaplastic lymphoma kinase (ALK) gene, located on 2p23, encodes a receptor tyrosine kinase involved in regulating cell proliferation and survival. Chromosomal rearrangements of the 2p23 locus can result in the production of an ALK fusion gene. One example of this is the EML4-ALK fusion gene, which is thought to be found in 3-5% of non-small cell lung cancers (NSCLC). Tyrosine kinase inhibitors (TKIs) such as Crizotinib (Xalkori, Pfizer) have been identified as potential therapies for EML4-ALK positive NSCLCs. This fusion oncogene therefore represents a significant target for novel treatments, necessitating a robust diagnostic test for evaluation of the ALK gene status in this subset of lung cancer patients. The gold standard method for identification of the EML4-ALK fusion gene is fluorescent in situ hybridisation (FISH) on formalin fixed paraffin embedded (FFPE) tissue sections. The North East of Scotland Cytogenetics laboratory has offered an ALK FISH service since March 2011, using the Vysis LSI ALK Break Apart probe. We have employed a three-tiered testing strategy in conjunction with our Pathology department, which has resulted in a selected cohort of patients tested (36 diagnostic cases) with a high abnormality rate observed (76%). In addition we have also participated in the European Thoracic Oncology Platform (ETOP) Lungscape study which has involved analysing over 40 cases so far. We present details outlining our experience of the service to date, including our technical and analytical experience, as well as examples of different abnormal cases.

3346T

Identification of novel Fusion transcripts in ALL patients. *M. Walter[1], C. Schroeder[1], C. Kyzirakos[2], P. Lang[2], M. Sturm[1], P. Bauer[1], O. Riess[1], M. Bonin[1].* 1) Dept Medical Genetics, Inst Human Genetics, Tuebingen, Germany; 2) Pediatric Hematology and Stem Cell Transplantation, Childrens Hospital, Tuebingen.

We describe a general workflow for the identification of tumor specific genetic variants as potential personalized therapeutic targets for the treatment of terminally ill ALL patients. Exom and transcriptome sequencing was performed to identify genetic differences between Tumor and normal tissue. One hallmark of cancer is genetic instability leading to large deletions, duplications or rearrangements, thereby generating novel fusion transcripts. these fusions very often are responsible for disease outbreak or progression. however, besides their important role in tumor biology, these fusions are found exclusively in the tumor and thus represent potential targets for treatment. In our cohort of 10 ALL patients we could identify and validate 5 gene fusion events in 4 samples. All gene fusions were copy number neutral an most likely result from balanced translokations. Three gene fusions were well-known, previously described fusion events in ALL patients. However, even in this very small cohort of 10 samples, we could identify two novel yet undiscribed gene fusions, which have a high probability of being responsible for tumorigenesis underlining the importance of individual tumor characterization.

3347F

Whole Genome Sequencing of Li-Fraumeni Families Reveals Heterogeneous Mutational Signatures. *K. Bhutani[1,4], D. Quarless[2,4], Q. Peng[4], K. Standish[2,4], E. Scott[4,5], S. Head[6], S. Williams[9], T. Kunicki[9], P.B. Hedlund[8], D. Nugent[9], N.J. Schork[3,4,7].* 1) Bioinformatics Graduate Program, University of California San Diego, La Jolla, CA; 2) Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA; 3) Department of Psychiatry, University of California San Diego, La Jolla, CA; 4) The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA; 5) Department of Biology, The Scripps Research Institute, La Jolla, CA; 6) Next Generation Sequencing Laboratory, The Scripps Research Institute, La Jolla, CA; 7) Department of Molecular and Experimental Medicines, The Scripps Research Institute, La Jolla, CA; 8) Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA; 9) Center for Inherited Blood Disorders, Children's Hospital of Orange County, Orange, CA.

Li-Fraumeni syndrome is a rare, autosomal dominant disease caused by a germline p53 mutation. We report the first human genome-wide investigation of somatic mutations arising in individual carriers of Li-Fraumeni-associated p53 mutations in two unrelated nuclear families. We sequenced eight whole genomes (two children and both parents in each family) using Complete Genomics, Inc technology, and conducted gene expression profiling to characterize the molecular profiles of the family members. The sequence data demonstrate a dramatic increase in somatic mutation rates among carriers of p53 mutations, with Non-negative Matrix Factorization analysis revealing differing mutational signatures between carriers of zinc-finger stabilization and DNA-binding mutations. Additionally, we explored the potential of genomic analyses to influence targeted therapies for individuals using germline sequence, somatic variants, tumor sequence variants, and tumor and blood gene expression profiles.

3348W

A role for telomere length in the genetic etiology of hematological malignancies. *N.B. Blackburn[1], J.C. Charlesworth[1,2], J.R. Marthick[1], T.D. Dyer[2], T.A. Thornton[3], R.J. Thomson[1], E.M. Tegg[4,5], K.A. Marsden[1,4], V. Srikanth[6], J. Blangero[2], R.M. Lowenthal[1,4], S.J. Foote[7], J.L. Dickinson[1].* 1) Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia; 2) Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Royal Hobart Hospital, Hobart, Tasmania, Australia; 5) School of Medicine, University of Tasmania, Hobart, Tasmania, Australia; 6) Department of Medicine, Monash University, Melbourne, Victoria, Australia; 7) Australian School of Advanced Medicine, Macquarie University, New South Wales, Australia.

Haematological malignancies (HMs) are blood cancers including leukemia, lymphoma and myeloma. One of the striking features of HMs is their propensity to cluster within families indicating that genetic variation is an important risk factor for disease. Given that telomere length has been shown to be heritable in families and genetic variants are known to influence telomere length it is hypothesised that telomere length contributes, at least in part, to the familial risk of HMs. Telomere length is a known risk factor for a range of cancers with excessive telomere shortening leading to genetic instability and resulting in the malignant transformation of cells. Increasingly telomere biology is being shown to be important in the development of HMs.

Using our Tasmanian familial HMs resource, a collection of large families with multiple cases of HMs, we measured telomere length using the monochrome multiplex quantitative PCR method in 49 familial HM cases, 169 unaffected relatives of familial HM cases, 66 non-familial HM cases and 669 population controls. Variance components modeling of telomere length in SOLAR, taking into account familial relationships, was used to determine factors that contribute to variation in telomere length in the study samples.

Through SOLAR, the heritability of telomere length was calculated to be 64.9% ($P=3.2\times10^{-4}$, SE=0.15) indicating that around 2/3 of the variation in telomere length in these families is influenced by genetics. Adjusting for age and sex we found that overall HM cases had shorter telomeres ($P=7.3\times10^{-6}$) than unaffected relatives and population controls and in contrast to our hypothesis this was observed across both familial (P=0.001) and non-familial HM cases ($P=6.9\times10^{-6}$). This finding supports the involvement of telomere length in the genetic etiology of both familial and non-familial HMs. Most recently we have used SNP data from Illumina 610 Quad arrays from the HM families and a genome wide quantitative trait association analysis using both the QM-QXM method and SOLAR to explore whether genetic variation associated with telomere length in these families is linked to disease. Our work provides further evidence for the role of telomere length in the genetic etiology of HMs.

3349T

Identification of novel mechanisms of drug resistance in BRCA1-deficient cancer cells by exome and RNA sequencing. *K.K. Dhillon[1], T. Walsh[2], S. Gulsuner[2], T. Taniguchi[1,3].* 1) Divisions of Human Biology and Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute.

Acquired resistance to platinum-based therapy is an obstacle for effective treatment of ovarian and breast cancers. BRCA1 or BRCA2 loss is commonly observed in these cancers and platinum compounds, such as cisplatin and carboplatin, are initially effective for the treatment of BRCA-deficient cancers. We have shown that re-expression of BRCA1/2 due to secondary BRCA1/2 mutations in recurrent ovarian cancers is associated with platinum resistance. However, restoration of BRCA1/2 does not account for all occurrences of platinum resistance. Therefore, we hypothesize that restoration of DNA repair even in the absence of functional BRCA1/2 may lead to platinum resistance in cancer cells. To test this hypothesis, we developed an in vitro model of cisplatin resistance using the BRCA1-mutated breast cancer cell line, HCC1937. We generated cisplatin-resistant clones by culturing cells in cisplatin. Surprisingly, none of the resistant clones showed BRCA1 re-expression or secondary BRCA1 mutations. However, consistent with our hypothesis, a subset of clones restored DNA damage-induced foci formation of CtIP, RAD51 and FANCD2, which are required for DNA repair and normally require functional BRCA1. To identify mechanisms of cisplatin resistance, we performed exome and RNA-sequencing of parental HCC1937 cells and four cisplatin-resistant clones. Data from cisplatin-sensitive HCC1937 parental cells were used as a baseline to identify genetic and expression variations unique to cisplatin-resistant clones. We found that a subset of DNA repair genes is differentially expressed in cisplatin-resistant clones that have restored DNA repair foci. In the non-DNA-repair-foci-restored clone, we identified missense mutations in RAD18 and XPC, genes involved in translesion synthesis and nucleotide excision repair, which are critical for cellular resistance to cisplatin. Additionally, we observed global up-regulation of a vast majority of histone genes across all cisplatin-resistant clones, suggesting that changes in DNA packaging may contribute to cisplatin-resistance. Currently, we are conducting functional studies to assess the role and mechanism of these observed changes in cisplatin resistance. In summary, our studies showed that altered expression of DNA repair genes and histones may lead to restoration of DNA repair and cisplatin resistance in BRCA1-deficient cancer cells.

3350F

Rhabdoid tumour in a 13 year old with Ring 22: A special case of the two-hit hypothesis. *H. Druker[1, 2, 4], L. Zahavich[1], D. Malkin[1, 3, 4, 5], S. Meyn[1, 2, 3, 5, 6].* 1) Cancer Genetics Program, Hospital Sick Children, Toronto, ON, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 3) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 4) Division of Oncology, Hospital for Sick Children, Toronto, ON, Canada; 5) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 6) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, ON, Canada.

We present a case of a severely developmentally delayed, hypotonic child with dysmorphic features and a de novo ring chromosome abnormality, 46,XX,r(22)(p11.1q13.2), who developed an abdominal rhabdoid tumour at 13 years of age. We propose a genetic mechanism and surveillance considerations for individuals with ring chromosome 22. Rhabdoid tumour syndrome is caused by germline mutations in SMARCB1, a tumour suppressor gene located at 22q11.2. Classically, young children with a germline SMARCB1 mutation have a high risk for developing atypical teratoid rhabdoid tumours of the central nervous system as well as renal and extra-renal rhabdoid tumours. Imaging is suggested for SMARCB1 mutation carriers from birth until four to six years of age for early detection and improved prognosis. Ring chromosome 22 is a rare human constitutional cytogenetic abnormality. Karyotype and FISH analyses of the patient's blood revealed an imbalance consistent with monosomy for chromosome region 22q13.2 to 22qter but retention of the SMARCB1 gene (22q11.2) within the ring. FISH and microarray analyses of tumour tissue revealed a total loss of SMARCB1 and flanking genes in the first allele and a partial deletion of SMARCB1 in the second allele (exons 6-9). Our findings are consistent with a specific modification of the classic two hit model of tumourogenesis in which the first hit is somatic loss of the unstable ring chromosome during mitosis. In those somatic cells that have lost the ring chromosome 22, a mutation in the SMARCB1 gene on the remaining chromosome 22 (the second hit) can result in tumour development. This mechanism has been previously proposed in relation to patients with ring chromosome 22 who developed NF2 associated tumours. Increased tumour risks have also been observed in carriers of other constitutional ring chromosomes, such as r(11) and Wilms' tumour (WT1) and r(13) and retinoblastoma (RB1). We conclude that tumours can arise by biallelic loss of a tumour suppressor gene through the combination of loss of a ring chromosome carrying the gene and a pathogenic second mutation. Our findings suggest that patients with a ring chromosome 22 should be monitored for SMARCB1 and NF2- related tumours, and possibly CHEK2-related tumours. As the timing of the loss of the ring may be unpredictable, surveillance should be considered in these individuals beyond what is normally considered as an appropriate age cut-off.

3351W

ATM and MDC1 independently modulate resection of DNA Double-Strand Breaks for Homologous Recombination Repair in Human Cells.
P.S Bradshaw[1,2], M. Komosa[1], M.S Meyn[1,2]. 1) Dept Gen & Genomic Biol 15-601F, Sick Kids Res Inst, Toronto, ON, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada.

Ataxia-Telangiectasia (A-T) is a human multisystem genetic disorder caused by mutation in the ATM gene and characterized by progressive ataxia, immunological defects, radiation sensitivity, cancer predisposition and genomic instability. Current models suggest that the ATM protein plays a major role in the activation of the cellular DNA Damage Response (DDR) following detection of DNA Double Strand Breaks (DSBs). Establishment of a protein platform encompassing megabase regions of modified chromatin distal to the DSB thought to facilitate the ATM-mediated DDR. We find that the ATM protein is not a member of the megabase protein platform. ATM belongs to a novel class of DDR proteins whose spatial-temporal behavior following DSB induction differs from members of the megabase protein platform and from proteins directly involved HRR. Following DNA damage induction ATM rapidly, but transiently, localizes to sites of DSBs and forms small discrete foci at damage sites that, unlike foci of γH2AX and Mre11, do not diffuse across megabase chromatic regions. While MRN is required for ATM foci formation, ATM and MDC-1 independent MRN foci do not co-localize 1 hour post-irradiation, suggesting that MRN is not required for the retention of ATM at sites of DSBs. DNA damage-induced ATM foci do not co-localize with foci of the HRR-associated proteins RPA and Rad51, suggesting that ATM is associated with unresected DSBs and may not play a direct role in HRR after DSB resection has begun. RPA Ionizing Radiation Induced Foci (IRIF) in ATM- cells are of normal intensity but fewer in number. This effect is dependent on CtIP, a protein implicated in DNA resection and a target for ATM phosphorylation. In addition, depletion of CtIP in ATM+ cells results in a decrease in the number of both RPA and Rad51 IRIF and an increase in ATM IRIF. Taken together, these data support a model where ATM-mediated activation of CtIP initiates resection followed by generation of RPA-coated single-stranded DNA and the concomitant displacement of ATM from DSB sites. While loss of ATM appears to affect the proportion of DSBs that undergo resection but not the extent of resection, depletion of MDC1 increases the intensity of RPA IRIFs without affecting their number. This result suggests that ATM and MDC1 independently modulate HRR through control of DSB resection: ATM facilitating the CtIP-dependent decision to undergo resection and MDC1 limiting the extent of resection.

3352T

Identification of an extragenic mutation of *BUB1B* gene for PCS (MVA) syndrome and functional analysis using TAL effector nucleases.
T. Miyamoto[1], H. Ochiai[1,2], A. Kanai[3], K. Hosoba[1], S. Kume[1], T. Sakuma[2], T. Kajii[4], T. Yamamoto[2], S. Matsuura[1]. 1) Genetics & Cell Biol, RIRBM, Hiroshima University, Hiroshima, Japan; 2) Math & Life Sci., Grad. Sch. Sci., Hiroshima University, Higashi-Hiroshima, Japan; 3) Mol.Oncol., RIRBM, Hiroshima University, Hiroshima, Japan; 4) Hachioji, Tokyo 192-0023, Japan.

Cancer-prone syndrome of premature chromatid separation with mosaic variegated aneuploidy (PCS (MVA) syndrome) is a rare autosomal recessive disorder, characterized by growth retardation, microcephaly, and childhood cancer. Monoallelic mutations in *BUB1B* encoding BUBR1, a mitotic spindle assembly checkpoint regulator, has been found in individuals with the syndrome, but no second mutation was found in the second allele although a decrease of their transcript was identified. To uncover the molecular basis of the second allele, we searched for mutation in a 200-kb genomic region using a next generation sequencer.

A novel single nucleotide substitution was identified in an extragenic region of *BUB1B* in the second allele. We then introduced it into human cultured cell lines by genome-editing technique using TAL effector nucleases (TALENs). The cell clones obtained all showed reduced *BUB1B* transcripts, suggesting that the nucleotide substitution is the causal mutation for the disease. Combination of next-generation sequencing and TALEN-mediated genome editing is a very powerful approach to study the molecular pathology of the disease with hitherto undetectable mutations.

3353F

Immunohistochemical Expression of Afadin-6 (AF-6) in Sporadic Neurofibromas.
M. Sulaiman[1], A. Dodson[2], T. Helliwell[2], C. Kudi[1], I. Hussaini[1], S. Ojo[1], J. Hambolu[1]. 1) Vet Anatomy, Ahmadu Bello University, Zaria, Nigeria; 2) Institute of Translational Medicine, University of Liverpool, UK.

Background Neurofibromatosis type 1 (NF1) - is a neurocutaneous single-gene (neurofibromin) disorder which is mapped on the long arm of chromosome 17(17q12).NF1 is inherited in an autosomal dominant manner with an incidence of about 1 in 3000. Tight junctions are specialized cell-cell point of adhesion that contributes to the regulation of differentiation and proliferation. The loss of epithelial differentiation in tumours often correlates with mutations in small GTPase Ras. The AF-6 protein is a protein that contains two potential Ras binding domains. The AF-6 functionally links the cytoskeleton through the intercellular junctions. This study was carried out to demonstrate the immunolocalization and cellular expression of AF-6 in sporadic neurofibroma by immunohistochemistry. Methods Informed patient consent was obtained before surgery and the study has an ethical approval (06/1505/137) of Liverpool Research Ethics Committee. Standard Operating Procedure of the Institute of Translational Medicine, University of Liverpool was used in the immunohistochemistry technique. Both the test and control tissues were immunostained with Rabbit Anti-AF-6 polyclonal antibody diluted at 1:100-1:200 at (pH 7.0) (Cat # 433280, Invitrogen). Slides were visualised under light microscopy. Results The AF-6 moderately localizes to the membrane of endothelial cells and at perineurial fibroblast cell-cell junction. Additionally, the expression of AF-6 in Schwann cells of all the Sporadic NF1 was nuclei in localization. Conclusion The study suggests that AF-6 may be involved in cell proliferation in neurofibroma cells and therefore becomes a target protein in the management of neurofibromatosis type 1.

3354W

Fragile site FS2 instability stimulates mitotic recombination in the yeast *Saccharomyces cerevisiae*.
K. Kapellas, S. Miller, D. Rosen, E. Younkin, A. Casper. Department of Biology, Eastern Michigan University, Ypsilanti, MI.

Common fragile sites (CFS) are regions of chromosomes in mammalian cells that are prone to DNA breaks under replication stress. CFS are prevalent throughout the human genome and some lie near or within tumor suppressor genes or oncogenes. Breaks at CFS can lead to gene deletions and amplifications that can result in the genesis of cancer cells. We hypothesize that fragile site breaks also contribute to tumorogenesis by stimulating mitotic recombination events, including reciprocal crossovers (RCOs) and break-induced replication (BIR), leading to loss of heterozygosity at tumor suppressor genes. We have examined the role of fragile site instability on the stimulation of mitotic recombination using a yeast model system. The yeast *Saccharomyces cerevisiae* possesses the native fragile site FS2 on chromosome III. Yeast fragile site FS2 exhibits double strand DNA breaks in response to DNA replication stress, but differs in structure to mammalian CFS, being composed of simple inverted repeats. We used a galactose-inducible promoter to regulate expression of the *POL1* gene of S. cerevisiae to control replication stress by galactose concentration in the medium. Three diploid yeast strains were used: an experimental strain with the native FS2 structure (AMC310), one control strain (Y382) with the FS2 region interrupted and inactivated by the insertion of a drug resistance gene, and a second control strain with the native *POL1* promoter and native FS2 (AMC324). Under low-galactose conditions that cause replication stress in this strain, experimental diploid AMC310 had a RCO frequency of $15.6*10^{-4}$, a BIR frequency of $23.8*10^{-4}$, and a chromosome loss frequency of $5.7*10^{-4}$. Under no-galactose conditions, the control strain with stabilized FS2 region, Y382, exhibited a RCO rate 5-fold lower than that of AMC310, a BIR rate 12.5-fold lower than that of AMC 310, and no chromosome loss was observed. Similar results were obtained from control strain AMC324. The results indicate that FS2 is a hotspot for mitotic recombination events that lead to LOH. Future research will determine whether human CFS instability contributes to an increase in mitotic recombination and LOH in tumors.

3355T

SirT7 promotes genomic stability and adequate DNA damage response. *B. Vazquez[1], J. Thackray[1], S. Bunting[2], J.A. Tischfield[1], L. Serrano[1].* 1) Human Genetics Institute of New Jersey Rutgers University, Piscataway, NJ; 2) Dept. of Molecular Biology and Biochemistry, Rutgers University Piscataway, NJ.

The Sir2 family of proteins, or Sirtuins, is a major player in sensing and coordinating cellular stress response by modulating chromatin structure, cell cycle progression and mitochondrial function. Mammals have seven Sirtuins, denoted SirT1-7. They exhibit different enzymatic activities such as deacetylation, have a myriad of substrates (histone and non-histone proteins) and present a diverse pattern of cellular localization. Sirtuin deficiency in mice has also been associated with increased mutagenesis, defective DNA repair pathways, tumorgenesis and accelerated aging. SirT7 is the least studied mammalian Sirtuin. Opposing roles have been attributed to SirT7 in cancer development. Up-regulation of SirT7 expression has been observed in breast and thyroid cancer and human hepatocellular carcinomas, indicating it may be involved in the promotion and/or maintenance of oncogenic features. However, down regulation of SirT7 expression is also observed upon human and mouse cell transformation suggesting that it may have a role in cancer prevention. We investigate the role of SirT7 in the maintenance of genome integrity by characterizing SirT7 knockout mice. Our results show that SirT7-/- pups are born at sub-mendelian ratio and a high proportion died within the first month of age. In vivo mutagenesis analysis show increased mutant frequency and polyploidy in SirT7 deficient cells. Moreover, SirT7-/- fibroblasts showed increased levels of DNA double strand breaks measured by immune detection of γ-H2AX foci and comet assays. Furthermore, SirT7-/- thymocytes present reduced survival upon X-ray irradiation, and mature B cells are reduced in number and present impaired immunoglobulin class switching. Overall, the loss of SirT7 is associated with embryonic lethality, increased levels of spontaneous DNA damage and an impaired DNA damage response supporting the importance of SirT7 in the maintenance of genome integrity. Acknowledgments: The authors are supported by a grant from the Human Genetics Institute of New Jersey. BN Vazquez is supported by post-doctoral fellowship EX-2010-278 from the Spanish Ministry of Education, Culture, and Sports.

3356F

The Fanconi anemia pathway regulates ALT telomere maintenance in human cells. *H. Root[1], M. Komosa[1], A. Larsen[1,2], D.P. Bazett-Jones[1,2], M.S. Meyn[1,3].* 1) Genetics & Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 2) Biochemistry, University of Toronto, Toronto, ON, Canada; 3) Molecular Genetics, University of Toronto, Toronto, ON, Canada.

Mutations in 15 different genes give rise to Fanconi anemia (FA), a genome instability syndrome characterized by bone marrow failure, malformations and cancer. FA proteins play roles in DNA repair, replication, meiotic recombination and telomere maintenance. The alternative lengthening of telomeres (ALT) pathway is a telomerase-independent recombination based mechanism active in cancer and immortalized human cells. We find that siRNA knockdown of FANCD2, FANCA, FAN1 or FANCP all cause rapid and dramatic increases in ALT telomeric DNA, demonstrating a central role for the FA pathway in the regulation of ALT.

Depletion of FANCD2 in ALT cells leads to increases in telomere length and extrachromosomal telomeric repeat (ECTR) DNA content. ECTR DNA molecules were analyzed using our novel halo-FISH technique, wherein cells are deproteinated and treated with NaOH prior to Q-FISH analysis. NaOH disrupts plectonemic interactions between ECTR DNA molecules, which normally tether ECTR DNAs together within ALT-associated PML bodies (APBs). APBs are putative sites of telomeric recombination, as well as ECTR DNA generation and storage. Electron spectroscopic imaging of APBs primarily reveals non-nucleosomal DNA, likely extrachromosomal in nature. FANCD2-depletion leads to APBs that also contain chromatin intrusions, potentially representing telomeres themselves, within APBs. FISH analysis shows that in FANCD2-depleted cells the frequency and stability of interactions between telomeres and ECTR DNAs is increased, suggesting that FANCD2 normally suppresses interactions between ALT telomeric DNAs.

FANCD2 regulates a mechanism of telomeric recombination that is independent of RAD51, but requires BLM and FANCJ expression. Knockdown or inhibition of DNA-PKcs, or KU80 does not limit telomeric DNA amplification in FANCD2-depleted ALT cells, suggesting that the critical role of FANCD2 in ALT is not to limit aberrant end-joining reactions. Interestingly, we also find that depletion of ATR does not result in increased telomeric DNA synthesis in ALT cells, but rather ATR expression is required to mediate the FAND2-depletion phenotype. This suggests that the role of FANCD2 in ALT is not tied to the response to stalled/collapsed replication forks. Together, our results indicate a direct role for FANCD2 and other members of the FA pathway in the regulation of recombination in the ALT pathway.

3357W

Anaplastic rhabdomyosarcoma in TP53 germline mutation carriers. *S. Hettmer[1], N. Archer[1], G. Somers[2], A. Novokmet[3], A. Wagers[4], L. Diller[1], C. Rodriguez-Galindo[1], L. Teot[5], D. Malkin[3].* 1) Department of Pediatric Oncology, Dana-Farber Cancer Institute and Division of Pediatric Hematology/ Oncology, Boston Children's Hospital, Boston, MA 02115, USA; 2) Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada M5G 1X8; 3) Division of Oncology, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada M5G IX8; 4) Howard Hughes Medical Institute, Department of Stem Cell and Regenerative Biology, Harvard University, Harvard Stem Cell Institute, Cambridge, MA, and Joslin Diabetes Center, Boston, MA 02115, USA; 5) Department of Pathology, Boston Children's Hospital, Boston, MA 02115, USA.

Rhabdomyosarcoma (RMS) represents a diverse category of myogenic malignancies with marked differences in molecular alterations and histology. To determine if RMS predisposition due to germline TP53 mutations correlates with certain RMS histologies, we performed a retrospective review of RMS diagnosed in 8 consecutive children with TP53 germline mutations at Boston Children's Hospital and the Hospital for Sick Children, Toronto. All 8 tumors exhibited non-alveolar, anaplastic histology as evidenced by the presence of enlarged hyperchromatic nuclei with or without atypical mitotic figures. Additionally, TP53 germline mutations were found in 3 out of 7 consecutive children with anaplastic RMS (anRMS) and previously unknown TP53 status. Thus, the frequency of TP53 germline mutations was 73% (11 out of 15 children) in pediatric anRMS patients. AnRMS was the first malignant diagnosis for all 11 TP53 germline mutation carriers in this cohort, and median age at diagnosis was 40 months (mean 40±15 months, range 19-67 months). The frequency of germline TP53 mutations in children with anRMS was 100% (5 out of 5 children) for those with a family cancer history consistent with Li-Fraumeni syndrome (LFS), and 80% (4 out of 5 children) for those without an LFS cancer phenotype. Our data suggest that individuals who harbor germline TP53 mutations are predisposed to develop anRMS at a young age; we propose extending the Chompret criteria for LFS to include children with anRMS irrespective of family history.

3358T

Polymorphisms of MTHFR, prothrombin and factor V (Leiden) genes in children with acute lymphoblastic leukemia. *E. Maly[1], O. Zajac-Spychala[2], J. Kedzierska[2], J. Nowak[3], D. Januszkiewicz-Lewandowska[1,2,3].* 1) Department of Medical Diagnostics, Poznan, Poland; 2) Department of Pediatric Oncology, Hematology and Transplantation, University of Medical Sciences, Poznan, Poland; 3) Institute of Human Genetics Polish Academy of Sciences, Poznan, Poland.

The onset of thrombosis in children with acute lymphoblastic leukemia (ALL) in the course of treatment varies according to different authors, from 2.5% to 11.6% of the patients. In addition to factors related to the underlying disease and its treatment (steroids, L-ASPA, MTX, central catheter implantation), it is not without significance the presence of genetic predisposition leading to the development of thrombosis. The most common cause of congenital thrombophilia outside deficiency of protein S, C, and antithrombin, are the gene mutations of factor V Leiden, prothrombin, and MTHFR. The aim of this study was to evaluate the prevalence of polymorphisms C677T and A1289C of MTHFR, as well as prothrombin gene mutation G20210A and factor V G1691A mutation. The study was performed in 29 children (19 boys, 10 girls) treated for ALL in the Department of Pediatri Oncology, Hematology and Transplantation, University of Medical Sciences Poznan. Homozygous C677T polymorphism of MTHFR gene was found in five (17.9%) and heterozygous in 11 patients (39.2%), whereas homozygous A1289C polymorphism of MTHFR gene was observed in 4 (14.3%) and heterozygous in 11 patients (39 , 3%). Heterozygous mutation G1691A of factor V was found in 5 (17.2%) children. No G20210A prothrombin gene mutation was observed in the study group. The simultaneous presence of two or more polymorphisms of studied genes were found in 11 (38%) children. 3 (27%) of these patients had symptomatic thrombosis requiring treatment with unfractionated heparin. In conclusion the frequency of homo-and heterozygous gene polymorphisms in children with ALL was similar to the data for Caucasians. Only the simultaneous presence of two or more polymorphisms of studied genes seems to be a predisposing factor for symptomatic thrombosis in the course of treatment of children with ALL.

3359F

Offspring of Couples Who Both Survived Cancer. *J.J. Mulvihill[1], J.F. Winther[2], L. Madanat-Harjuoja[3,4], P.M. Lähteenmäki[3,4], J.D. Boice Jr.[5,6].* 1) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; 2) Danish Cancer Society Research Center, Copenhagen, Denmark; 3) Finnish Cancer Registry, Helsinki, Finland; 4) Turku University Hospital, Turku, Finland; 5) National Council on Radiation Protection and Measurements, Bethesda, MA, USA; 6) Vanderbilt University, Nashville, TN, USA.

To date and contrary to intuition and mouse studies, no environmental exposure has been proven to cause new heritable disease in human beings, not among children born to survivors of the American atomic bombs in Japan nor to survivors of cancer in childhood, adolescence, or young adulthood who received chemotherapy, radiotherapy or both. In our population-based studies of reproduction by cancer survivors in Denmark and Finland, we had the opportunity to study the birth defects, cancer, and possibly genetic disorders in offspring of that rare event when BOTH parents were cancer survivors. We identified all cases of connubial cancer in two population-based cancer survivor cohorts comprising 54,349 cancer survivors in the Danish or Finnish cancer registries with cancer under age 35 years between the start of the respective registries (1943 in Denmark and 1953 in Finland) through 2004, who survived to reproductive age. Offspring were linked to health registries to identify malformations, cancer, deaths (including stillbirths), selected single gene disorders and chromosomal abnormalities. A total of 58 connubial cancer families with a total of 110 offspring have been identified. Only 14 children were born (in 4 families) after the cancer diagnoses of BOTH parents: In 3 families each offspring had an adverse outcome (pyloric stenois, cleft lip and palate, and brain ependymoma, the last in a Li-Fraumeni syndrome family); in the fourth family, 2 of 11 offspring had an abnormality (branchiogenic cyst, terminal phalangeal aplasia of toes 1-3). In the 9 other affected offspring (born after just one parent had cancer), the father was the cancer survivor in 8 cases. No offspring has a *de novo* mendelian trait. Although numbers are small, they are reassuring that therapy that is reputed to be highly mutagenic in somatic cells (seen as second cancers and chromosomal breakage in the survivors) is not causing an excess of germ cell mutations.

3360W

Pleuropulmonary blastoma (PPB) type and distant metastases are significantly associated with disease-free survival: a report from the International PPB Registry. *D.R. Stewart[1], P.S. Rosenberg[2], G.M. Williams[3], J.R. Priest[3], KA.P. Schultz[3], A. Harris[3], L. Doros[4], D.A. Hill[4], L.P. Dehner[5], Y.H. Messinger[3].* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 3) International PPB Registry and Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 4) Department of Pathology, Children's National Medical Center, Washington, DC; 5) Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO.

BACKGROUND. Pleuropulmonary blastoma (PPB) is the most common primary malignancy of the pulmonary parenchyma in childhood. Three types of PPB have been recognized: type I (purely cystic), type II (cystic and solid) and type III (purely solid). Type I PPB may also regress (or not progress), resulting in a cyst with the architecture of type I PPB but without the primitive cells (type Ir). Familial PPB is an autosomal dominant, pleiotropic, tumor-predisposition disorder with incomplete penetrance that arises secondary to mutations in *DICER1*, a gene critical in microRNA biogenesis; mutations in *DICER1* are also observed in sporadic PPB. Since 1987 the International PPB Registry (IPPBR) has collected clinical and follow-up data on centrally-reviewed pathology from PPB cases collected world-wide. The last IPPBR report (1997) on the natural history, treatment and survival experience of PPB was based on 50 cases. We now report analysis of disease-free survival, prognostic factors and overall survival from 350 cases. METHODS/ RESULTS. IPPBR pathologists (LPD, DAH) performed central review of all cases; PPB was excluded in 20%. Of the 350 confirmed cases, there were 88 type I, 26 type Ir, 124 type II, 21 type II/III and 91 type III PPBs from 187 males and 162 females (1 unknown gender). In 97 PPB patients tested, 64 (66%) had germline *DICER1* loss-of-function mutations. Disease-free survival 3 years post PPB-diagnosis was significantly different by PPB type (P-value = 1.6×10^{-10}): 86% for type I, 93% for type Ir, ~64% for type II and II/III and 36% for type III. Type III had a 2.3-fold worse disease-free survival when compared to type II (relative hazard, 95% CI: 1.5-3.5; P-value = 5.0×10^{-5}). Within PPB types II, II/III and III, presence of distant metastases was significantly associated with poorer overall (P-value < 0.002) and disease-free survival (P-value < 0.01). Preliminarily, *DICER1*-mutation status did not influence survival in the limited number of patients tested to date. CONCLUSIONS. We have demonstrated, for the first time, significantly worse disease-free survival in type III vs type II PPB. In the largest cohort to date, we observed that PPB clinical subtype confers significant differences in disease-free and overall survival, suggesting biologic differences between the cystic types I and Ir and within the solid types II, II/III and III.

3361T

Improved detection of FLCN mutations in patients with Birt-Hogg-Dubé Syndrome. *J.R. Toro[1, 2], B. Friedman[3], S. Bale[3].* 1) National Cancer Institute, Bethesda, MD; 2) Dermatology Department, Veterans Affairs Medical Center, Washington, DC, USA; 3) GeneDx, BioReference Laboratories, Inc., Gaithersburg, MD Maryland, USA.

Background: Birt-Hogg-Dubé syndrome (BHDS) (MIM 135150) is an autosomal dominant predisposition to the development of follicular hamartomas (fibrofolliculomas), lung cysts, spontaneous pneumothorax, and kidney neoplasms. Germline mutations in FLCN are associated with the susceptibility for BHDS. To date 153 FLCN germline mutations have been reported in the online Folliculin sequence variation database. Objective: To characterize methods for improved FLCN mutation detection and novel mutations. Methods: Initial screening was conducted with direct bidirectional DNA sequencing of the coding regions and splice sites of exons 4-14 of FLCN. If no mutation was identified by sequencing analysis, large intragenic insertion and deletion mutations were screened by RQ-PCR and targeted arrays comparative genomic hybridization with exon-level resolution. Results: The FLCN mutation detection rate by direct sequencing was 89 percent. We detected 56 unique novel FLCN germline mutations: 22 deletions, 11 insertions, 13 missense,7 nonsense, 2 splice site and 1 deletion/insertion. To date only eight large unique intragenic mutations have been reported. We identified one whole gene FLCN deletion and eleven unique large FLCN intragenic deletions: four involving exon 1, one in exon 6 and six encompassing: exons 1-6, exons 2-5, exons 2-13, exons 6-14, exons 7-8 and exons 10-14. Including this report, to date there are 214 unique FLCN mutations identified: 76 deletions, 32 insertions, 99 substitutions and 7 deletion/insertion. A comprehensive worldwide review of published FLCN mutations and current ongoing efforts to detect novel BHDS causing mutations will be discussed. Conclusion: A systematic approach combining accurate and sensitive methods to detect FLCN mutations provides evidence that most patients with BHDS have mutations in FLCN.

3362F

Gene Expression in P53 and Bcl-2 in Biopsy Samples of Ulcerative Colitis and Colon Cancer in Iraqi Patients. *Z. Jaafar.* Biotechnology Center, Ministry of Science and Technology - Iraq, Baghdad, Baghdad, Iraq.

The investigation done on the biopsy samples from patients suffered from ulcerative colitis (UC)and colon cancer. The aim of the study is to detect the genetic alteration for P53 and Bcl-2 which consider prognostic factor for chronic ulcerative colitis that show a relation between chronic ulcerative colitis and increased risk with colon cancer by using in situ hybridization technique . The results showed an increase in genetic expression for p53 and Bcl-2 in chronic UC patients as a result of accumulative mutation in P53 gene and Bcl-2 when compared with colon cancer and it can be considered as a biomarker to detect the increase sensitivity to carcinogenesis . The extent of gene Bcl-2 in UC was 8.4% and 25% and 66.6%in low ,moderate, high grade respectively while in colon cancer gave 25% in moderate grade and 75 % in high grade . The extent in P53 in UC was 8.3 and 27.7 and 64 % in low , moderate , high grade respectively . The intensity of Bcl-2 in UC was 21.3 % in low grade and 50.8% in moderate grade and 17.8 % in high grade , while the intensity in colon cancer was 3% in low grade and 56% in moderate grade and 42% in high grade . The intensity of P53 in UC was 33.3% and 47.2% and 18.47 % in low, moderate ,high grade respectively while the intensity in P53 in colon cancer gave 8.3% and 36.1% and 55.6% and in low, moderate ,high grade respectively .

The conclusion study indicate there were an increase in gene expression in P53 and Bcl-2 genes which can be used as prognostic factor to follow the chronic stage of UC and detect the UC -associated neoplasm lesion in the early stage and therefore decrease the risk of developing to colon cancer .

3363W

Integrative analysis of acute myeloid leukemia with genetic, epigenetic, and transcriptional data. *S. Li[1], F. Garrett-Bakelman[1], J. Patel[2], T. Hricik[2], M. Guzman[1], M. Carroll[3], A. Brown[4], R. D'Andrea[4], R. Levine[2], A. Melnick[1], C. Mason[1].* 1) Weill Cornell Medical College, New York, NY; 2) Memorial Sloan-Kettering Cancer Center, New York, NY; 3) University of Pennsylvania, Philadelphia, PA, USA; 4) The Queen Elizabeth Hospital, Woodville, South Australia.

Understanding the molecular mechanisms that lead to a patient's recurrence of cancer is one of the most important aspects of cancer research. Leukemia often presents as an abrupt cancer at diagnosis, and even after treatment, many patients experience relapse. The relapse of leukemia is hypothesized to occur when sub-clones of leukemia cells resist the chemotherapy during the treatments and then emerge to re-populate the cancerous cells in the patient. This relapse of the leukemia represents one of the greatest challenges towards improving the clinical outcome for patients. The current study has identified the evolution of molecular changes in tumors at diagnosis and relapse stages in acute myeloid leukemia (AML) patients. We observe thousands of diagnosis-specific and relapse-specific changes at the genetic, epigenetic and transcriptome level. We have detected both known and novel somatic variants from Exome-seq data and validated them using Enhanced Reduced representation bisulfite sequencing (ERRBS) and RNA-seq with a high validation rate (96-99%). We also found known/novel fusions and differentially expressed genes from RNA-seq data. Also, we detected differentially methylated CpGs and regions from ERRBS data. Furthermore, we found distinct DNA methylation patterns associate with gene expression from CG poor to CG rich promoters. These molecular changes are potential candidates of prognostic factors and targets for relapse treatment and provide a wealth of new information about the evolution of cancer, which will be useful for AML as well as other tumors.

3364T

Tissue-of-origin chromatin organization shapes the mutational landscape of cancer. *P. Polak[1,2], R. Karlic[3], A. Koren[4,2], B. Thurman[5], R. Sandstrom[5], A. Reynolds[5], E. Rynes[5], J. Stamatoyannopoulos[5], S. Sunyaev[1,2].* 1) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Broad Institute, Boston, MA; 3) University of Zagreb, Zagreb, Croatia; 4) Harvard Medical School, Boston, MA; 5) University of Washington, Seattle, WA.

Density of somatic mutations in cancers varies greatly along the genome, and chromatin organization has been implicated as a major determinant of the mutational landscape in cancer. However, both the regional variation of somatic mutations and chromatin organization are highly cell type specific properties. We analyzed genomes of five major types of cancer representing a range of tissues of origin and mutational mechanisms and signatures. We compared the genomic distribution of cancer mutations to a large set of epigenetic properties that included chromatin accessibility as measured by DNaseI hypersensitivity, histone modifications and DNA replication timing from multiple cell types. In order to evaluate the contribution of different epigenetic properties to the local variation in mutation density, we used a machine learning tool called Random Forest regression. Random Forest provides a framework to rank individual predictors and to identify the set of strongest predictors of local mutation density, and also numerically scores the contribution of each individual predictor. For each cancer type, we focused on the predominant mutation type and initially combined all mutations from the different genomes of that same cancer type. We analyzed mutation density variation at 1Mb scales. Remarkably, epigenetic marks explain a large fraction of variance in mutation density for all cancer types. Variance explained ranges from 65% to 88%. This is substantially higher than in earlier studies, and indicates that, at least for several cancer types, we have identified a set of epigenetic variables that almost fully predict the mutation variability along the genome. Furthermore, chromatin marks from the matched tissue of the cell-of-origin of a cancer, e.g. melanocytes and melanoma, explain a much larger proportion of the variance than the chromatin marks from non-matched tissues.

3365F

Association of eNOS 4 a/b polymorphism in Mexican patients with Breast Cancer. *R. Ramírez[1,2], MP. Gallegos[1], A. Ramos[1,3], L. Gómez'Flores[1,2], D. Carrillo[1,2], O. Soto[1,3], I. Gutíerrez[1,2], I. Delgado[5], AM. Puebla[5], LE. Figuera[4], RP. Mariaud[6], GM. Zúñiga[7].* 1) Laboratorio de Genética Molecula , Division de Médicina Molecular, CIBO, IMSS, Guadalajara, Jalisco, Mexico; 2) Doctorado en Genética HUmana, CUCS, Universidad de Guadalajara; 3) Doctorado en Farmacología, CUCS, Universidad de Guadalajara; 4) División de Genética, CIBO, IMSS; 5) Laboratorio de Inmunofarmacologia, CUCEI, Universidad de Guadalajara; 6) Departamento de Clínicas Odontológicas Integrales, Centro Universitario de Ciencias de la Salud, Universidad Guadalajara; 7) Laboratorio de mutagénesis, CIBO, IMSS.

Breast cancer (BC) is one of the most common diseases in developing countries in the world. It is estimated that there are millions of symptomatic women affected by BC and millions more currently asymptomatic that will develop cancer. Several polymorphisms in the gene for endothelial nitric oxide synthase (eNOS) have been associated with the different diseases including cancer, one of the most studied has been the 4a/b, and however the results have been contradiction in different parts of the world. The aim of this study was to determine the association of eNOS 4a/b polymorphism in BC patients. Were included samples from 429 BC patients and 281 controls from the general population, and were genotipified to eNOS 4 a/b polymorphism in the intron 4 of gene. The allele identification was performance by polyacrylamide gel electrophoresis after staining with silver nitrate. The observed genotype frequencies for controls and BC were 0.6% and 0.7% for a/a, 12% and 22% for a/b, 0.6% for allele b/c only control group and 87% and 77% for b/b, respectively. The frequencies of the genotype aa-ab of eNOS 4 a/b polymorphism showed significant differences ($p < .05$) when comparing the study groups. We conclude that the genotypes a/a-a/b of the eNOS 4 a/b polymorphism contribute significantly to breast cancer susceptibility in the analyzed sample from the Mexican population.

3366W

Analysis of methylation pattern of candidates genes regulated by TDG in patients with germline mutations in TP53 gene. *F. Fortes[1], H. Kuasne[2], F. Marchi[2,4], S. Rogatto[2,3], M. Achatz[1].* 1) A.C. Camargo Cancer Center. Department of Oncogenetics. São Paulo/SP- Brazil Sao Paulo, Brazil; 2) CIPE - NeoGene Laboratory, AC Camargo Cancer Center, Fundação Antonio Prudente, São Paulo, São Paulo, Brazil; 3) Department of Urology, School of Medicine, UNESP - São Paulo State University, Botucatu, São Paulo, Brazil; 4) Inter-institutional Grad Program on Bioinformatics, Institute of Mathematics and Statistics, USP - São Paulo University, São Paulo, São Paulo, Brazil.

Li Fraumeni syndrome (LFS) is a rare autosomal dominant syndrome caused by germline mutations in the tumor suppressor gene *TP53* that predispose to hereditary cancer. In Brazil, a LFS variant form is often due to the occurrence of a founder effect, characterized as a mutation which corresponds to the exchange of an arginine by a histidine at codon 337 (p.R337H mutation). It is known that p53 regulates several cellular pathways important to cell cycle regulation, including Thymine DNA glycosylase (TDG), a specialized glycosylase that protects the hypermethylation of CpG islands and activates demethylation of promoters and tissue-specific enhancers. TDG is transcriptionally regulated by p53 and is responsible for epigenetic regulation of several "sentinel" genes. DNA methylation is one of epigenetic alterations common in human cancer. The aim of this study was to evaluate the methylation pattern of "sentinel" genes regulated by TDG: JUN, OCT4, SOX2, HOXD8, SOX17, FOXA1, NKX2.2 and repetitive regions ALUyB8 of LFS patients with and without TP53 mutations. Using quantitative bisulfite pyrosequencing the pattern of gene methylation was evaluated from DNA extracted of peripheral blood (five groups with 10 patients) and DNA extracted of tumor tissue (one group with 7 LFS patients). The 5 groups were composed by patients with (1) cancer and p.R337H mutation, (2) without cancer and p.R337H, (3) cancer and other mutation in TP53, (4) relatives without cancer and without mutation in TP53, (5) controls without cancer history and (6) 7 tumors of patients with LFS. It was not found any statistical difference between the peripheral blood group, but the genes SOX17 and HOXD8 showed altered methylation levels in the tumor group compared to peripheral blood groups. The methylation patterns had not changed when the tissue analyzed was blood and possibly these genes are epigenetically regulated by other factors than TDG. Regarding to the tumor tissue, the role of TDG in regulating HOXD8 and SOX17 methylation cannot be assumed and it will be better assessed in other groups with and without TP53 mutation. These results reinforce the importance of other mechanisms which may be related to carcinogenesis in LFS.

3367T

Functional characterization of mutations in the spliceosomal component SF3B1 in uveal melanoma. *A. Bowcock[1, 2], M. Sentmanat[1], H.N. Anbunathan[2], E.D.O. Roberson[3], J.W. Harbour[4].* 1) Department of Genetics, Washington University, St. Louis, MO; 2) National Heart & Lung Institute, Imperial College London, London, UK; 3) Division of Rheumatology, Washington University School of Medicine, St. Louis, MO; 4) Bascom Palmer Eye Institute, University of Miami, Miami, FL.

Spliceosome components have recently been implicated as oncogenic drivers of tumor progression for a variety of human cancers. We recently reported recurrent mutations at amino acid 625 of splicing factor 3b subunit 1 (*SF3B1*) in ~20% of uveal melanomas. SF3B1 is a component of the U2 small nuclear ribonucleoprotein complex and the minor U12-type spliceosome. The presence of an *SF3B1* mutation is a key genetic indicator of prognostic outcome in uveal melanoma. *SF3B1* mutations are rarely seen in tumors that metastasized and their presence is inversely correlated with the presence of deleterious mutations in BRCA1 associated protein 1 (*BAP1*) that are associated with strong likelihood of metastasis. Mutations in *SF3B1* have also been reported for chronic lymphoid leukemia, myeloid dysplastic syndromes, breast cancer, and other solid tumors - emphasizing the need to functionally characterize the effects of these *SF3B1* mutations. To achieve this we have performed knockdown and overexpression studies of *SF3B1* in cell culture followed by expression analysis with array-based approaches and RNA-Seq. Knockdown of *SF3B1* in uveal melanoma cell lines from primary tumors results in the differential regulation of several oncogenes with roles in cell proliferation, angiogenesis and apoptosis. Transient overexpression of mutant forms of SF3B1 (p.R625C, p.R625H, p.R625L, and p.R625G) found in primary uveal melanomas leads to the upregulation of tumor promoting transcripts such as *E2F3*. No overlap of differentially regulated transcripts was observed between knockdown and *SF3B1* mutant-expressing lines consistent with SF3B1 mutations being gain-of-function alterations. *CLK4,* which regulates effectors involved in splice-site recognition was among the upregulated transcripts identified in SF3B1-mutant expressing lines. Data on these cell lines are being compared with RNA-Seq data from *SF3B1* mutant and wildtype uveal melanoma tumors to identify critical changes mediated by these mutations that are involved in tumorigenesis.

3368F

Role of rs9620497 with CRYBB2 expression levels in African American women with invasive breast cancer. *S.M. Brown[1], C.D. Shriver[2], R.E. Ellsworth[3].* 1) Windber Research Institute, Windber, PA; 2) Walter Reed National Military Medical Center, Bethesda, MD; 3) Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD.

Expression levels of crystallin, beta-B2 (CRYBB2) has been found to be significantly higher in sex-derived tumors such as prostate and breast from African American compared to Caucasian individuals. Because African Americans have less favorable outcomes compared to Caucasians, overexpression of CRYBB2 may contribute to these outcome disparities. SNP rs9620497 has previously been associated with increased risk of prostate cancer in African American men, thus, we evaluated the association between rs9620497 and breast cancer risk in African American women (AAW). Genomic DNA from AAW with invasive breast cancer (n=200), AAW controls (n=244) and Caucasian women with invasive breast cancer (n=243) was genotyped using a TaqMan SNP assay. Gene expression levels were evaluated by qRT-PCR using RNA isolated from breast tumors from 41 CW and39 AAW. Genotype frequencies were analyzed using chi-square analysis and gene expression levels by a Mann-Whitney U-test; P<0.05 was used to define significance. Genotype frequencies did not differ significantly between any of the groups with GT/TT genotypes found in 19% of AAW cases, 14% of AAW controls and 21% of CW cases. Gene expression levels did not differ significantly (P=0.70) between patients with GG and GT/TT genotypes. Thus, these data do not support findings presented at an earlier ASHG meeting that rs9620497 is associated with cancer risk in African Americans. In addition, this SNP is not correlated with expression levels of CRYBB2 and whether higher levels of CRYBB2 in tumor from AAW is associated with differences in tumor etiology and outcome disparities or whether this is the result of population stratification must be determined.

3369W

Functional characterization of melanoma-associated common variants in *PARP1*. *J. Choi[1], M.M. Makowski[1], W.J. Kim[1], T. Zhang[1], M.H. Law[2], M. Xu[1], M. Kovacs[1], H. Parikh[1], L.G. Aoude[2], M. Gartside[2], H.H. Yin[3], J.M. Trent[3], S. Macgregor[2], N.K. Hayward[2], K.M. Brown[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Queensland Institute of Medical Research, Brisbane, QLD, Australia; 3) Translational Genomics Research Institute, Phoenix, AZ.

Recent genome wide association studies (GWAS) identified several new loci for melanoma susceptibility. While these results highlight potential pathways predisposing to melanoma, functional risk variants in these regions, as well as mechanism by which they influence risk, have yet to be elucidated. To nominate potential functional variants we assessed whether SNPs in melanoma loci affect flanking gene levels by expression quantitative trait loci (eQTL) analysis. Transcript levels were measured in 62 melanoma cell lines using Affymetrix U133Plus2 expression microarray. SNPs were then typed on Illumina OmniExpress arrays and further imputed to 1000 genomes (v3) using IMPUTE2. Regional copy number variations were also estimated and adjusted in the analysis using mach2qtl. Among 16 loci two chromosome 1 loci exhibited significant cis-eQTL. Subsequent validation using Taqman quantitative PCR (qPCR) demonstrated that Poly [ADP-ribose] polymerase 1 (*PARP1*) expression is significantly associated with the GWAS lead SNP (p=0.03). Namely, the risk allele is correlated with an increased *PARP1* transcript levels. We then further interrogated the genotype-expression correlation by Taqman allele discrimination qPCR in 21 melanoma cell lines heterozygous for the lead SNP. The results demonstrated significantly higher proportion for the risk allele in *PARP1* transcripts (p=0.0001). These findings were also cross-validated using RNA sequencing on a subset of the heterozygous cell lines. To identify functional risk variants mediating these effects we annotated *PARP1* locus using ENCODE database. Among 43 SNPs of strong linkage disequilibrium with the GWAS lead SNP (r2>0.8), three exhibited strong evidence as potential transcriptional enhancers in melanoma relevant cell types. Luciferase assays on these candidate SNPs demonstrated that the protective allele of rs1417765 displayed four-fold higher reporter gene expression (p=0.001). Consistent with these results, Electro Mobility Shift Assays indicated stronger nuclear protein binding for the same allele. Protein identification using antibody super-shift and mass-spectrometry will provide further explanation for the correlation of this SNP with *PARP1* expression. Our data suggest that increased *PARP1* levels are correlated with melanoma risk. Further analyses will better elucidate *PARP1* function in melanoma susceptibility.

3370T

Genetic polymorphisms in DNA base excision repair gene XRCC1 and the risk of Head and Neck Cancer. *K. Chukka[1], Z. Vishnuvardhan[2], U. Radhakrishna[3].* 1) Department of Biotechnology, Acharya Nagarjuna University, Guntur, India; 2) Department of Botany & Microbiology, Acharya Nagarjuna University, Guntur, India; 3) Green cross Pathology and Molecular Biology laboratory, Paldi, Ahmedabad.

Introduction: Head and neck cancers (HNC) are among the most common types of cancer and represent a major health problem; there are approximately 540,000 new cases and 271,000 deaths annually worldwide for a mortality of approximately 50%. Single-nucleotide polymorphisms (SNPs) are the most common form of genetic variations found in the human population. The aim of this study was to evaluate the association between polymorphism XRCC1 gene and patients with head and neck cancer. Materials and Methods: 30 head and neck cancer patients selected from the local government general hospital. An age- and sex-matched cancer-free control group (n-30) was used to compare the frequency of polymorph variants. DNA was extracted from peripheral blood. SNP were genotyped by direct sequencing. Results: Data showed no significant allelic associations for XRCC1 347bp/445bp. 80% (n-24) of the head and neck cancer cases were Heterozygous only 20% (n-6) of the cases were Homozygous. All the controls were Heterozygous. Conclusion: XRCC1 SNPs 347bp/445bp may not be biomarker for head and neck cancer. The complex analysis of these factors may provide the basis for personal risk assesement and an opportunity for individualised therapy.

3371F

Characterization of immunogenetics variants of the *CCR2*, *CCR5*, and *HLA-G* genes as potential targets for diagnosis, prognosis and treatment in women with sporadic and familial breast cancer. *C. de Oliveira Giongo[1], A.P. Carneiro Brandalize[2], P. Ashton Prolla[1,2], J.A. Bogo Chies[1].* 1) Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil; 2) Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

Breast cancer is the most common cancer among women. It is estimated that 5 to 10% of the breast cancers are represented by familial breast cancers and 90 to 95% are represented by sporadic breast cancers. Mutations can lead to a change or loss of expression a different genes and this allows the appearance of genetic and phenotypic features which contribute to tumor progression. Among these features is the ability of tumor cells to evade from the immune cells or even use immune cells in the promotion of a inflammatory microenvironment promotion which may help angiogenesis and, later, metastasis. The aim of our study was to evaluate four polymorphic variants of genes which encode important immune system molecules, two related genes encoding chemokine receptors, *CCR2* (rs1799864) and *CCR5* (rs333), and two related to *HLA-G* gene (rs1704 and rs1063320) in 105 women with familial breast cancer, 83 with sporadic breast cancer and 151 without cancer and family history of cancer (control group), such as potential markers for diagnosis and prognosis of breast cancer. Rs1799864 and rs1063320 polymorphisms were genotyped by PCR-RFLP. Rs333 and rs1704 polymorphism were genotyped by PCR. Allelic, genotypic and haplotypic frequencies were estimated and compared using the Chi-square test or Fisher's exact test and subsequently were associated to diagnostic and prognostic factors. We observed a higher allelic frequency of the *CCR2* wild type allele, *Val* (p=0.040, OR 0.61, CI 95%=0.38-0.98) e *CCR5* wild type allele, *Wt* (p=0.032, OR 0.46, CI 95%=0.23-0.94) and higher haplotype frequency of the double wild type variants (*Wt/Val*) of these same genes in women on the control group (p=0.030) compared to women with familial breast cancer. All polymorphisms were evaluated together with the clinical parameters and it was observed that women with breast cancer showed sporadic cancer latter (57.29±8.457 years and 44.23±12.092 years for women with sporadic and familial breast cancer respectively, p<0.001) and more invasiveness (p=0.001) as compared to women with familial breast cancer. Moreover, the rs1704 and rs1799864 polymorphism showed a positive association with tumor aggressiveness in women with sporadic breast cancer (p=0.039 and p=0.005, respectively). Our data suggest that invasive cancers may be associated with increased immune cells infiltration and inflammation in the tumor microenvironment mediated by both CCR2 receptor and HLA-G molecule. Financial support: CNPq and FIPE.

3372W

A common germline deletion of APOBEC3B alters the somatic mutation profile in breast cancer. *R.J. Delahanty[1], Y. Zhang[1], Y. Guo[2,4], Y. Shyr[3,4], W. Zheng[1], J. Long[1].* 1) Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN; 2) Department of Cancer Biology, Vanderbilt University Medical Center, Nashville, Tennessee; 3) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee; 4) Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, Tennessee.

Breast cancer is the most commonly diagnosed cancer and the leading cause of cancer deaths in women worldwide. Members of the APOBEC3 cytosine deaminase family (A3s) have been identified as the mutational source in breast cancer and likely other cancers. We recently found a strong association between a common deletion in the *APOBEC3B* gene with increased breast cancer risk. However, the biological mechanism underlying this association is unknown. In this study, we hypothesized that the deletion may alter the balance of gene expression levels of the APOBEC3 family members and the mutation spectrum of breast cancer. Using data from TCGA, the *APOBEC3B* deletion was called based on the Affymetrix SNP 6.0 data in blood and tumor samples of 842 breast cancer patients. Among them, 170 women were heterozygous and 30 were homozygous for the deletion in tumor. RPKM (gene expression in reads per exon kilobase per million mapped sequence reads) values were used to represent the expression level based on RNA-seq (level 1) data from 788 tumor and 108 matched normal samples. Among those tumor samples with RNA-Seq data, somatic mutation profile (level 2) was available for 762 samples. As expected, the *APOBEC3B* deletion was associated with decreased expression of the *APOBEC3B* gene. After adjustment for the *APOBEC3B* expression level, tumors with homozygous deletion of the *APOBEC3B* gene had higher level of somatic mutations, including overall mutation load, C-to-T mutation, and the mutations located in known breast cancer genes. We did not see any difference regarding the mutation load between the heterozygous deletion and those without deletion, suggesting that breast cancer risk and somatic mutations in women carrying homozygous deletion may occur by mechanisms other than APOBEC3B. This study identifies possible mechanisms for the association between the *APOBEC3B* gene deletion and breast cancer risk.

3373T

Functional Variants at the 5q11 Breast Cancer Risk Locus Regulate MAP3K1 Expression through Long-Range Regulatory Elements. *S.L. Edwards[1,5], D.M. Glubb[1], M.J. Maranian[2], K.B. Meyer[3], K.A. Pooley[4], K. Michailidou[2], K.M. Hillman[1], S. Kaufmann[1], G. Chenevix-Trench[1], D.F. Easton[2,4], J.D. French[1,5], A.M. Dunning[4], Breast Cancer Association Consortium.* 1) Functional Cancer Genomics, Queensland Institute of Medical Research (QIMR), Brisbane, Australia; 2) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK; 4) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 5) School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Australia.

Despite the success of genome-wide association studies (GWAS) in identifying common genetic variants associated with complex diseases, it has been difficult to demonstrate which variants are causal and how they contribute to disease risk. GWAS have previously identified a variant on 5q11.2 near MAP3K1, which is associated with breast cancer risk (rs889312; OR = 1.13; 95% CI 1.10-1.16; p=7×10-20). MAP3K1 is a key component of the mitogen-activated protein kinase (MAPK) signal transduction pathway and is responsible for regulating important cancer genes such as c-Myc, c-Jun and c-Fos. In an attempt to determine the causal variant(s) underlying this association, we analysed 300 genotyped and 609 imputed variants within the 305kb surrounding rs889312 in 89,050 Europeans in 41 case-control studies within the Breast Cancer Association Consortium (BCAC). We identified three independent association signals for estrogen receptor (ER)-positive but not ER-negative breast cancer. Using ENCODE and other published ChIPseq data, we have identified three putative regulatory elements (PREs) corresponding to the three association signals and each containing several candidate causal variants. Chromatin conformation studies showed long-range physical interactions between the PREs and MAP3K1, their likely target gene. The strongest signal mapped to a transcriptional silencer element in which the risk alleles of candidate causal variants increased MAP3K1 promoter activity in luciferase reporter assays. Electrophoretic mobility shift assays (EMSAs) have detected allele-specific protein-DNA interactions for several candidate causal variants. In conclusion, we have identified candidate variants at 5q11 that are likely to be causally related to breast cancer risk and act by controlling MAP3K1 expression.

3374F

Transcriptional regulation and prostate cancer risk loci. *R.J. Klein[1], J. Hayes[1], X. Xu[1], J. Farber[1], R.-M. Väänänen[2], P. Taimen[3], J. Vijai[1], H. Lilja[4,5], K. Pettersson[2], K. Offit[1].* 1) Clinical Genetics Service, Department of Medicine, and Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Division of Biotechnology, University of Turku, Turku, Finland; 3) Department of Pathology, University of Turku and Turku University Hospital, Turku, Finland; 4) 5.Department of Laboratory Medicine, Surgery, and Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 5) Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK.

While genome-wide association studies (GWAS) have identified numerous loci at which common single nucleotide polymorphisms (SNPs) are clearly associated with the risk of developing prostate cancer, the mechanism by which these variants influence disease is not clear. Based on the observation that regions of the genome with marks of transcriptional regulatory elements harbor SNPs that are more likely to be under negative selective pressure, we hypothesized that many prostate cancer associated SNPs, or their correlated proxies in linkage disequilibrium, function by altering regulation of nearby genes through alteration of transcription factor binding sites. As a first test of this hypothesis, we previously identified several cis-acting expression quantitative trait loci (eQTLs) in which prostate cancer risk SNPs correlate with expression of nearby genes in prostate tumors. Here, we further test this hypothesis by 1) asking to what extent prostate cancer risk SNPs can be explained by variants that alter transcriptional regulatory elements and 2) asking to what extent prostate cancer risk SNPs are correlated with nearby gene expression in normal tissue. Using data from the ENCODE Project, we find that prostate cancer risk SNPs overlap DNase hypersensitive sites more often than expected by chance in prostate-derived cell lines and several other tissue types. Many of these sites are predicted to alter binding of specific transcription factors. We further find that while many prostate cancer risk SNPs are associated with expression of nearby genes in tumor tissue, fewer such associations are observed in benign tissue. No significant association with distal genes (trans eQTLs) were found. Overall, these data support the hypothesis that many GWAS-identified risk SNPs alter transcriptional regulation of nearby genes though further studies are needed to demonstrate gene functionality and causality of prostate cancer.

3375W

Drug-sensitivity assay for sequence variants verification. *K. Tao, S. Tavtigian.* Huntsman Cancer Institue, Salt Lake City, UT.

[Background] Individual sequence variant may confer different level of drug sensitivity and affect the outcome of treatment. Emerging poly (ADP-ribose) polymerase (PARP) inhibitor-based synthetic lethal approach is beneficial for the patients with defective function in homologous recombination repair system. High-throughput sequencing technologies provide us numerous sequence variants and majority of them are missense substitutions that are difficult to interpret. However, the assay to determine which sequence variant create clinical significance or increase drug sensitivity has not been well established. Here, we developed a drug-sensitivity assay system to verify sequence variants of genes involved in HR pathway. [Method] Total twelve sequence variants were analyzed in this method. Stable cell lines that have either a wild-type or mutant construct were generated using PiggyBac (PB) transposase (PB 5' terminal repeat_EF1alpha promoter_CDS_IRES_RFP_PB 3'terminal repeat). For accurate and high-throughput screening, the cells expressing the same level of fluorescence intensity were sorted and plated in 96 well plates using FACSAria cell sorter. Cross-linking agent (Mitomycin, Cisplatin), poly ADP-ribose polymerase (PRRP) inhibitor (AZD-2281, MK4827), and Topoisomerase I (Camptothecin) were used individually or as combination. Following the drug exposure for 5-7 days, cell viability was assessed based on the amount of ATP. The ATP-generated luminescent signal of each well was measured in EnVison 2104 plate reader. [Result] Three sequence variants (5'UTR variant, missense substitution, and frameshift deletion) showed an increase sensitivity to the drugs compared to wild-type, These three mutants were sensitive to both cross-linking agent and PARP inhibitor and the combination of these two agents augumented drug effect. . [Conclusion] It is critical to select appropriate treatment for individual cancer patient and for that, the development of system for accurate measurement of drug sensitivity in each sequence variant is urgent. Our strategy allows to see the effect of one-week treatment, which is necessary to determine the effect of PARP inhibitor and three mutants were confirmed as pathogenic variants. However, pathogenicity of the remaining mutants are not yet clear. Further improvement of sequence verification system is necessary to detect smaller functional difference and it is ideal to conduct the assay using isogenic human mutant cells.

3376T

SNP Variation in MicroRNAs Targeting the Tumor Suppressor Gene *PTEN*. *V.A. Ware, A. Jones, C. LaViolette, C. Corcoran, A. Maletz, J.A. Wilder.* Northern Arizona University, Flagstaff, AZ.

MicroRNAs play important roles in a broad range of biological processes and previous studies have demonstrated that SNP variation in microRNA sequences can affect pathways associated with cancer progression through diverse mechanisms. Here we describe sequence variation in microRNAs that target *PTEN* (Phosphatase and Tensin Homolog), an important tumor suppressor gene in a number of cancers including prostate, breast and lung cancer. We screened 41 Native American samples to find SNPs in microRNA encoding genes unique to Native American populations. A SNP (rs13136737) was found in the pri-miRNA of mir302D and mir367, which we are evaluating via quantitative PCR for potential allele-specific patterns of expression. Further evaluation of microRNA-encoding genes may uncover variation affecting cancer progression.

3377F

Missense Mutation of the Last Nucleotide of Exon 1, CDH1 c.48 G>C (Q16H), Contributes to Cancer Predisposition through Disruption of Normal Splicing and Generation of Missense Mutation. *L. Zhang[1], A. Xiao[2], J. Ruggeri[1], R. Bacares[1], S. Melo[3], J. Figueiredo[3], J. Simões-Correia[3], R. Seruca[3], M. Shah[4].* 1) Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY; 2) Northwestern University, Evanston, IL; 3) IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, 4200-465 Porto, Portugal; 4) Department of Medicine, Weill Cornell Medical College, New York, NY.

Purpose: The objective of this analysis was to assess the molecular mechanisms through which the CDH1 c.48 G>C (Q16H) missense mutation contributes to cancer predisposition. Mutation screening of CDH1 that contributes to the Hereditary Diffuse Gastric Cancer (HDGC) is important for patients with early-onset gastric cancer and/or with strong family history. Classification of variants in CDH1 is critical for cancer patient management. A particular challenge is the classification of rare non-truncating CDH1 sequence variants because it is not known whether these subtle changes can affect E-cadherin protein function sufficiently to predispose cells to cancer development. Patients and Methods: Our case is a 22 year old woman who was diagnosed with gastric cancer at the age of 18. Her grandmother, one great aunt and one great uncle were diagnosed with gastric cancer in their 30s and died of this disease in their 30s or 40s. The other great aunt was diagnosed with abdominal cancer at 16 and died subsequently. The CDH1 c.48G >C variant was identified in the patient and her living great uncle. RT-PCR products from the patient and her great uncle were cloned into the TOPO vector and subsequent sequencing was performed on individual clones from these two patients. In vitro functional studies were performed by transient transfection of CHO cells with the vectors encoding the WT and the mutant protein, as well as the empty vector (Mock), as a control. Upon transfection, cells were analysed for E-cadherin expression and localization and more importantly, for the two main functions of E-cadherin: cell-cell adhesion and invasion suppression. Results: Using RT-PCR and subsequent cloning strategy, we were able to detect a low level of mutant transcripts (5/27 clones had the 'C' allele, Q16H) in our patient. However, we were unable to detect any mutant transcripts in her great uncle (0/17 clone had the 'C' allele). In vitro functional studies of the Q16H missense mutation demonstrated it did not affect the expression, localization and cell-cell adhesion function of the E-cadherin. Instead, we observed loss of anti-invasion function in the E-cadherin protein with this substitution. Conclusion: These results support the conclusion that CDH1 c.48 G>C (Q16H) variant is a deleterious mutation and contributes to HDGC through disruption of normal splicing and inducing increased cell invasion.

3378W

The K-Type Human Endogenous Retroviral Element Encodes Fusogenic Activity in Melanoma Cells. *J. Dong[1], G. Huang[1], Z. Li[2], X. Wan[3], Y. Wang[4].* 1) Department of Pathology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX, USA; 2) Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Pathology, Peking University School of Oncology, Beijing Cancer Hospital & Institute, Beijing, China; 3) Department of Laboratory Medicine, Beijing Tongren Hospital, Capital Medical University, Beijing, China; 4) The Center for Medical Genetics, 7400 Fannin St, Suite #700, Houston, TX, USA.

Nuclear atypia with features of multinuclei have been detected in human melanoma specimens. We found that the K-type human endogenous retroviral element (HERV-K) is expressed in such cells. Since cellular syncytia can form when cells are infected with retroviruses, we hypothesized that HERV-K, which is expressed in melanoma cells and may play a role in melanomagenesis, contributes to the formation of multinuclear cells in melanoma. We specifically inhibited HERV-K using RNAi and monoclonal antibodies and observed dramatic reduction of intercellular fusion of cultured melanoma cells. Importantly, we identified loss of heterozygosity (LOH) of D19S433 in a cell clone that survived and proliferated after cell fusion. Our results support the notion that proteins encoded by HERV-K can mediate intercellular fusion of melanoma cells, which may generate multinuclear cells and drive the evolution of genetic changes that provide growth and survival advantages.

3379T

Detection of P16, Cyclin D1 and Bcl-2 Expression in Pancreatic Neoplasms. *S. Lai[1,2], X. Zhou[3].* 1) Pathology, Michael E. DeBakey VA Medical Center, Houston, TX; 2) Baylor College of Medicine, Houston, TX; 3) Internal Medicine, University of Texas Medical School at Houston, Houston, TX.

P16 and cyclin D1 genes play a critical role in the regulation of the G1-S transition of the cell cycle, and their expression are frequently altered in several neoplastic entities. Bcl-2 gene in apoptosis pathway has important function in tumorigenesis of cancers. Analysis of the protein products of these genes by immunohistochemical methods provides information on their functional status and allows for their phenotypic evaluation of tumor cells. Studies of the correlation between these gene expression levels and clinicopathologic characteristics of pancreatic neoplasms are not ample. We constructed a tissue microarray block from 24 pancreatic neoplasms including 17 adenocarcinoma, 6 neuroendocrine neoplasm and 1 intraductal papillary mucinous neoplasm, and stained for P16, cyclin D1 and Bcl-2 by immunohistochemistry. Protein expressions were correlated with tumor histologic type, grade, perineural invasion, lymphovascular invasion and lymph node status. Loss of P16 expression was correlated with malignancy ($p<0.001$), and lymphovascular invasion ($p=0.032$). Cyclin D1 overexpression was also associated with malignancy ($p=0.035$). Our study indicates the absence of p16 in most of the malignant neoplasms are associated with an early tumorigenic event. High P16 expression level in pancreatic tumors predicts a better prognosis. P16 and cyclin D1 immunoreactivity may be used as a prognostic marker for pancreatic neoplasms.

3380F

Exome-based method to determine cancer tissue of origin. *K. Robasky[1], E. Aronesty[1,2], W.D. Jones[1,3].* 1) Expression Analysis a Quintiles Company, Durham, NC; 2) Bioinformatics Program, Johns Hopkins University, Washington D.C; 3) Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC.

Identifying the source of newly presenting tumor fails to yield definitive results in a significant number of cases, with cancer of unknown primary organ (CUP) accounting for 3-5% of all cancer diagnoses. Tumors are commonly classified via visual inspection alongside of immunohistochemistry and expression profiling, neither of which can consistently identify a molecular signature by which to guide treatment. Misclassifying a cancer's primary site can have wide-ranging effects, including: 1) masking the true primary cancer, 2) misinforming the patient treatment plan 3) obfuscating clinical trial results by enrolling misclassified subjects. Here we present a method for using cost-effective, high-throughput sequencing exome data to find the molecular signatures of various tissue-specific cancers. We do so by first scoring the genes from individual exomes based upon the level of mutational burden. We then perform an unsupervised analysis to reduce the dimensionality of the scored genes. Using exomes from The Cancer Genome Atlas, we apply this method to classify cancers into tissue-specific clusters. This model can thus be used with new tumor exomes to aid in identifying the tissue-of-origin of the primary cancer.

3381W

A novel statistic method for drug response prediction with big RNA-seq data. *M. Xiong[1], L. Ma[1], M. Chen[1], S. Guo[2,1].* 1) Dept Biostatistics, Univ Texas Hlth Sci, Houston, TX; 2) Fudan University, Shanghai, China.

Digital transcriptome analysis by next-generation sequencing discovers substantial mRNA variants. Variation in gene expression underlies many biological processes and holds a key to unraveling mechanism of common diseases and drug response. However, the current methods for drug response prediction using overall gene-expression are originally designed for microarray expression data and overlook a large number of variations in gene expressions. The challenge for application of gene expression to pharmacogenomics is how to accurately measure variation in gene expression and how to develop statistical methods to fully explore information on expression variation for predicting response of a patient to a particular treatment regimen with extremely big RNA-seq data. To fully utilize expression information at genomic positional level we use sufficient dimension reduction (SDR) techniques which project the original high dimensional data to very low dimensional space while preserving all information on response phenotypes, functional principal component analysis which further reduce the dimension of the big RNA-seq data, and penalization methods for optimal feature selection as a powerful tool to develop novel statistical methods for classifying drug response of patients with RNA-seq data. The proposed methods are applied to ovarian cancer drug response RNA-seq and microarray expression data from TCGA database where expressions of 163 sensitive and 70 resistant to drug ovarian cancer samples were measured by RNA-seq and microarray. Tenfold cross validation was used to evaluate the performance of classifier. We can reach almost 100%, 100% and 100% average classification accuracy, sensitivity and specificity, respectively, in the test datasets with RNA-seq data, and 63.54%, 88.85% and 10.43% average classification accuracy, sensitivity and specificity, respectively, in the test datasets with microarray-measured expression dataset. Our results strongly demonstrate that the prediction based on RNA-seq substantially outperforms the prediction based on the microarray expressions and open a new avenue for cancer pharmacogenomics studies.

3382T

Leukemia relapse in donor cells ten years after allogeneic hematopoietic stem cell transplantation for acute myeloid leukemia. *G. Calabrese[1,2], L. Militti[1], R. DiGianfilippo[2], M. Alfonsi[1], P. Guanciali-Franchi[1], D. Fantasia[1], P. Bavaro[3], P. DiBartolomeo[3], G. Palka[1,2].* 1) Dept S Med, Oral & Biotec, Univ Chieti, Chieti Scalo, Italy; 2) Medical Genetics Dept, Pescara Hospital, Pescara, Italy; 3) Bone Marrow Transplantation Center, Pescara Hospital, Pescara.

Leukemic relapses after allogeneic hematopoietic stem cell transplantation (HSCT) in most instances arise from the original recipient leukemia clone. Exceptionally, leukemic relapse can originate from the donor hematopoietic cells. A recent review listed 65 cases of donor cell leukemia, with an estimated risk to be about 12 per 10,000 HSCT. Exposure to radiant therapy and to chemotherapy are regarded as involved in donor cell leukemogenesis, though causes are still unclear. In 2000, a female patient, 42-yrs-old, was diagnosed with acute myeloid leukemia (AML) M2 subtype having a diploid karyotype. After achieving remission by induction chemotherapy and total body radiotherapy the patient underwent a HSCT with cells from her brother. Six months later the patient showed a complete chimerism with a normal male karyotype which persisted for nine years. In 2010 the patient was treated with metabolic radiotherapy for a thyroid carcinoma. Six months later showed pancytopenia and FISH analysis with X/Y chromosome probes revealed a partial chimerism (0.08% recipient cells). Six months after loss of complete chimerism the patient showed clinical-hematologic features of leukemic relapse. Cytogenetic analysis performed on bone marrow aspirate revealed an abnormal male karyotype with chromosomes 8, and 13 clonal anomalies, and including a marker chromosome in all cells analyzed. This result was confirmed by FISH. The patient deceased two weeks later. In the present case AML occurred in donor cells ten years after HSCT with a phenotype overlapping patient original disease. Although rarely reported, the chromosomal anomalies in donor cells were specific of therapy-related AML. Radiotherapy for thyroid carcinoma and possible immune system disturbance might be associated with leukemogenesis. Chimerism study with the finding of clonal chromosomal anomalies confirm cytogenetic analysis role as of relevance for both HSCT follow up surveillance and investigation of biological mechanism of leukemogenesis.

3383F

Photodynamic therapy effect in EGF pathway gene expression in glioblastoma as a biological model. *L.B. de Paula[1,2], F.L. Primo[1,3], N.T.A. Peres[2,4], N.M. Martinez-Rossi[2,4], A. Rossi[2,4], A.C. Tedesco[1,3], Supported by CNPQ, FAPESP, FAEPA.* 1) Center of Nanotechnology and Tissue Engineers - Photobiology and Photomedicine research Group, School of Philosophy, Sciences and Letters of Ribeirao Preto, University of Sao Paulo, Sao Paulo, Ribeirao Preto, Brazil; 2) Department of Genetics, Medical School of Ribeirao Preto, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Ribeirao Preto, Brazil; 3) Department of Chemistry, School of Philosophy, Sciences and Letters of Ribeirao Preto, University of Sao Paulo, Sao Paulo, Ribeirao Preto, Brazil; 4) Laboratory of Laboratory of Genetics and Molecular Biology of Fungi, Medical School of Ribeirao Preto and University of Sao Paulo, Sao Paulo, Ribeirao Preto, Brazil.

Glioblastoma multiforme (GBM), the most malignant human brain tumor, may develop de novo or through progression from low-grade or anaplastic astrocytoma. The deadly nature of GBM originates from explosive growth and invasive behavior, which are fueled by dysregulation of multiple signaling pathways. Dysregulated epidermal growth factor receptor (EGFR) pathway is frequently in high-grade gliomas via gene amplification or dominant-active mutation, is a leading cause of gliomagenesis. New methods to achieve widespread distribution of therapeutic agents have shown a significant improvement in brain tumor therapy. Photodynamic therapy (PDT) has been used mainly as an anticancer therapy, which relies on the absorption and retention of a photosensitizer molecule in the tumor cells associated with further irradiation with an appropriate visible light. We investigated whether similar changes in gene expression EGF pathway in GBM cell lines pre and post treatment with PDT could be detected and correlating the incorporation of this nanotechnology as an alternative treatment for drug-resistant tumors.Gene expression EGF pathway resulted in hipoexpression EGFR, SOS1, SOS2, STAT1, PIK3R1, PIK3C2A, PIK3CA, PIK3CB, MAP2K4, MAP3K1, MAPK1, MAPK10, and MAPK8 MAPK9 in the post-PDT. The complex EGFR-TK (tyrosine kinase) has key role in many processes that affect tumor growth and progression, including proliferation, dedifferentiation, apoptosis inhibition, invasiveness and loss of dependence on adhesion. Phosphorylation of tyrosine kinase residues in EGFR functions as binding domain Grb2/Sos complex, which activates the signaling cascade Ras/Raf/ MAPK pathway, which influences the proliferation, migration and differentiation. Members of the MAPK pathway have been associated with various functions within the process of tumorigenesis, since this pathway is responsible for the regulation of proliferation, differentiation, survival, angiogenesis and metastasis in many tumor types. These findings confirmed that the engineering of nanocarriers associated with PDT procedures led to hipoexpression of the genes that are directly involved in the tumor process the EGF pathway in vitro by advanced protocols that can be useful for future in vivo trials available to clinical oncology. With the development of protocols associating photodynamic therapy and cancer, we aim to develop advanced treatments with nanotechnological tools for highly resistant diseases such as human gliomas.

3384W

Elevated expression of maspin mRNA as a predictor of survival for stage II Gallbladder Cancer. *K. Baghel[1,2], HR. Kazmi[1], S. Raj[1], A. Chandra[1], RN. Srivastava[2].* 1) Department of Surgical Gastroenterology, King George's Medical University, Lucknow, Uttar Pradesh, India; 2) Department of Orthopaedic Surgery, King George's Medical University, Lucknow, Uttar Pradesh, India.

Introduction: Prognostic significance of Maspin gene has reported in various cancers but has not been evaluated for Gallbladder Cancer (GBC). In present study, we investigated maspin mRNA expression in normal, cholelithiasis and GBC patients along with its prognostic importance in cancer patients. Methodology: Gallbladder tissue from normal (n=25), cholelithiasis (n=25) and cancer patients (n=31) were analysed for maspin mRNA expression by semi-quantitative reverse transcriptase PCR and quantitative real time PCR. Statistical analysis was carried out using student t test and ANOVA. Kaplan-Meier method was used for survival analysis and the difference between survival curves was analyzed by log-rank test. Results: Significant increased(P<0.0001) expression of maspin mRNA was observed in GBC as compared to cholelithiasis, whereas no expression was found in normal tissues. Significant correlation (Pearson's coefficient(r)=-0.829;P<0.0001) was observed between relative quantification of maspin mRNA and survival of cancer patients after surgery. The Kaplan-Meier survival curves demonstrated that overall survival was significantly shorter(P=0.0002) in GBC patients with RQ >1.5 as compared to patients with RQ≤1.5. Patients with RQ >1.5 had a hazard ratio [HR] of 5.527 for death as compared to patients with RQ≤1.5 (95% Confidence interval [CI]=2.223-13.75). Statistically significant difference(P=0.0371) in overall survival was observed for stage II patients with RQ >1.5 in comparison to patients with RQ≤1.5. For stage III, there was no significant difference(P=0.126) in overall survival between patients with R.Q.>1.5 and R.Q.≤1.5. The HR for stage II and III disease was 5.236(95% CI =1.104-24.83) and 4.608(95% CI = 0.6507 - 32.63) respectively. Conclusion: Higher expression of maspin mRNA in gallbladder cancer has prognostic significance, especially for resected stage II cancer, which needs to be investigated further.

3385T

Prognostic signature in papillary thyroid carcinoma patients that undergone total thyroid ablation. *M.C. Barros Filho[1], F.A. Marchi[2], C.A. Pinto[1], S.R. Rogatto[1,3], L.P. Kowalski[1].* 1) AC Camargo Cancer Center, Sao Paulo, Brazil; 2) Sao Paulo University, Sao Paulo, Brazil; 3) Faculty of Medicine/UNESP, Botucatu, SP, Brazil.

Purpose. Papillary thyroid carcinoma (PTC) is the most common endocrine malignancy. Despite of the low mortality rate, recurrence is frequent and associated with patient morbidity. Currently, clinical and pathological procedures are insufficient to determine the risk of relapse. Although *BRAF* mutation is often associated with tumor aggressiveness, no molecular marker is used in clinical routine. The aim of this study is to identify molecular markers able to predict recurrence in PTC patients. **Patients and Methods.** Eighty-three patients submitted to total thyroid ablation were retrospectively enrolled. Primary fresh-frozen tumors (RNA integrity score> 7) from 58 cases were evaluated by expression oligoarray using Sure Print G3 8x60K slides (Agilent Technologies). Cases without suspicion of active disease (normal imaging screening and serum Tg) with at least five years of follow up (group 1= 46) were statistically compared with cases with confirmed recurrent disease (group 2= 12) by SAM (FDR< 20%) and t test (P< 0.005). Six transcripts were selected for further RT-qPCR confirmation in a technical validation group (samples previously evaluated by microarray) and in an independent validation group (group 1= 15 and group 2= 10) **Results.** By expression oligoarray, two genes were underexpressed (*FBXL5* and *F2RL2*) and 13 overexpressed (*FOXP2, SLC2A4, GADD45B, C6ORF223, LYSMD3, UTP23, C10ORF79, NCRNA00203, FLJ45950, S1PR1, DNTT, MBD2* and *PROM2*) in the recurrence group. Four out of six transcripts evaluated by RT-qPCR were highly correlated with microarray data (Pearson correlation test, r> 0.75 and P< 0.001). *GADD45B* remained associated with recurrence in the validation test (log rank, P< 0.004). A multivariable analysis identified *GADD45B* as an independent marker of disease free survival (Cox regression, P= 0.015). Compared to well-established clinical parameters, as lymph node involvement and extrathyroidal extension, *GADD45B* presented superior value in predicting recurrence (AUC= 0.741; CI 95%= 0.615-0.868). **Conclusion.** *GADD45B* was identified as an independent marker of poor outcome and can be useful to better stratify PTC patients according to the risk of recurrence.

3386F

LOX-1, A Novel Metabolic Target in Human Breast and Colon Cancer. *M. Biancolella, M.J. Zonetti, T. Fisco, P. Mazzarelli, F. Sangiuolo, L. Saieva, M. Murdocca, G. Novelli, S. Pucci.* Dept Biomedicine and Prevention, Univ Rome Tor Vergata, Rome, Italy.

Targeting the metabolic pathways of cancer is a hot topic for drug discovery. An emerging theme in cancer biology is that a number of lipid metabolic genes, consistently over-expressed in human cancers, are critical to cellular transformation and in maintaining the transformed state. OLR1 gene, encoding the cell membrane receptor LOX-1 (lectin-like oxidized low density lipoprotein receptor), emerged as one of the most important for maintaining the malignant state. LOX-1 may have at least two independent pro-oncogenic mechanisms of action, one based on activation of NF-kB signaling pathway and the other as regulator of lipogenesis.. We observed that LOX-1 is overexpressed in human breast and colorectal carcinomas, at different stages of disease, as compared to normal tissues. Moreover, we demonstrated that the metabolic oncogene FASN (fatty acid synthase), is overexpressed in the breast and colon neoplastic cells. Silencing ORL1 gene in a colon cancer cell line (DLD1), we observed a strong down regulation of FASN and an evident inhibition of the secreted isoform of Clusterin (sCLU) a cytoprotective protein, involved in tumor progression. Conversely, a strong up regulation of the nuclear clusterin isoform (nCLU), was observed. These data suggest a role of LOX-1 in the cancer insurgence and progression. Therefore, this protein could be considered a novel target for cancer therapy.

3387W

Increased frequencies of GSTM1 and GSTT1 null genotypes in Indian patients with Leukemia.
S. Caplash[1], S. Kaur[1], R. Arora[2]. 1) Human Genetics, Punjabi University, Patiala, Punjab, India; 2) Oswal Cancer Hospital, Ludhiana.

Glutathione S-transferases (GST) isoenzymes play a significant role in phase II biotransformation and detoxification of many xenobiotics including environmental carcinogens, pollutants and drugs. Both the genetic polymorphisms and expression pattern of GST genes may have a major impact on cancer susceptibility, inter-individual variability in the prognosis, drug effects and toxicity. Of these, GSTT1 and GSTM1 isoenzymes are highly polymorphic with homozygous deletion of either or both genes resulting in absence of enzyme activity. We carried out a trial case-control study involving 142 Leukemia patients and 60 normal healthy controls from Punjab (North India). Multiplex PCR was carried out to determine GSTM1 and GSTT1 polymorphism. The frequency of individuals carrying GSTM1 and GSTT1 null genotypes was higher among Leukemia patients (42 % and 23 %) as compared to the control group (38 % and 17 %), but the difference found was not statistically significant (p>0.05). Follow up of patients is in process to analyze the association of GST genotypes with differential chemotherapy drug response.

3388T

Two different *BRCA2* mutations found in a multigenerational family with a history of breast, prostate, and lung cancers.
D. Caporale, E. Swenson. Biology, University of Wisconsin-Stevens Point, Stevens Point, WI.

Breast and lung cancer are two of the most common malignancies in the United States, causing approximately 40,000 and 160,000 deaths each year, respectively. Approximately 5 to 10 percent of breast cancer cases are hereditary, with about 84 percent of those cases due to mutations in two large breast cancer predisposition genes, *BRCA1* and *BRCA2*. These are normally tumor-suppressor genes associated with DNA repair. Since the discovery of these two genes in 1994 and 1995, respectively, several other breast cancer predisposition genes have been identified, including the *CHEK2* gene, which is another tumor-suppressor gene. Recently, studies have begun investigating the roles of *BRCA1* and *BRCA2* in lung cancer. We conducted two case studies, one on an extended family of Italian heritage with several cases of breast cancer and associated cancers through multiple generations, and another on a non-blood relative of Scottish/Irish descent who was consecutively diagnosed with breast and lung cancer. Cancer history and environmental risk factors were recorded for each family member. To investigate possible genetic risks, we screened for mutations in specific hypervariable regions of the *BRCA1*, *BRCA2*, and *CHEK2* genes. DNA was extracted and isolated from the individuals' hair follicles. PCR, allele-specific PCR (AS-PCR) and DNA sequencing were performed to identify and verify the presence or absence of mutations in these regions. Genotypes of each family member were determined and carriers of mutations were identified. Specifically, three Italian family members were found carriers of the *BRCA2-3036del4* mutation, a 4-nucleotide deletion in an exon, which is a nonsense mutation that causes a frameshift in the genetic code, rendering the BRCA2 product nonfunctional. The individual with breast and lung cancer was not a carrier of this mutation, but rather a carrier of the *BRCA2-6503delTT*, which is also a nonsense mutation but more common in the Irish heritage.

3389F

Negotiating a Minefield: Which Variants to Return in a Large Prospective Whole Exome Sequencing Project?
A. Church[1,2], E. Van Allen[3,4,6], E. Hiller[3], I. Rainville[3], H. Rana[3,4], D. Treacy[3], K. Karalis[2,5], F. Huang[3,4,6], M. Giannakis[3,4,6], F. Wilson[3,4,6], E. Stover[3,4,6], J. Bohkari[1], L. Sholl[1], N. Lindeman[1], J. Garber[3,4], N. Wagle[3,4,6], L. Garraway[3,4,6]. 1) Pathology, Brigham and Women's Hospital, Boston, MA; 2) Pathology, Boston Children's Hospital, Boston, MA; 3) Medical Oncology, Dana Farber Cancer Institute, Boston, MA; 4) Medicine, Dana Farber Cancer Institute, Boston MA; 5) Wyss Institute, Boston, MA; 6) Broad Institute, Boston, MA.

The CanSeq study is a prospective whole exome sequencing project currently enrolling patients at the Dana Farber Cancer Institute and Brigham and Women's Hospital. Whole exome sequencing is performed at the Broad Institute on both somatic cancer tissue as well as germline DNA derived from blood. As a large collaborative effort, the aims of the project include identification of somatic and germline alterations, the identification of potentially actionable and consequential variants, and to prioritize events for treatment. Enrolled patients, who have undertaken informed consent, receive curated results of both somatic and germline sequencing.

Herein we describe the process we have undertaken to identify genes and variants of clinical significance to return to patients and their treating physicians. Lists of genes include somatic (122 genes), germline risk and carrier (149 cancer and 35 non-cancer genes), pharmacogenomics (70 cancer and 18 non-cancer genes). Our process has evolved over years and incorporates available literature, committee guidelines including the recent ACMG recommendations, ethical considerations such as beneficence, non-maleficence and justice, and the consensus of our expert collaborators.

3390W

Analysis of *BRAF* mutations (*BRAF* ^{V600E}) in gliomas.
L. DA SILVA[1], P. VIDIGAL[2], N. BINDA[1], P. COUTO[1], L. BASTOS-RODRIGUES[1], L. DE MARCO[1]. 1) INCT of Molecular Medicine (INCT-MM), Faculty of Medicine, Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG, Brazil; 2) Faculty of Medicine, Universidade Federal de Minas Gerais (UFMG) - Belo Horizonte, MG, Brazil.

Background: Gliomas are the most common tumors of the Central Nervous System and are classified to grades I-IV according to the WHO. Despite remarkable progress in characterizing the molecular pathogenesis of gliomas, these tumors remain incurable and, in most cases, refractory to treatment, due to its molecular heterogeneity. To improve the survival rate of gliomas patients, a better understanding of tumor biology is required as well as the subsequent development of new therapeutic strategies. *BRAF* represents one of the most frequently mutated protein kinase genes in human tumors. *BRAF* is a key mediator of the RAS-RAF-MAP kinase signaling pathway and is involved in a wide variety of cellular functions, including cell proliferation, cell cycle arrest, terminal differentiation and apoptosis. Ninety percent of the reported mutations of *BRAF* are the somatic missense mutation V600E. *BRAF*^{V600E} are present in roughly 25% of low grade gliomas but the overall impact of these mutations on clinical outcome in gliomas remains unclear. Methods: In this study, we investigated the prevalence and histopathological grade of brain tumors and correlated with the incidence of mutations in *BRAF*. The target regions of exon 15 of the *BRAF* gene were amplified using specific primers, purified and sequenced in 30 fresh tumor samples. Using immunohistochemistry, we analyzed 38 FFPE brain tumors, including 14 gliomas (grade I), 11 (grade II), 3 (grade III) and 11 glioblastomas (grade IV). Statistical analyses were made by Student's t test. Results and conclusions: In our sample, no *BRAF*^{V600E} mutations were found. The cytoplasmic *BRAF* expression was higher in neoplastic cells compared to benign cells, suggesting that the higher the grade, the greater is the immunoreactivity of *BRAF*. Several studies have shown genetic alterations in *BRAF*, mainly *BRAF*^{V600E}, in low but not in high-grade brain tumors. We found an increased expression of *BRAF* in high-grade gliomas compared to low-grade. Although we unexpectedly did not find *BRAF* mutations in any of the fresh tissue we analyzed, the increased expression suggests that other proteins and/or transcription factors-related pathways may be altered.

3391T

MtDNA copy number alteration and association with clinicopathological features in Breast Cancer patient. *M. Ghaffarpour, M. Houshmand.* Medical Genetics, National Institute of Genetic Engineering and Biotechnology, Tehran, Tehran, Iran.

Large-scale depletion in mitochondrial DNA (mtDNA) has been extensively detected in various human cancers. However, it still remains unclear whether the alterations in mtDNA content are related to the clinicopathological parameters and patient with breast cancer. To explore the role of mtDNA copy number in breast cancer etiology and association between clinicopathological features, in the present study, we performed the copy number of mtDNA in 59 cases of breast tumors and paired non-tumor tissues using quantitative real-time PCR. Our data showed that the level of mtDNA was significantly increased in tumor tissues in compared to the adjacent non-tumor counterparts (P<0.05). The increased copy number in mtDNA was not significant association with clinicopathological features such as age ,grade,stage, lymph node involvement, ER, PR and Her-2/neu expression.However, there was a significant association between decreased overall Survival (5yr %)and increased copy number in mtDNA(p=0.049). Together, our results suggested that high copy number of mtDNA may be involved in breast neoplastic transformation and mtDNA content might be potentially used as a tool to prognosis factor with the other risk factors.

3392F

TNFA-308 G>A is associated with HER-2/neu expression in women with breast cancer. *L. Gomez-Flores-Ramos[1,2], A. Escoto-De Dios[1], A. Ramos-Silva[1,3], R. Ramirez-Patiño[1,2], I. Gutierrez-Hurtado[1,2], MP. Gallegos-Arreola[1].* 1) Laborotorio de Genetica Molecular, CIBO, IMSS, Guadalajara, Mexico; 2) Doctorado en Genetica Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara; 3) Doctorado en Farmacologia, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara.

Introduction. The tumor necrosis factor-alpha gene (*TNFA*) plays an important role in cell proliferation, differentiation, apoptosis, lipid metabolism, coagulation, insulin resistance, and endothelial function. *In vitro* and *in vivo* experiments have shown the tumorigenic effects of TNF, and elevated plasma levels of TNF have been associated with a poor prognosis in Breast Cancer (BC). The TNFA-308G>A (rs1800629) polymorphism has an effect on the gene expression, increasing the production of TNF protein. The aim. To determine the association between the TNFA-308G>A in Mexican women with BC with Her-2/neu expression. **Methods.** Blood samples were collected after a written informed consent approved by the ethical committee from 371 controls and 465 patients with clinical and histological confirmation of BC living in Guadalajara, Mexico. Genotyping was performed by RFLPs. The genotype frequencies at the control group were in Hardy-Weinberg equilibrium. Statistical analyses were performed using PASW Statistic Base 18 software, 2009 (Chicago, IL). **Results and discussion.** Our results show an association between the genotypes GA-AA and the expression of Her-2/neu in BC cells (OR 1.6, 95% CI 1.06-2.4, p=0.025). The patients with the genotypes GA-AA at tumor stages I-II were also associated with Her-2/neu positivity (OR 2.5, 95% CI 1.31-4.8, p=0.004) as well as patients with BMI >30>40 and Her-2/neu (OR 2.8, 95% CI 1.2-6.6, p = 0.016). In BC patients the presence of HER-2/neu has been correlated with metastases, insulin resistance and overexpression of fatty acid synthase, and its expression is regulated by metformin in *in vitro* models. HER-2/neu serum concentrations in patients with diabetes type 2 were associated with serum TNF receptor 1, thus TNF might be behind concomitant insulin resistance and EGF resistance, leading to increased circulating HER-2/neu levels. Although the mechanisms of how the overexpression of HER-2/neu oncogene induces resistance to TNF is unknown; different studies have been observed their overexpression is associated with poor prognosis in BC, because it induces metastatic potential and resistance in cancer cells. Another mechanisms proposed is that the overexpression of Her-2/neu is mediated by signaling pathways and cytotoxicity, suggesting that the overexpression of TNF receptor-1 is important in TNF sensitivity in HER-2/neu-overexpressing cancer cells.

3393W

The anti-cancer activity of propolis in colorectal cancer. *M. Gunduz[1,2], G. Nas[1], M. Acar[1], O.F. Hatipoglu[1], B. Yilmaz[1], G. Kaya[1], E. Gunduz[1].* 1) Medical Genetics, Turgut Ozal University Medical Faculty, Ankara, Turkey; 2) Department of Otolaryngology Head and Neck Surgery, Ankara, Turkey.

Colorectal cancer is the third most commonly diagnosed cancer in the world. It originates from the epithelial cells lining the colon or rectum. Colorectal cancer is commonly observed in developed countries and is the second deadliest type of cancer. . Colorectal tumorigenesis proceeds through an accumulation of specific molecular alterations and one of the most important of these mechanisms is inhibition of apoptosis. The Inhibitors of Apoptosis (IAP) are a family of functionally and structurally related proteins, which serve as endogenous inhibitors of programmed cell death (apoptosis). Survivin is a member of the IAP family and functions as an inhibitor of caspase activation, thereby leading to negative regulation of apoptosis. This has been shown by disruption of Survivin induction pathways leading to increase in apoptosis and decrease in tumor growth. Propolis is a resinous mixture that honey bees collect from tree buds, sap flows, or other botanical sources. It is used as a sealant for unwanted open spaces in the hive and possesses anti-microbial, anti-oxidative, anti-ulcer and anti-tumor activities. The survivin protein is expressed highly in most human tumors but is completely absent in terminally differentiated cells. Our study investigated the effect of propolis onHT29 colorectal cancer cells. Propolis exhibits anticancer activity in a dose dependent manner by inhibition of survivin gene expression. Decreased Survivin expression results in a decrease in tumor growth and prevents carcinogenesis by inducing programmed cell death. Consequently, Propolis may become a strong candidate molecular agent for cancer therapy.

3394T

Identification of Complete Hydatidiform Mole Pregnancy-associated MicroRNAs in Plasma. *Y. Hasegawa[1], K. Miura[1], A. Higashijima[1], S. Miura[1], J. Tsukamoto[1], S. Abe[1], A. Kinoshita[2], H. Mishima[2], K. Yoshiura[2], H. Masuzaki[1].* 1) Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 2) Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, , Nagasaki, Japan.

Objective: The aim of this study was to identify the complete hydatidiform mole (CHM) pregnancy-associated microRNAs in plasma. Methods: First, by comparative analysis of next generation sequencing-generated microRNAs expression profiles, we selected the candidate CHM pregnancy-associated microRNAs in plasma, which showed a higher expression in CHM tissues than in normal villous tissues, but no expression in blood cells. Then, expression levels of these microRNAs in CHM tissues (n:14) and normal villous tissues (n:20) were investigated to identify CHM-associated microRNAs with increased expressions in CHM tissues than in normal villous tissues. Subsequently, when the plasma concentrations of cell-free CHM-associated microRNAs were significantly higher in CHM pregnant women (n:14) than in uncomplicated pregnant women (n:20), these microRNAs were identified as CHM-associated microRNAs in plasma. Finally, to identify CHM pregnancy-associated microRNAs in plasma, circulating levels of CHM-associated microRNAs in plasma samples from CHM pregnant women (n:14) were measured before and after evacuation. Results: First, hsa-miR-520b, hsa-miR-520f and hsa-miR-520c-3p were selected as candidate CHM pregnancy-associated microRNAs in plasma, and then all of them were confirmed as CHM-associated microRNAs in tissue samples. Subsequently, all CHM-associated microRNAs were identified as CHM-associated microRNAs in plasma. Finally, the circulating levels of all CHM-associated microRNAs in plasma were decreased significantly after evacuation (Wilcoxon signed rank test, p=0.001), suggesting that all of these microRNAs were pregnancy-associated molecules. Conclusions: As CHM pregnancy-associated microRNAs in plasma, hsa-miR-520b, hsa-miR-52Of and hsa-miR-520c-3p were identified. Circulating levels of three CHM pregnancy-associated microRNAs in plasma can be potential molecular marker for CHM pregnancy.

3395F

Eukaryotic translation initiation factor 4E (eIF4E) expression is associated with breast cancer tumor phenotype and predicts survival after anthracycline chemotherapy treatment. *T. Heikkinen[1], T. Korpela[2], R. Fagerholm[1], S. Khan[1], K. Aittomäki[3], P. Heikkilä[4], C. Blomqvist[5], O. Carpén[2, 6], H. Nevanlinna[1].* 1) Department of obstetrics and gynecology, Helsinki university central hospital, Helsinki, Finland; 2) Department of pathology and Medicity laboratory, University of Turku, Turku Finland; 3) Department of medical genetics, Helsinki university central hospital, Helsinki, Finland; 4) Department of pathology, Helsinki university central hospital, Helsinki, Finland; 5) Department of oncology, Helsinki university central hospital, Helsinki, Finland; 6) Turku university hospital, Turku Finland.

Abnormal translation of mRNAs, which frequently occurs during carcinogenesis, is among the mechanisms that affect the expression of proteins involved in tumor development and progression. Eukaryotic initiation factor eIF4E is a key regulator of translation of many cancer-related transcripts, including VEGF, Cyclin D1, and MMP-9. It is over-expressed in many cancers and has been associated with worse survival of cancer patients. The protein levels of eIF4E were determined with immunohistochemistry (IHC) in 1,233 paraffin-embedded breast tumors on tissue microarrays. The effects of the IHC expression level were analyzed on tumor characteristics and patient survival. The survival analyses were also stratified by tumor properties and adjuvant chemotherapy treatment. A number of 1,085 tumors were successfully stained. A high level of eIF4E protein expression was associated with several characteristics of aggressive breast cancer, namely grade, estrogen and progesterone receptor negativity, HER2 receptor positivity and high expression of p53 and Ki67 (p<0.001). High eIF4E expression was also associated with worse breast cancer-specific survival with a hazard ratio (HR) of 1.99 (95% CI:1.32-3.00, p=0.0008). In multivariate analysis high expression of eIF4E was found to be an independent prognostic factor. High eIF4E expression was also associated with worse survival after detection of distant metastasis (HR=1.88, 95% CI:1.20-2.94, p=0.0060). In the stratified survival analyses the survival effect was strongest among patients treated with anthracycline chemotherapy (HR=3.34, 95% CI:1.72-6.48, p= 0.0002), whereas no effect was seen among patients who had not received anthracycline with significant difference in heterogeneity between the two groups (p=0.0358). Our results confirm that high expression of eIF4E is associated with adverse tumor characteristics and with poor prognosis of breast cancer patients. The strongest survival effect was seen in patients treated with anthracycline-based chemotherapy, suggesting eIF4E to have predictive value as a biomarker. These results highlight the importance of the regulation of mRNA translation in the biology of breast cancer and emphasize the potential of eIF4E as a therapeutic target. Further studies are required to validate eIF4Eas a treatment predictive factor in breast cancer.

3396W

Personalized Genomics of Metformin Therapy for Improved Cancer Survival. *C.C. Iverson[1,2], H. Xu[3], Q. Chen[4,5], A. Shah[5], Q. Dai[6], J. Warner[5,7], N.B. Peterson[7], L. Olson[1], D.C. Crawford[1], D.M. Roden[7,8], J.C. Denny[5,7], M.C. Aldrich[1,2,6].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University, Nashville, TN; 3) The University of Texas School of Biomedical Informatics at Houston, Houston, TX; 4) Department of Biostatistics, Vanderbilt University, Nashville, TN; 5) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 6) Division of Epidemiology, Vanderbilt University, Nashville, TN; 7) Department of Medicine, Vanderbilt University, Nashville, TN; 8) Department of Pharmacology, Vanderbilt University, Nashville, TN.

Increasing evidence supports a role for metformin as a therapeutic for improved cancer survival. To determine the potential use of metformin in a personalized genomic medicine approach for the treatment of cancer, we performed a genome-wide association study of overall survival in cancer patients. Using the Vanderbilt University large-scale biorepository linked to electronic medical records, we identified 1,106 individuals (674 with type 2 diabetes (T2D), 432 non-diabetics) of European ancestry diagnosed with cancer between 1995 and 2010. Among cancer patients with T2D, 461 individuals were on metformin therapy and 213 were treated with other T2D medications. Patients were followed a median of 2.3 years (IQR: 1.3-3.8) for overall mortality or loss to follow-up. Participants were genotyped using either the HumanOmni5-Quad or Illumina OMNI-Quad genotyping platforms. We estimated hazard ratios (HRs) and accompanying 95% confidence intervals (CI) for overall mortality using Cox proportional hazard models, adjusted for age and sex, to evaluate associations between genetic variants and mortality, using an additive model. Genomic control was applied to correct for potential population stratification occurring in the study. Among cancer patients who were metformin users, we identified one marker, rs153047, near genome-wide significant and associated with a two-fold higher mortality (HR=2.22, 95% CI: 1.91-2.54, $p=5.6 \times 10^{-7}$). This SNP was not associated with mortality in T2D cancer patients on other drugs or non-diabetic cancer patients (p>0.20 for all groups). Assessment of gene-environment interactions demonstrated that the association with rs153047 may vary by metformin use (p=0.15). Among individuals homozygous for the reference allele, metformin use was associated with a significantly reduced mortality compared to non-metformin use (HR=0.58, 95% CI: 0.25-0.92, p=0.0023). Metformin use was not significantly associated with improved survival for individuals with the variant allele (HR=0.98, 95% CI: 0.47-1.49, p=0.95). This marker is located within *SNTB2* on chromosome 16q22.1, which has been previously shown to be associated with invasive breast cancer. Additional SNPs on chromosomes 3, 10, 12 13, and 16 had $p<10^{-6}$ and deserve further investigation. Our findings suggest the survival benefit of metformin therapy may be modulated by specific genomic loci. If replicated, this finding may be useful to target metformin therapy in cancer patients.

3397T

Comparison of mRNA expression profiles in familial and sporadic breast cancers in Finnish population. *S. Khan[1], P. Heikkilä[2], K. Aittomäki[3], C. Blomqvist[4], H. Nevanlinna[1].* 1) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 2) Department of Pathology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 4) Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland.

Although majority of breast cancers initiate from somatic alterations, familial aggregation and twin studies have suggested that hereditary predisposing factors are involved in up to one third of all breast cancers. The underlying inherited mutations in breast cancer families may also affect the progression and mRNA expression profiles of the tumors. In order to investigate the differences in mRNA expression profiles of familial and sporadic tumors, we performed a comparative study of mRNA levels in 211 familial and sporadic tumors from Southern Finland. Altogether we had mRNA expression data (Illumina HumanHT-12 platform) from 139 sporadic and 72 familial samples. The expression data was quantile normalized and the comparison in mRNA expression levels between familial and sporadic samples were performed by empirical Bayes t-test. By comparing the gene expression in familial vs. sporadic breast tumors overall we did not find any genes with expression greater than 1.2-fold above or below expression threshold. Since the estrogen receptor (ER) status has been shown to affect gene expression profiles in breast cancer, we further analysed the tumor samples according to the ER status. In the ER positive subgroup, we found 3 genes upregulated and 2 genes downregulated above and below 1.2-fold expression threshold in the familial vs. sporadic subgroups. The largest differences between the familial and sporadic subgroup were obtained in ER negative subgroup where we identified 25 up- and 25 downregulated genes (9 up- and 6 downregulated genes with greater than 2-fold expression threshold). The results help pinpointing genes with aberrant expression in putative germline mutation carriers compared to sporadic cases, especially in the ER defined subgroups and thus might indicate novel susceptibility genes.

3398F

Discovering latent cancer characteristics predictive of drug sensitivity.
D.A. Knowles, A. Battle, D. Koller. Stanford University, Stanford, CA.

Several recent studies have involved assaying the sensitivity of a library of cancer cell lines to an array of anti-cancer compounds, in particular the Cancer Cell Line Encyclopedia (CCLE), Genomics of Drug Sensitivity in Cancer (GDSC) and the Heiser dataset on which the DREAM challenge was based. Along with growth curve measurements, molecular characteristics of the cell lines are assayed: CCLE for example includes gene expression microarrays (GE), copy number variation (CNV), and oncogene mutation status assays. We extend previous per drug regression analysis using biologically informed dimensionality reduction and performing joint analysis across drugs in order to improve statistical power. We perform two stages of dimensionality reduction on the cell line characteristics, both using biological knowledge. In the first stage, we use a one dimensional factor analysis model for each gene, so that each data modality is considered an observed variable, explained by the latent factor which we envisage as the activation level of this gene. This is distinct from simply averaging across the different modalities: for example, methylation might be negatively correlated with gene expression, which we can account for. In the second stage, we use a hand curated collection of 1987 known pathways, collected from resources including GO, KEGG, and published GWAS hits, to construct a per pathway activation level, again using a one dimensional factor analysis. This approach gives easily interpretable results since the predictive pathways are associated with particular functions. While cancer is an incredibly heterogeneous disease there are characteristics shared across many cancers: disruption of apoptosis, cell cycle regulation, or DNA repair mechanisms, or addiction to specific oncogene pathways. We develop a model that hypothesizes a finite set of such unobserved, latent characteristics, which explain the observed drug sensitivity patterns and which are encouraged to be related to the pathway activation levels found by the dimensionality reduction procedure. The presence or absence of these characteristics confers sensitivity or resistance to specific therapeutic compounds. Our initial results on the CCLE dataset suggest five such latent characteristics are key predictors of drug sensitivity across cell lines and drugs. The top two most frequently used pathways involve vasculature development and TGF-beta signalling.

3399W

Prognostic significance of syndecan-1 expression in colorectal carcinoma.
S.H Lee[1], E.J Choi[2], J.A Yoon[2], E.S Jung[1], S.Y Kim[2]. 1) Department of Hospital Pathology, College of Medicine, Seoul St. Mary's Hospital, The Catholic University of Korea, Seoul, South Korea; 2) Department of Pathology, College of Medicine, The Catholic University of Korea, Seoul, South Korea.

Background: Syndecan-1 (SDC1) is reported to modulate several key processes of tumorigenesis and have variable expression in many cancers. The cause provoking altered expression is not known to date. In this study, we compared SDC1 status with various clinicopathologic parameters and molecular markers to evaluate clinical implications in colorectal carcinoma. Methods: With 219 surgical specimens of primary colorectal carcinoma, treated between 2008 and 2010 at the Seoul St. Mary's hospital, the catholic university of Korea. The study group consisted of 136 men and 83 women. The mean age was 62.3 years (range 32-93 years). We screened SDC1 expression using immunohistochemistry and analyzed the relationship between SDC1 expression and various clinicopathological parameters and molecular markers. Results: The tumors were located mainly in the left colon (68.5%) and especially in the rectum (37.9%). There were 206 (94.1%) adenocarcinomas, 10 (4.6%) mucinous adenocarcinomas and 3 other tumors. Most of the carcinomas were pT3 and pT4 (n=150 and 49, respectively). Regional lymph nodes contained metastases in 138 patients. SDC1 expression was found in cancer cells in 212 cases (96.8%) of colon cancer. Of the SDC1 expression cases, 131 cases dominantly showed membranous immunopositivity, 81 cases showing cytoplasmic pattern. 154 cases showed mixed membranous and cytoplasmic pattern. In 93 cases, stromal SDC1 reactivity was noted. It was significantly associated with primary tumor (T) (p=0.000) and perineural invasion (p=0.043). In addition, it was correlated with EGFR immunohistochemical reactivity (p=0.004). On the contrary, it was not significantly correlated with lymph node metastasis (N), distance metastasis (M), lymphatic and vascular invasion, and K-ras mutation states Conclusions: We showed that SDC1 expression is associated with T stage, status of perineural invasion and EGFR immunoreactivity in colorectal carcinoma. The expression profiles of SDC1 may be of prognostic value in colorectal cancer and may help in identifying aggressive forms of colorectal carcinoma. Further studies are necessary to better understand the role of SDC1 in the progression and invasiveness of colorectal carcinoma.

3400T

Mutations and copy number changes identified in primary brain tumors using complementary analyses and formalin-fixed, paraffin-embedded (FFPE) tissues.
A. Ligon[1,2,3], B. Alexander[2,3], S. Ramkissoon[1,2,3], P. Wen[2,3], D. Reardon[2,3], E. Lee[2,3], M. Rinne[2,3], A. Norden[2,3], L. Nayak[2,3], S. Nuland[2], L. Doherty[2], D. Lafrankie[2], L. Brown[2], N. Arvold[2], S. Santagata[1,2,3], I. Dunn[2,3], N. Lindeman[1,3], L. MacConaill[1,2], B. Rollins[2,3], R. Beroukhim[2,3,4], K. Ligon[1,2,3]. 1) Department of Pathology, Brigham & Women's Hospital, Boston, MA; 2) Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Broad Institute, Cambridge, MA.

Routine and integrated genomic approaches for genotyping clinical oncologic specimens have not yet been well-established. Whole genome testing in of oncologic specimens can inform diagnosis, prognosis and genotypic stratification with respect to clinical trials. However, the accessibility of such data is constrained by the frequent need to rely on routine formalin-fixed, paraffin-embedded (FFPE) samples as starting material. In this study, whole genome copy number analysis was performed as a routine clinical test for 400 primary adult and pediatric FFPE brain tumors. The tumors, which included meningiomas and gliomas, were analyzed by array-based comparative genomic hybridization (aCGH) using stock 1x1M Agilent SurePrint G3 arrays. Twenty-three copy number aberrations relevant to diagnosis, prognosis or clinical management were scored for each tumor. In addition, as part of a clinical trials research program, we profiled somatic mutations using the Oncomap v4.4 platform (Sequenom), which queried 471 known cancer-related mutations in 41 genes. We identified *PIK3CA* mutations in grade I meningothelial subtype tumors and found that sequence mutations, including *IDH1* and *IDH2* mutations, were present in ~30% of gliomas, mostly in tumors of WHO grade II or III. Both mutation and copy number data were obtained for a subset of the gliomas, 68% of which were classified as glioblastoma (GBM). Copy number changes identified in GBM varied with the patient age and *IDH1* mutation status, and included polysomy 7, *EGFR* amplification, identification of the EGFRvIII structural variant, as well as losses of *CDKN2A*, *PTEN*, *RB1* and *TP53*. Copy number profiles for *IDH1*(p.R132H) mutant tumors generally showed less complexity than that of tumors with wild-type *IDH1*. 1p/19q co-deletions were readily identified in oligodendroglial tumors, as were single copy gains involving *BRAF* that are consistent with the *BRAF* duplication event observed in pilocytic astrocytomas and related tumors. In summary, we developed complementary assays for detecting common cancer-specific mutations and genome wide copy number changes from FFPE tissues. Together, these assays have been integrated to generate data that are diagnostic (e.g., 1p/19q co-deletion), prognostic (e.g., *IDH1* and *IDH2* mutation status), or that illuminate genomic aberrations (e.g., identification of EGFRvIII variant) for use in advancing clinical trial enrollment and therapeutic management.

3401F

A leukemic stem cell score associated with patient prognosis and tumor histopathology in ovarian cancer.
B.A. Logsdon[1], S.L. Battle[1], M.H. Rendi[2], R.D. Hawkins[1], S-I. Lee[1,3]. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Department of Anatomic Pathology, University of Washington Medical Center, Seattle, WA; 3) Department of Computer Science and Engineering, University of Washington, Seattle, WA.

We derive a novel leukemic stem cell (LSC) score based on the gene expression levels of ten genes associated with both patient prognosis and leukemic stem cells in acute myelogenous leukemia (AML). Using gene weights estimated from AML patients across multiple studies, we generate the LSC score in three independent gene expression data-sets for patients with ovarian cancer. The LSC score is significantly associated with poor patient prognosis in ovarian cancer (Cox proportional hazard inverse normal meta-analysis p-value: 8.55×10^{-7}). The LSC score is also significantly associated with the optimal debulking status of tumors (Fisher meta-analysis p-value: 4.93×10^{-4}). In data available from The Cancer Genome Atlas on patients with ovarian serous carcinoma we show that the LSC score is associated with abnormal tumor stroma, increased desmoplastic response, and increased tumor invasiveness. Additionally, there is evidence of increased abnormal angiogenesis in patients with high LSC scores, suggesting an opportunity for targeted interventions with anti-vascular endothelial growth factor therapies. This strategy is further motivated by recent efforts to target tumor vascularization with anti-angiogenic therapies in AML. Our results suggest the LSC score is a novel biomarker that is informative for cancer stem cell populations, abnormal tumor histopathology, and patient prognosis in ovarian cancer.

3402W

Coupled Use of Family-based Exome Sequencing and TCGA Germline Data Analysis to Identify Novel Breast and Ovarian Cancer Susceptibility Genes. *J.A. Martignetti[1], P. Dottino[1], M. Babcock[1], L. Jara[3], K. Moysich[2], L. Sucheston[2], S. Lele[2], K. Odunsi[2], S. Sanga[4].* 1) Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) School of Medicine and Biomedical Sciences State University of New York at Buffalo, Buffalo, NY, USA; 3) Human Genetics Program, University of Chile School of Medicine, Santiago, Chile; 4) Station X, Inc., San Francisco, CA, USA.

Family history is the strongest single predictor of a woman's risk for developing breast and/or ovarian cancers. Genetic studies seeking to identify breast and ovarian cancer susceptibility genes have therefore focused on those families with a high incidence of cancer across multiple generations. While they represent the strongest known genetic predictors, BRCA1/2 mutations account for less than half of all families containing two or more cases in first-degree relatives and explain less than half of the excess familial cancer risk. Our hypothesis is that the residual risk of familial breast and ovarian cancers is attributable to mutations and/or genetic variants in genes other than BRCA1/2. To identify novel susceptibility genes, we exome sequenced the germlines of a discovery cohort of selected families with hereditary breast and ovarian cancer but who lacked deleterious BRCA mutations. The discovery cohort was sequenced using Illumina technology at the Icahn School of Medicine at Mount Sinai and Beijing Genomics Institute. Permutations of BWA sequence alignment, GATK re-calibration/re-alignment, SRMA re-alignment, GATK variant calling, and VarScan2 variant calling were used to identify small genomic variations (SNVs, small INDELs). These variants were annotated, visualized, analyzed, and evaluated within GenePool™ (Station X, Inc., San Francisco, CA) for rare variants consistent with an autosomal dominant model of inheritance. All exome-defined candidate mutations were then triaged by leveraging biological annotations, filtering capabilities, and cross-referencing the interactive genome browser within GenePool™. Once this highest tier of candidates was established, they were then compared within and between families, and candidate genes mutated in multiple families were identified. To further refine this list and begin to gain an understanding of the extent of their mutation in a larger cohort, we compared our findings to germline and somatic variant allele frequencies derived from the most up-to-date TCGA breast and ovarian cancer genome data. The somatic and germline variants in the TCGA cohorts were called using bioinformatics pipelines consistent with those used for the discovery cohort, and made accessible through GenePool™ at varying granularities along with the relevant clinical metadata per sample. The final variant list of novel breast and ovarian susceptibility genes, and functional validation studies will be discussed.

3403T

KiSS1 suppressor metastasis gene rs12998 and rs5780218 polymorphisms in Mexican patients with breast cancer. *M.C. Moran Moguel[1], S.E. Flores Martínez[1], L.A. Juarez Aguilar[2], J. Sanchez Corona[1], I.P. Davalos Rodriguez[1], R.C. Rosales Gomez[1], S.A. Gutierrez Rubio[1], M.I. Torres Moran[3], E.G. Cruz Quevedo[1].* 1) Centro de Investigación Biomedica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, Mexico; 2) Doctorado en Genética Humana. CUCS. Universidad de Guadalajara. Guadalajara, Jalisco. Mexico; 3) IMAREFI. CUCBA. Universidad de Guadalajara. Guadalajara, Jalisco. Mexico.

Aims. KiSS1 is a metastasis suppressor gene that has been associated with inhibition of cellular chemotaxis and invasion attenuating the metastasis in melanoma and breast cancer cell lines. Along the KiSS-1 gene have been described at least 130 SNPs however the association of these polymorphisms as genetic markers for metastasis in breast cancer studies has not been investigated. Here we describe two simple PCR-RFLPs protocols to identify the 9DelT (rs5780218) and E20K (rs12998) KiSS1 polymorphisms and the allelic, genotypic and haplotype frequencies in Mexican general population (GP) and patients with benign breast disease (BBD) or breast cancer (BC) in any histological stage. Results. Individually, none of the polymorphisms were associated with breast cancer, since when comparing alleles and genotypes between the BC group vs BBD and GP groups, no differences were finded ($p > 0.05$); however, for polymorphism rs 12998 comparing GP vs case (BC and BBD) groups, statistically significant differences were found. The haplotype consisting of the wild-type allele of rs12998 and variant allele of rs5780218 polymorphisms (G/del) occurred more frequently in the BC group (0.4256) whereas haplotype G/T (both wild type alleles) was the most prevalent in BBD group (0.4674). Conclusions. Our data indicated that these polymorphisms individually do not confer susceptibility for development of breast cancer or metastasis in Mexican population, however, a possible role as genetic marker in breast cancer metastasis for H1 haplotype (Wt/Variant) in KiSS1 gene, must be analyzed in other populations.

3404F

Initial viral load in cases of single human papillomavirus 16 or 52 persistent infection is associated with progression of later cytopathological findings in the uterine cervix. *K. Ohashi[1], K. Miura[1], S. Abe[1], A. Kinoshita[2], S. Miura[1], D. Hamaguchi[1], KI. Yoshiura[2], H. Masuzaki[1].* 1) Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Nagasaki, Japan; 2) Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Nagasaki, Japan.

Objectives: The aim of this study was to investigate the relationship between viral load in single human papillomavirus (HPV) 16 or 52 persistent infection and the progression of later cytopathological findings in the uterine cervix. **Methods:** Cervical cytological tests and HPV genotyping tests were repeated within 3-6 months in 305 women with oncogenic HPV. Then, 24 cases of a single HPV52 persistent infection and 24 cases of a single HPV16 persistent infection were identified. Cases of later cytological findings showing progression were defined as 'progression' group, while cases of no-change or regression were as 'non-progression' group. Relative HPV DNA loads were determined by quantitative real-time polymerase chain reaction and expressed relative to human albumin (ALB) DNA. Differences between the two groups were evaluated. **Results:** The median relative HPV 52 DNA load was 2.211 in the progression group and 0.022 in the non-progression group (Mann-Whitney U test, $P=0.003$). The median relative HPV 16 DNA load was 4.206 in the progression group and 0.103 in the non-progression group ($P=0.001$). **Conclusions:** HPV 52 or 16 DNA load, which is measured by quantitative real-time methods, may be useful as a short-term marker to identify women at high risk for progression of cervical cytological findings.

3405W

Conventional and Molecular Cytogenetic Techniques in Comparison with mRNA and DNA Based Quantitative Real-Time PCR to Monitor Minimal Residual Disease in Chronic Myeloid Leukemia. *I.S. Pagani[1,4], C. Pirrone[1], D. Pigni[1], O. Spinelli[2], C. Boroni[2], T. Intermesoli[2], U. Giussani[2], F. Pasquali[1], F. Lo Curto[1], A. Lanfranchi[3], F. Porta[3], A. Rambaldi[2], G. Porta[1].* 1) Department of Clinical and Experimental Medicine, Insubria University, Varese, Italy; 2) Paolo Belli Haematology Division, Papa Giovanni XVIII Hospital, Bergamo, Italy; 3) Hematology and BMT Unit Children's Hospital Spedali Civili Brescia, Italy; 4) Dept. of Experimental Medicine and Surgery, University of Rome Tor Vergata, Italy.

The Imatinib mesylate (IM) is the first line therapy against Chronic Myeloid Leukemia (CML), through the inhibition of the BCR-ABL1 proliferation pathway. Despite its efficacy in prolonging overall survival, discontinuation of treatment is associated with molecular relapse. So, to maintain an operational cure IM is required indefinitely, despite financial cost to the community and side effects. To evaluate the degree of response to therapy and to highlight the persistence of the disease after treatment, patient should be monitored routinely. The gold standard for diagnosing CML is the cytogenetic analysis, a direct not-sensitive method to detect Ph-positive cells. Quantitative real-time RT-PCR (qRT-PCR) based on mRNA is the most sensitive technique available for the detection of BCR-ABL1 transcripts and it is used to follow the progression of CML after initial diagnosis and treatment. However mRNA levels are not directly related to number of leukemic cells and negative results are difficult to interpret, because undetectable levels of chimeric transcript can reflect either an effective elimination of leukemic cells, or the presence of a leukemic cell transcriptionally silent. Here we will propose a new sensitive approach to directly detect the number of leukemic cells using a DNA-based biomarker specific for each patient. For the first time we developed a DNA Q-PCR assay based on the genomic breakpoint found with Next-generation sequencing and a formula to calculate the number of Ph+ cells. We monitored eight CML patients in their early chronic phase and in follow-ups up to 8 years under Imatinib treatment. We carried out patient specific Q-PCR assays to monitor minimal residual disease, testing the same samples in parallel by cytogenetic analysis and by standard qRT-PCR. In all positive samples for chimeric transcript we measured positive levels of corresponding genomic DNA, confirming the sensitivity of the method. Furthermore, we showed the persistence of leukemic cells transcriptionally-silent by Q-PCR in 33% of samples with undetectable levels of mRNA. Finally, we applied our technique in the evaluation of BCR-ABL1 presence in CD34+ sorted cells in order to enrich the cancer stem cells. In conclusion the DNA Q-PCR is a sensitive and direct technique to identify leukemic cells and patients that could be candidate to stop the therapy. Thanks to AIRC, AIL.

3406T

ALK gene copy number gain in a series of 350 NSCLC: a single institution experience. *L. Pecciarini[1], A. Talarico[1], G. Grassini[1], E. Dal Cin[1], S. Foti[2], C. Lazzari[2], V. Gregorc[2], M.G. Cangi[1], C. Doglioni[1].* 1) Pathology Unit, San Raffaele Scientific Institute, Milan, Italy; 2) Oncology Department, San Raffaele Scientific Institute, Milan, Italy.

Anaplastic lymphoma kinase (ALK) gene translocations are involved in the tumorigenesis of a small group of non-small cell lung carcinomas (NSCLCs) and identify patients sensitive to ALK inhibitors. Therefore fluorescence in situ hybridization (FISH) analysis for ALK rearrangements is routinely performed in NSCLC patients for correct therapy selection. ALK gene copy number (GCN) gains are known to play an oncogenic role in tumors such as neuroblastoma, but they are poorly characterized in NSCLC. Indeed in our experience ALK GCN changes represent common incidental findings of the ALK FISH test in NSCLCs. In this study we report the prevalence of ALK GCN gain and its correlation to ALK protein expression, epidermal growth factor receptor (EGFR) gene and v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) gene mutational status, and clinicopathological data in a series of 350 NSCLC patients treated at our Institution. ALK translocations and gene status was evaluated by fluorescence in situ hybridization (FISH), using a commercial split-apart ALK (2p23) probe. Specimens positive for a ALK split signal were studied for echinoderm microtubule-associated protein-like 4 (EML4, 2p11) gene involvement using a three color FISH probe. Gain of ALK GCN was defined in 2 categories: low gain = 4<ALK<12 copies in ≥40% of tumor cells and high gain = ALK ≥12 copies in ≥5% of tumor cells. ALK expression was assessed by immunohistochemistry. EGFR and KRAS mutational status were also assessed by both direct sequencing and pyrosequencing. ALK translocations were identified in 10% of the analyzed cases and none of them were either EGFR or KRAS mutated. Interestingly ALK GCN gains were observed in 15% of the ALK translocation negative cases: 10% with low gain and 5% with high gain. EGFR gene and KRAS gene were mutated in respectively 20% and 8% of the samples with ALK GCN gains. ALK translocations, but not ALK GCN gains were associated with ALK protein overexpression. In order to better characterize ALK GCN gains in NSCLCs, correlation with clinico-pathological data and both EGFR gene and KRAS gene mutations are under evaluation and results will be presented.

3407F

Shortened telomere length and survival in bladder cancer patients. *A. Russo[1,2], F. Modica[1], S. Guarrera[1], G. Fiorito[1,2], A. Allione[1], B. Pardini[1], R. Critelli[1,2], F. Ricceri[1], A. Bosio[3], G. Casetta[4], G. Cucchiarale[5], P. Destefanis[3], P. Gontero[4], L. Rolle[3], A. Zitella[4], D. Fontana[6], P. Vineis[1,7], C. Sacerdote[8], G. Matullo[1,2].* 1) Human Genetics Foundation, Turin, Turin, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy; 3) Department of Urology 2, San Giovanni Battista Hospital and University of Turin, Turin, Italy; 4) Department of Urology 1, San Giovanni Battista Hospital and University of Turin, Turin, Italy; 5) Department of Urology, Clinica Cellini, Turin, Italy; 6) Department of Epidemiology, Biostatistics & HTA, Radboud University Medical Centre, NL-6500 HB, Nijmegen, The Netherlands; 7) Imperial College London, UK; 8) CPO Piemonte, Turin, Italy.

Telomeres are the specialized DNA-protein structures composed of a variable number of simple repetitive nucleotide repeats (TTAGGG) that cap the ends of linear eukaryotic chromosomes. They are responsible for the protection of the chromosome ends from nucleolytic degradation, end-to-end fusion, irregular recombination, and other lethal events to a cell. Replicating somatic cells in vivo lose ~50-105 bp per year, because of DNA polymerase's inefficient replication of linear DNA ends. Excessive or accelerated telomere shortening results in critically short telomeres, inducing cellular senescence and apoptosis. If protective mechanisms, such as the TP53 tumor-suppressor gene, are inactive, thus allowing continued proliferation, telomeres become extremely short and dysfunctional; end-to-end fusions ultimately cause chromosomal instability that may contribute to the development of several types of cancer. Moreover, as a marker of cumulative cellular aging, short telomere length may also be associated with increased risk of early death in elder individuals. In the present study, we analyzed blood telomere length (TL) in 463 bladder cancer patients followed for up to 18 years using quantitative PCR to evaluate its association with risk of death. Patients with non-invasive tumor had significantly longer telomeres than patients with invasive tumor (RQ=0.78 vs. 0.71, p=0.03). Decreasing quartiles of TL were associated with decreasing survival after cancer (log-rank test, p=3.9×10-4). Finally, Cox model adjusted by age, cancer extension, radical cystectomy, radiotherapy and chemotherapy, showed an independent effect of TL on bladder cancer survival (HR=2.03 95% CI, 1.27-3.23, p=0.003). In conclusion, our results suggest that TL inversely correlates with tumor extension and that shortened telomeres act as an independent prognostic predictor for survival of bladder cancer patients.

3408W

Generating Hypotheses for Targeted Therapies with Cancer In silico Drug Discovery Tools. *F.A. San Lucas[1,2], J. Fowler[2], S. Kopetz[1,4], E. Vilar[1,3,4], P. Scheet[1,2].* 1) The Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston, Houston, TX; 2) Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 3) Clinical Cancer Prevention, The University of Texas M.D. Anderson Cancer Center, Houston, TX; 4) GI Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX.

We have developed *Cancer In silico Drug Discovery (CIDD) Tools* to empower clinical researchers to perform preliminary analyses for drug discovery projects by facilitating the setup, execution and evaluation of *in silico* cancer drug experiments. CIDD is a toolset that integrates genomic mutation data, tumor gene expression profiles and drug-induced gene expression profiles from public data sets to help researchers generate hypotheses for three general problems: (1) to determine if a particular somatic mutation or a set of mutations in a cancer is functional, producing unique gene expression signatures, (2) to find candidate drugs to treat, or repress, these alterations in gene expression, and (3) to identify cell-lines for subsequent lab experimentation that most closely represent the cancer being studied. CIDD integrates publicly available experimental data and annotation databases, such as data from *The Cancer Genome Atlas (TCGA)*, the *Connectivity Map (CMap)* and the *Cancer Cell-line Encyclopedia (CCLE)* to perform the in silico experiments. CIDD characterizes generated tumor expression signatures using *MSigDB* and describes candidate drugs by integrating data from annotation databases such as *DrugBank* and *SuperDrug*. An end result of a CIDD execution is a statistically derived, biologically interpretable candidate drug list. This empowers clinical researchers to make biologically informed decisions on candidate drugs. To illustrate its use, we applied CIDD to the study of BRAF-mutant colorectal cancers. CIDD identified BRAF-mutant (V600E) samples from the TCGA colorectal cancer (CRC) project, extracted RNA-seq and microarray gene expression data for BRAF-mutant and BRAF-wildtype samples, verified a BRAF-mutant gene expression signature that was previously proposed by Popovici et al (2012, Journal of Clinical Oncology, 30:1288-1295), and then identified and characterized a list of candidate drugs to target BRAF-mutant CRCs. Through CIDD, we identified EGFR and proteasome inhibitors to target BRAF-mutant CRCs, and CIDD proposed 7 possible cell-lines from the CCLE as the best candidates on which to test the drugs. CIDD is written in Python and requires R. CIDD will be available at http://scheet.org/software.

3409T

Decreased gene expression and localization of alkaline phosphatase (L/B/K) to the microvillus membrane of renal cells in renal cell carcinoma. *U. Sharma[1], D. Pal[1], S.K. Singh[2], N. Kakkar[3], A.K. Mandal[2], R. Prasad[1].* 1) Biochemistry, PGIMER, Chandigarh, Chandigarh, India; 2) Urology, PGIMER, Chandigarh, chandigarh, India; 3) Histopathology, PGIMER, Chandigarh, chandigarh, India.

The present study was conducted to explore the expression and localization of Liver/Bone/kidney alkaline phosphatase (L/B/K ALP) on renal tubular cells in clear cell renal cell carcinoma (RCC). A total of 50 patients of histopathologically confirmed cases of RCC were included in this study. The L/B/K ALP protein level was determined by immunohistochemistry, immunofluorescence and flow cytometry in renal cell carcinoma and adjacent normal renal parenchyma tissue. The mRNA expression of L/B/K ALP was detected using real time PCR. All the parameters were statistically analyzed. Immunohistochemistry showed mean immunoreactivity in all normal renal sections (n = 50) was 2.9. Whereas, mean L/B/K ALP immunoreactivity in all tumor samples (n = 50) was 0.7. Similarly, immunofluorescence showed mean L/B/K ALP immunoreactivity in all normal renal sections (n = 50) was 3.2. Whereas, it was 0.5 in tumor (n=50).The flow cytometric studies documented a significant reduction in the ALP presentation on microvillus membrane of RCC (291.9±16.8 vs 191±15.9; P<0.05). Further, Real time PCR analysis revealed a significant depreciation in the ALP mRNA transcript in the RCC tissue (0.65±0.09 vs 1.0±0.18; P<0.01). These findings conclude that reduced activity of ALP in BBM could be associated with reduced ALP gene expression and as well as decreased localization to microvillus membrane of RCC cells.

3410F

Hif1a and PHD gene expression levels in CML patients. *A. TOYLU[1], O. Altiok Clark[1], O. Salim[2], A. Timuragaoglu[2], M. Okur[3], OK. Yucel[2], M. Ulubahsi[3], K. Eker Guler[3], N. Sayin Ekinci[3], I. Karadogan[2], L. Undar[2].* 1) Department of Medical Genetics, Akdeniz University Medical Faculty, ANTALYA, Turkey; 2) Department of Hematology, Akdeniz University Medical Faculty, ANTALYA, Turkey; 3) Laboratory of Hematology, Akdeniz University Hospital, ANTALYA, Turkey.

Hif1a is a transcription factor regulating the expression of multiple genes responsible for the cancer development and metastasis. The expression of Hif1a gene is regulated by transcriptional and post-transcriptional mechanism which is related with the proteosomal degradation of Hif1a protein by PHD family proteins. It has been reported that Hif1a protein itself also regulates the expression of PHD family genes by a feedback mechanism. The high level of Hif1a expression was found in various types of cancer but the expression levels of PHD family genes were remain unclear. Recent studies on hematologic malignancies indicated the differential expression of Hif1a gene in leukemia cells. In particular, chronic myeloid leukemia (CML) stem cells express Hif1a gene accompanied by increased survival capacity. But still there is no data about the expression level of the PHD genes in leukemia cells. Our study was performed to investigate the expression of Hif1a and PHD genes in CML cells. The expression levels of Hif1a and PHD genes compared among each other and the relationship between the BCR-ABL which is the marker gene of CML, and Hif1a expression analyzed. Peripheral blood and bone marrow samples of 29 CML patients, who have been receiving imatinib therapy, were obtained to investigate mRNA expression of Hif1a, PHD1 and PHD2 genes. RPL13A gene was used as a reference gene. Real time quantitative PCR and REST analysis were performed to measure expression levels. Hif1a and PHD2 gene expressions were higher in BCR-ABL negative CML patients compared to BCR-ABL positive patients. There was no difference in the expression of PHD1 gene among these two groups. The patients who have a high level of Hif1a expression also showed high level of PHD2 expression but PHD1 expression levels were unchanged. These results indicate that Hif1a and PHD2 genes have a similar expression pattern among CML patients. According to our data, the expression of PHD1 and PHD2 genes is most likely regulated by Hif1a. But there might be another mechanism that is independent from Hif1a for the regulation of both PHD1 and PHD2 genes. The relationship between Hif1a and PHD family genes needs to be further analyzed in CML.

3411W

Applicability of high-resolution multicapillary electrophoresis for molecular characterization of immune gene rearrangement profiles in acute lymphoblastic leukemia (ALL). *H. Trautmann[1], M. Kozulic[2], A.-W. Kruse[1], M. Kneba[1], M. Brüggemann[1].* 1) Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany; 2) QIAGEN Instruments AG, Hombrechtikon, Switzerland.

Analysis of clonally rearranged immunoglobulin (IG)- and T-cell receptor (TCR) genes is used to identify markers for minimal residual disease (MRD) diagnostics in ALL. Therefore, products of standardized IG/TCR PCR assays are applied to GeneScan, dHPLC and/or heteroduplex analysis to distinguish polyclonal from clonal rearrangements. In this setting, we tested the applicability of an automated high-resolution capillary electrophoresis, the QIAxcel Advanced system (QIAGEN), as alternative option for fragment analysis. The performance of QIAxcel for analysis of IG/TCR PCR products was compared to GeneScan and dHPLC. A panel of 15 multiplex PCRs (BIOMED2), targeting different immune gene rearrangements (6 IGH, -L, -K and 9 TCRB, -G, -D PCRs) was applied to 10 B-lineage-ALL samples resulting in a total of 150 PCR products that were analyzed by QIAxcel, GeneScan and dHPLC. IG/TCR profiles were validated by sequence analysis. Additional sensitivity testing was performed via dilution experiments with patient samples in polyclonal DNA. Just as dHPLC and GeneScan, QIAxcel accurately detected all amplifiable clonal IG- and TCR-rearrangements (61/150) as well as all polyclonal signals (89/150). In addition, depending on the individual PCR target, the detection limit was 10-5% of clonal cells in a polyclonal background. The applicability of QIAxcel to identify IG/TCR markers in ALL is on a par with dHPLC and GeneScan. In addition, it reduces manual handling errors, eliminates the need for gel preparation and ensures electronic documentation of data. Our results indicate that it is a reliable, cost-effective and accurate high-throughput tool for IG/TCR marker characterization in ALL.

3412T

Investigation of Resveratrol/AT-101`s Molecular Targets of Apoptotic Effect to the Hormon Sensitive and Hormon Nonsensitive Prostate Cancer Cell Lines. *Y. YUKSELTEN[1], N. AKTEPE[2], A. KISIM[3], H. ATMACA[3], B. KARACA[4].* 1) Department of Medical Biology, Medicine Faculty of Ankara University, Ankara, Turkey; 2) Department of Nursing, Health High School of Artuklu University, Mardin, Turkey; 3) Department of Molecular Biology and Genetics, Faculty of Arts and Sciences- Celal Bayar University, Manisa, Turkey; 4) Department of Oncology, Medicine Faculty of Ege University, Izmir, Turkey.

Objective: Resveratrol, which is used in the cancer treatment, shown to have occurred DNA damage repair features and inhibits cancer cell proliferation by including anticancer activity, is a phytoalexin derived from the skin of grapes. AT-101, an (-)- enantiomer of gossypol, is a potent anticancer agent that it was reported to be an inhibitor of Bcl-2/Bcl-XL. In this study, the possible synergistic cytotoxic and apoptotic effects of Resveratrol in combination with AT-101 were investigated in human hormone refractory prostate cancer cell lines, PC-3 , DU- 145 and LNCaP in a time- and dose-dependent manner.

Material and Methods: Cytotoxicity was determined by XTT Cell Proliferation Kit (Roche). Drug synergy was assessed by using CalculSyn 2.1 software (Biosoft). Apoptosis was detected by Cell Death Detection Elisa Plus Kit (Roche) and confirmed by Caspase-Glo 3/7 Assay. The expression levels of apoptotic proteins were assessed by human apoptosis antibody array. Expression levels of proteins associated with apoptosis were investigated by Human Apoptosis Array Kit (R&D Systems, UK).

Results: The IC50 values of AT- 101 and Resveratrol were found 5.51 µM, 10 µM, 7.5 µM and 114 µM, 166 µM, 150.5 µM in human prostate cancer cell lines DU-145, PC3 and LNCaP, respectively. Combined treatment (100 µM + 5 µM AT-101) was shown to have strong synergistic cytotoxic and apoptotic effects in DU-145 cell line, at 72 hours. The combined use of AT-101 and Resveratrol resulted in an increase at Bax, Fas/TNFSF6, Pro-Caspase-3, Cleaved-Caspase-3, FADD, Phospho p53 (S46), p21/CIP1/CDNK1A, Cytochrome c, Smac/DIABLO and a decrease in Bcl-2, Bcl-x, XIAP, p27/Kip1, cIAP-1 proteins in DU-145 prostate cancer cells.

Conclusion: Our data revealed that the combination of AT-101 with Resveratrol may hold great promise for development as a novel chemotherapeutic approach to human prostate cancer. In determined synergism combination rates, it needs confirmation of promising cytotoxic and apoptotic effects in other prostate cancer cell lines; and also clarification of impact mechanism and confirmation of these data in animal experiments.

3413F

Cytogenetic and FISH monitoring of t(16;21)(p11.2;q22) FUS-ERG fusion gene in an AML NOS patient. *L. Militti[1,2,3], D. Fantasia[1], A. DiTecco[3], D. Onofrillo[4], A. Spadano[4], G. Calabrese[1,3], G. Palka[1,3].* 1) Dip. S. Mediche, Orali & Biotec, University of Chieti, Chieti, Italy; 2) AIL Associazione Italiana contro le Leucemie - Linfomi e Mieloma, Pescara Section, Pescara, Italy; 3) Medical Genetics, Pescara Hospital, Pescara, Italy; 4) Hematology Dept., Pescara Hospital, Pescara, Italy.

Reciprocal t(16;21)(p11.2;q22) is a rare chromosomal abnormality in acute myeloid leukemia (AML), being about 60 cases described to date, associated with dismal prognosis. In this rearrangement the ERG gene on chromosome 21 and the TLS/FUS gene on chromosome 16 result in a TLS/FUS-ERG fusion gene which is thought to be responsible for leukemogenesis. We report on a AML NOS new case (WHO 2008) with t(16;21)(p11.2;q22) investigated with different approaches. At diagnosis the karyotype of the 16 year-old male patient, as from conventional cytogenetic, SKY-FISH and MCB, was: 46,XY,t(16;21)(p11.2;q22)[12]/46,XY,inv(11)(p14q22),t(16;21)(p11.2;q22)[5]/46,XY,t(11;15)(p15;q13),t(16;21)(p11.2;q22)[3]/46,XY[1]. FISH analysis with a novel FUS (16p11) break-apart probe (Kreatech, NL) revealed 77% of cells with split signals. The patient received induction chemotherapy according to AIEOP protocol, resulting in normalization of karyotype and reduction of leukemic cells by FISH up to 1.35%. Five months from diagnosis he underwent sex-matched allogeneic bone marrow transplantation (BMT). At this time he had normal karyotype, while FISH revealed 0.4% of split signal cells. Five months later he relapsed with a new abnormal clone: 46,XY,der(10)add(10)(p15)dup(10)(q23q26),t(16;21)(p11.2;q22)[17]/46,XY[7]. FISH showed 20% of cells bearing FUS split signal, confirmed by quantitative RT-PCR and STR polymorphisms analysis. The patient had chemotherapy using FLAN protocol and achieved complete hematological and cytogenetic remission. FISH monitoring displayed progressive reduction of FUS splitted cells being 0.47% of cell population at time of 2nd BMT, 4 months from relapse. Six-month 2nd BMT follow up using FISH showed lowering of leukemic clone up to complete molecular remission. Present report shows MRD monitoring feasibility using FISH analysis for this rare translocation, and demonstrates genetic instability of FUS-ERG rearrangement cells as evidenced by sequential occurrence of three different leukemic clones during disease course.

3414W

Intrafocal Heterogeneity of Gene Rearrangements in Prostate Cancer: Implications to Genomic Targeting Therapeutics. *I. Tereshchenko[1,2], H. Zhong[1], U. Santanam[1], W. Petrosky[1], N. Kane-Goldsmith[2], J.A. Tischfield[2], R.S. DiPaola[1].* 1) The Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ; 2) Department of Genetics, Rutgers University, Piscataway, NJ.

Efforts to develop targeted therapy based on genomic findings are complicated by the polygenic nature of drug resistance and genetic heterogeneity between separate biopsy cores. These findings have led to the conclusion that analysis of multiple biopsy cores and consideration of multiple targeted agents will be important. Because of our concern that these assessments may be missing an additional layer of heterogeneity within even a single biopsy core, which sequencing alone may not identify, we studied the genetic heterogeneity of cells in focal areas within single biopsy cores. Using the TMA slides and confocal microscopy, we analyzed the most common and prostate cancer specific gene rearrangement (TMPRSS2-ERG fusion) in neighboring single cells in a tumor focus without changing the stromal micro-environment. We evaluated 55 patients that underwent radical prostatectomy at CINJ with pathological assessment, grading, and ERG rearrangement status. A minimum of 50 cells in each TMA core was scored using the ERG break-apart FISH assay. Overall, ERG rearrangements occurred in 24/55 (44%;) patients. The most frequent mechanism of rearrangements was deletion (58%), while a split event was confirmed in 25% of the ERG-rearranged cases. In 88% of cases the ERG rearrangement was associated with TMPRSS2-ERG gene fusion. A copy number gain of the non-rearranged ERG gene occurred in 10/55 (18%) cases. A heterogeneous mix of ERG deletions, translocations, and copy number changes occurred in 8/24 (33%) patients, with a combination of deletions and translocations alone occurring in 4/24 (17%) cases. These results support the notion that TMPRSS2-ERG gene fusions may arise independently within the same tumor focus and that additional heterogeneity may be present even in a focus of a single biopsy core. These data are extremely important and urgently require further study to understand heterogeneity of genetic changes, because clinical studies have already been launched using genetic changes such as TMPRSS2-ERG fusions found in a single core biopsy to direct a specific therapy.

3415T

Trisomy of chromosome 8 in children with haematological malignancies. *D. Januszkiewicz-Lewandowska[1,2,3], O. Zajac-Spychala[2], E. Maly[3], J. Nowak[1].* 1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) Department of Pediatric Oncology, Hematology and Transplantation, University of Medical Sciences, Poznan, Poland; 3) Department of Medical Diagnostic, Poznan, Poland.

Trisomy of chromosome 8 is one of the most common numerical aberrations in adults with AML and MDS, which is present in 5-20% of the cases. Much less frequently is observed in ALL and CML (up to 5% of patients). According to the IPSS cytogenetic classification of MDS trisomy of chromosome 8 is an indirect risk marker. There is no clear data on the prevalence and role of trisomy 8 in pediatric haematological malignancies. In 2009-2012 trisomy of chromosome 8 was found in 5 (4.4%) patients out of 113 children (82 ALL, 26 AML and 5 MDS) treated at the Department of Pediatric Oncology, Hematology and Transplantation in Poznan. In two of them were diagnosed ALL (2/82 - 2.4%), in one AML-M4/M5 (1/26 - 3.8%) and in other to MDS (2/5 - 40%). In three children, trisomy 8 was isolated aberration, in the remaining two, accompanied by monosomy of chromosome 7 (children with MDS) or t(9;22), transcript p210/230 BCR/ABL (child with ALL). In 4 patients (one child with ALL was treated only by ALL IC BFM 2002), out of chemotherapy, treatment included HSCT. Only one boy with relapsed leukemia with transformation of primary pre T-ALL to AML-M0 failed treatment. In general observed frequency of trisomy of chromosome 8 in haematological malignancies in children (4.4%) appears to be lower as compared to adults. The decision about the early use of HSCT in the treatment of studied children could be the cause of received good treatment results.

3416F

Joint effect of multiple common SNPs predicts melanoma susceptibility in place of or beyond traditional risk factors. *S. Fang[1], J. Han[2], M. Zhang[2], L. Wang[1], Q. Wei[1], C. Amos[3], J. Lee[1].* 1) University of Texas MD Anderson Cancer Center,Houston, TX; 2) Harvard Medical School, Boston, MA; 3) Dartmouth College, Lebanon, NH.

Genome-wide association studies have identified several genes that are associated with melanoma risk. However, most common genetic variants identified by those studies have only weak effects on melanoma risk, as well as on pigmentation or nevi. We hypothesized that the joint analysis of multiple single-nucleotide polymorphisms (SNPs) may detect a larger effect and improve the predictive value of models using standard phenotypic risk factors. In this study, we analyzed 11 SNPs that were associated with melanoma risk in previous studies and were also genotyped in The University of Texas MD Anderson Cancer Center melanoma case-control study and the Harvard Medical School Nurse Health and Health Professionals Follow-Up studies. All 11 SNPs were replicated in the MD Anderson study, but only 5 in the Nurse Health study and 2 in the Health Professionals Follow-Up study were confirmed. Participants who carried 15 or more risk alleles were more than five times as likely to have melanoma as compared to those carrying ≤6 risk alleles in the MD Anderson study. A weighted polygenic risk score (PRS) was constructed using the 11 SNPs to evaluate their joint effect on melanoma risk. In a data set pooled from all 3 studies, per unit increase of PRS led to 1.12 times more likely to have melanoma disease with adjustment for age, sex and pigmentation(95%; CI, 1.06-1.18, P= 4.63×10-5); Individuals in the highest PRS tertile were 1.69 times more likely to have melanoma than were those in the lowest PRS tertile (95%; CI, 1.28-2.25; P = 2.24 × 10-4), after adjustment for age, sex, and pigmentation, in the pooled dataset. Furthermore, PRS significantly improved the predictive value and discriminability of a conventional phenotypic melanoma risk model, and PRS alone had almost the same predictive value as the conventional model in the MD Anderson study. These results indicate that PRS may be used to replace standard phenotypic variables for melanoma risk assessment; Inclusion of PRS contributes to a small increase of risk predictive value of standard phenotypic model.

3417W

Radiogenomics: Using Genetics to Identify Cancer Patients at Risk for Development of Adverse Effects Following Radiotherapy. *S.L. Kerns[1,2], H. Ostrer[2,3], B.S. Rosenstein[1,4,5], Radiogenomics Consortium.* 1) Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY; 3) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 4) Department of Radiation Oncology, New York University School of Medicine, New York, NY; 5) Departments of Dermatology and Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Radiogenomics (RGx) research aims to identify genetic predictors of response to cancer radiotherapy. RGx focuses on common clinical endpoints, such as fibrosis in irradiated breast tissue, pneumonitis in irradiated lung tissue, and urinary and bowel symptoms following radiotherapy of pelvic cancers. It parallels pharmacogenomics with respect to clinical aims and research methods. RGx is a relatively young discipline with many opportunities for discovery and clinical impact, but to date, the majority of RGx papers have been published in the radiation oncology literature. This abstract aims to increase understanding of RGx among human geneticists and genetic epidemiologists by addressing 4 major points. 1) What is the evidence of a genetic basis for RGx phenotypes? The strongest evidence that common genetic variation underlies such phenotypes comes from a study in breast cancer patients comparing intra- with inter-patient variation in skin toxicity, which attributed 80-90% of the variability to genetics. 2) What have RGx studies shown to date and where is the field going? Early RGx studies focused on SNPs in DNA repair, inflammation and radiation response pathways. This approach yielded conflicting results, and the field has recently shifted to a genome-wide approach with an effort to build large cohorts. To date, two published GWAS have identified SNP associations with adverse effects from radiotherapy for prostate cancer, including rs7120482 and rs2268363 that reached genome-wide significance for association with rectal bleeding and erectile dysfunction (ED) respectively, and rs17779457 that approached genome-wide significance for association with urinary morbidity. 51 additional loci have been replicated in multiple validation cohorts at levels nearing genome-wide significance. 3) What types of models have been built? Robust multi-SNP models have been developed using re-sampling or external validation, and show stronger association signals compared to single-SNP models. Addition of SNPs improves models based on clinical predictors. 4) What are the plans for building big GWAS in RGx? Future directions focus on large, well-powered GWAS, meta-analysis, fine-mapping and polygenic models incorporating gene-environment interaction. This has been enabled by formation of the NCI-supported Radiogenomics Consortium (http://epi.grants.cancer.gov/Consortia/single/rgc.html), which aims to foster collaboration and encourage translation to clinical practice.

3418T

RLIP76 expression levels in CML patients. *O. Altiok Clark[1], A. Toylu[1], O. Salim[2], A. Timuragaoglu[2], M. Okur[3], OK. Yucel[2], N. Sayin Ekinci[3], M. Ulubahsi[3], K. Eker Guler[3], I. Karadogan[2], L. Undar[2].* 1) Department of Medical Genetics, Akdeniz University Medical Faculty, Antalya, Turkey; 2) Department of Hematology, Akdeniz University Medical Faculty, Antalya, Turkey; 3) Laboratory of Hematology, Akdeniz University Hospital, Antalya, Turkey.

The Ral-interacting protein (RLIP76) is a member of the non-ABC transporter group of multiple drug-resistance proteins, involved in the export of intracellular drugs, GSH conjugates, and lipid peroxication end products out of the cell. While it is believed that this protein might be responsible for the drug resistance observed in various types of cancers, the expression of RLIP76 has yet to be studied in hematologic malignancies. Furthermore, while it is known that the RLIP76 gene is expressed in the HL60 and K562 leukemia cell lines, the role of the RLIP76 protein in both the pathology of leukemia and chemotherapeutic resistance remains largely unknown. As the prevalence of the drug resistance phenotype, and the resultant grim prognoses, both increases among leukemia patients, the need to better understand the role of the RLIP76 protein is imperative. It was thought that the expression levels of RLIP76 might be important for giving accurate prognoses for CML cases that can't otherwise be explained by any of the other known drug-resistance genes. Therefore, based on the potential at RLIP76 to be a new and useful prognostic marker for investigating drug-resistance in CML, the expression levels of the RLIP76 gene were measured in CML patients that were receiving imatinib, in this study. Analysis of RLIP76 mRNA showed that BCR-ABL positive patients had a lower level of RLIP76 than did BCR-ABL negative patients and this could be due to the effects of patient age, duration of treatment, and BCR-ABL expression levels.

3419F

Circulating microRNAs in acute lymphoblastic leukemia during chemotherapy treatment. *T.M. Marques[1], L.H. Sakamoto[2], C.P.C. Gomes[1], G.P. Oliveira Jr[2], G.R. Fernandes[1], R.W. Pereira[2].* 1) University of Brasilia, Brasilia, Distrito Federal, Brazil; 2) Catholic University of Brasilia, Distrito Federal, Brazil.

Circulating microRNAs have been found in all body fluids and could be associated with modulation of tumor microenvironment and progression of diseases. In the presence of cancer cells, active components present in normal microenvironment, become deregulated, promoting cell survival, disease progression and drug resistance. Several microRNAs act as participants in the process of development of leukemia either increasing or decreasing the expression levels. Here, we show data where circulating microRNAs were found in bone marrow serum from acute lymphoblastic leukemia (ALL) patients. We investigated, by smallRNAseq, the presence of microRNAs in tumor microenvironment from four patients, collected at diagnosis (phase one), during chemotherapy treatment (phase two and three). After statistical analysis, only microRNAs with expression difference between treatment phases with a p value under 0.05 ($p \leq 0.05$) were considered. In totally, 946 known microRNAs were identified, 83 presented significant difference expression between phase one and two, 57 between phase one and three and 56 comparing phase two to three. miR-146a, identified in disease progression, with expression inversely proportional to patient survival, appeared with reads count decaying through chemotherapy phases, suggesting a reduction in the expression within cells and subsequent reduction in the microenvironment. miR-181, described in studies as participating in leukemia development also presented significativity alteration during chemotherapy treatment. The presence of circulating microRNAs in bone marrow of ALL patients shows their potential as tumor microenvironment regulators and biomarkers.

3420W

microRNA expression profiling in metastatic cutaneous squamous cell carcinoma. *A.E. Toland[1,2,3], L.E. Skeeles[2], S.B. Peters[3], T. Teknos[3], T. Olencki[3], D.C. Allain[1,3].* 1) Human Cancer Genetics and Comprehensive Cancer Center, Ohio State University, Columbus, OH; 2) MVIMG, Ohio State University, Columbus, OH; 3) OSU Wexner Medical Center, Columbus, OH.

Cutaneous squamous cell carcinoma (cSCC) is the second most common form of cancer with approximately 700,000 cSCCs diagnosed in the United States each year. The majority of cSCCs are readily treated surgically; however, 2-6% of all cSCCs metastasize leading to approximately 2500 deaths annually. Metastatic cSCC is associated with a survival rate of 56% with distant metastases and a worse outcome in recurrent cSCCs and immunocompromised individuals. No published studies have been performed to date to assess microRNA (miRNA) profiles of metastatic cSCC samples. In order to develop specifically targeted therapies for metastatic cSCC and to determine which factors are predictive of metastatic cSCC from early stage lesions, it is crucial to understand the molecular events which lead to metastasis of cSCC. We hypothesized that aberrant expression of miRNAs enhances the metastatic potential of cSCCs. To test this hypothesis, we performed miRNA expression analysis using the nCounter miRNA panel of approximately 800 miRNAs in 48 samples including 10 trios of normal, primary tumor and metastatic samples and 9 paired normal/non-metastatic primary cSCCs. We identified ~225 miRNAs expressed in the skin and performed analysis on the top 100 expressed miRNAs. Multiple miRNAs showed significant differences between primary cSCCs that did not metastasize compared to metastatic cSCCs or their corresponding primary tumor. These include *miR-let-7f*, *miR-4286*, *miR-15*, *miR-16*, *miR-29a* and *miR-29b*; all of which have been previously associated with metastatic cancers, suggesting the validity of these results. Interestingly, *miR-4286* is up-regulated in metastatic melanoma and shows a 6 to7-fold up-regulation in our metastatic cSCC samples. *miR-21* has been shown to be associated with metastasis in multiple tumors including esophageal SCC. Differentially expressed miRNAs are currently being evaluated in a second cohort of cSCCs and metastatic samples. In summary, several miRNAs show differential expression between non-metastatic and metastatic cSCCs; these may be useful as biomarkers for cSCC metastasis or as targets for therapy.

3421T

Prognostic significance of K-ras codon 12 mutation with resected Gall Bladder Cancer. *H.R. Kazmi[1], A. Chandra[1], D. Parmar[2], N. M[1].* 1) Surgical Gastroenterology, King George's Medical University, Lucknow, Uttar Pradesh, India; 2) Developmental Toxicology Division, Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India.

Introduction: Prognosis of Gall bladder cancer (GBC) is dismal due to aggressive behavior and lack of effective treatment. High incidence and mortality rates have been observed in Northern India. In present study, we had analyzed K-ras mutation in normal (undiseased) and GBC tissue samples and investigated its prognostic significance. Materials and Methods: Histopathologically proven GBC (n=39, stage II: n=22 and stage III: n=17) and normal gall bladder (n=24) tissue samples were studied. DNA was isolated and mutation was detected by Restriction Fragment Length Polymorphism analysis. Statistical analysis was carried out using chi-square test with Yates correction and Fisher exact test. The effect of K-ras mutation on patient survival was estimated using the Kaplan-Meier method and the difference between curves was analyzed by log-rank test. Results: The mean age (years + S.D.) of normal and GBC patients were 41.41+13.42 (range 19-65) and 43.87+12.39 (range 22-67) respectively. K-ras codon 12 mutation was observed in 1/24 (4.17%) and 16/39 (41.03%) of normal and GBC tissue samples respectively with statistically significant difference (p= 0.001).The overall survival time was significantly shorter (p=0.003) in patients having K-ras mutation. The median survival for GBC patients was shorter with K-ras mutation as compared to patients without K-ras mutation (12.5 vs. 17 months). Statistically significant difference in overall survival was also observed for stage II (p=0.012, median survival -15 months) and Stage III (p=0.009, median survival -8 months) GBC patients with K- ras mutation. Overall K-ras mutation significantly affects prognosis for GBC (Hazard ratio [HR] = 3.54; 95% Confidence interval [CI] = 1.54-8.14). For stage II GBC, K-ras positive patients carry 4.31 times higher risk of failure as compared to K-ras negative cases (95% CI = 1.37-13.55). Similarly, the HR for stage III disease was 7.42 (95% CI = 1.66-33.11). Conclusion: High frequency of K-ras codon 12 mutation in GBC is present in Northern Indian population. Its presence indicates poorer outcomes in patients undergoing surgery for this disease.

3422F

"Genetic polymorphisms in the apoptotic-associated genes Fas (-670 ag) and Fas L (-844 tc) & acute myeloid leukemia risk". *Vuree. Sugunakar[1], C. Anuradha[1], Dunna. Nageswara rao[4], G. Manjula[1], EM. Prajitha[1], D. Raghnadharao[3], Sinha. Sudha[2], Atili. Venkat Satya Suresh[2], Satti. Vishnupriya[1].* 1) Genetics and Biotechnology, Osmania University, Hyderabad, Andhra Pradesh, India; 2) MNJ Institute of Oncology, Red Hills, Hyderabad; 3) NIMS, Panjagutta, Hyderabad; 4) Sastra University, Thanjavur, Tamilnadu.

Acute Myeloid Leukemia (AML) is a cancer of myeloid lineage of blood cells, characterized by the rapid growth of abnormal white blood cells that accumulate in the bone marrow and interfere with the production of normal blood cells. FAS (TNFSf6/CD95/APO-1) belongs to the family of tumor necrosis factor receptors, and binding to the receptor by the FAS ligand (CD95L) triggers receptor trimerization and subsequent assembly of the death-inducing signaling complex. This gene is silenced in many tumor types, resulting in an inability to respond to proapoptotic signals. The FAS promoter is polymorphic, including an A to G substitution at -670bp, and T to C substitution at -844bp, which occur within SP1 and signal transducers and activators of transcription 1 transcription factor binding sites, respectively. The present study on the association of Fas Receptor (Fas R) and Fas Ligand (Fas L) promoter polymorphisms with AML were analyzed by PCR-RFLP method. GG genotype of Fas -670 AG showed a significant association with respect to the AML as well as with epidemiological variables like Sex of the proband, Area of living (Urban), WBC (30,000), no significant trend was observed with platelet count and Complete Remission. Whereas in FasL-844TC, C allele of FasL-844 was elevated in cases when compared with controls. Further, MDR analysis was performed between the two promoter SNPs which showed a significant trend.

3423W

TRAPing Telomerase Activity Using Droplet Digital PCR (ddPCR). *D.N. Shelton[1], J. Lin[2], J.F. Regan[1], G. Karlin-Neumann[1], E.H. Blackburn[2].* 1) Bio-Rad Laboratories, Digital Biology Center, Pleasanton, CA; 2) UCSF School of Medicine, Dept. of Biochemistry and BioPhysics, San Francisco, CA.

The aim of this work was to develop a more sensitive and high throughput assay for measuring telomerase activity. Telomeres are the protective structures at the ends of chromosomes consisting of 6 bp repeat sequences. In young cells, these regions can be as long as 15kb and act as caps which protect the DNA ends. These ends naturally degrade with each passing cell division, usually losing 25-200 base pairs per division. Once they are shortened below a critical length (estimated to be 200-300 bp) the cells arrest and become senescent, or 'old'. Telomeres can be thought of as a cellular or mitotic clock. Once the clock has wound down, the cells either die or pass through crisis and become immortal. One of the mechanisms of immortality is the activation of the enzyme, telomerase. Telomerase is the endogenous reverse transcriptase responsible for adding repeats to telomeres, rewinding the clock and enabling a cell to continuously divide. Abundant telomerase activity is found in the majority of cancers, fetal and adult stem cells, and germ cells. It is also present at much lower levels in non-pluripotent cells, such as immune cells, but these levels are currently difficult to measure. The telomerase repeat amplification protocol (TRAP) measures the presence of active telomerase by measuring the activity of the enzyme on a starting template, which is then amplified by PCR. For samples with abundant telomerase activity, SYBRGreen® qPCR assays provide high throughput. However, the current most sensitive method of detection still uses radioactivity and laborious PAGE sequencing gels followed by densitometry to quantify telomerase. In this study, the PAGE detection method is replaced by single molecule counting of telomerase-extended templates using droplet digital PCR technology.

3424T

Splicing of HRAS exon 2 is vulnerable - The splicing efficiency of activating mutations in codons 12 and 13 determines Costello syndrome phenotype. *B.S. Andresen[1], A.M. Hartung[1], J. Swensen[2], I.E. Uriz[1], M. Lapin[1], J.C. Carey[3], A. Calhoun[4], P. Yu[2], C.P. Vaughn[2], S.F. Dobrowolski[5], M.R. Larsen[1], H. Hanson[3], D.A. Stevenson[3].* 1) Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark; 2) University of Utah School of Medicine, ARUP Laboratories, Salt Lake City, Utah; 3) Division of Medical Genetics, University of Utah School of Medicine, Salt Lake City, Utah; 4) Division of Pediatric Genetics and Metabolism, University of Minnesota, Minneapolis, Minnesota; 5) Department of Pathology, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania.

Costello syndrome (CS) is most frequently caused by a c.34G>A (p.G12S) activating mutation in HRAS with modest transforming activity. G12V mutations have the highest transforming activity, the greatest frequency in cancers, but are very rare in CS. So far, all CS patients with p.G12V mutations (c.35G>T; c.35_36delinsTT; c.35_36delinsTA), have had a severe, early lethal, phenotype. Sequence analysis of a 12-year-old boy with an attenuated CS phenotype revealed, to our surprise, a new germline p.G12V mutation, c.35_36delinsTG, without evidence of mosaicism. In silico analysis shows that exon 2 has a weak 3' splice site and that c.35_36delinsTG simultaneously abolishes exonic splicing enhancer (ESE) motifs and creates exonic splicing silencer (ESS) motifs indicating that it may disrupt splicing. Analysis of patient HRAS cDNA showed that c.35_36delinsTG results in exon 2 skipping and consequently little mutant protein, explaining the attenuated phenotype. Transfection of several different cell lines with HRAS mini genes with the four known p.G12V mutations and the common c.34G>A mutation showed that only c.35_36delinsTG results in exon 2 skipping. Deletions of nucleotides c.32-37 cause exon 2 skipping, indicating presence of an ESE in this region. Optimization of the weak 3' splice site corrected splicing from the mutants confirming vulnerability of exon 2. Testing of wild type and several mutant HRAS exon 2 sequences in different splicing reporter minigenes showed that ESE strength is increased by the other p.G12V mutations and by other mutations with strong transforming potential, whereas c.35_36delinsTG inactivated splicing. RNA affinity purification, ITRAQ labeling followed by MS/MS, showed that c.35_36delinsTG increases binding of hnRNPF/H splicing inhibitory proteins, whereas c.35G>T increased binding of several splicing stimulatory SR proteins consistent with the observed effect on splicing. Replacement in the HRAS minigene of the wild type sequence with hnRNPF/H binding ESS motifs confirmed that binding of hnRNPF/H results in exon 2 skipping. Our study illustrates that phenotype of Costello syndrome and likely somatic cancers are not only determined by the transforming potential of the mutant HRAS protein, but is also determined by the efficiency of exon 2 inclusion. This has important implications for our understanding of the correlation between genotype and phenotype in diseases caused by HRAS mutations and for development of new therapies.

3425F

Familial intracranial meningioma without *NF2* and *SUFU* mutations: searching for additional predisposing genes.
M. Aavikko[1], H. Ristolainen[1], E. Kaasinen[1], P.A. Koivisto[2], M. Pöyhönen[3], K. Claes[4], T. Van Maerken[4], M. Artama[5], E. Pukkala[5], L.A. Aaltonen[1], P. Vahteristo[1]. 1) Department of Medical Genetics, Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 2) Seinäjoki Central Hospital, Department of Neurology, Seinäjoki, Finland; 3) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 4) Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium; 5) The Finnish Cancer Registry, Helsinki, Finland.

Meningiomas are the most common tumors of the central nervous system (CNS) accounting for one third of all primary CNS tumors in adults. Majority of meningiomas are slowly growing benign tumors arising from the leptomeninges covering CNS. Most commonly they occur sporadically, but they are also commonly seen in individuals with Neurofibromatosis type 2 tumor suppressor syndrome.

We recently reported a candidate predisposing gene defect in the *Suppressor of Fused homolog* (*SUFU*) in a family of five siblings with intracranial meningiomas, four of whom had multiple lesions. All seven studied meningiomas from the family displayed loss of the wild type allele following the two-hit model for tumor suppressor genes. In addition, we showed that the mutant SUFU had lost its ability to bind GLI1 transcription factor causing dysregulated Hedgehog signaling. Mutations in *SUFU* have previously been identified in familial and sporadic medulloblastomas. Our results suggested that in addition to medulloblastomas *SUFU* defects may also predispose to a subset of meningiomas.

Here we describe a Finnish *NF2* and *SUFU* mutation negative family of five affected individuals with intracranial meningiomas. The family display autosomal dominant inheritance pattern with affected individuals in three generations. We have performed genome-wide SNP genotyping and linkage analysis in combination with exome and genome sequencing. We have identified candidate chromosomal regions and gene variants in this family and we are studying the pathogenicity of these variants in the meningioma tumors from the family and in additional familial and sporadic intracranial meningioma cases, these include a Belgian family with four affected individuals in four generations and Finnish familial and sporadic meningioma patients identified through Finnish Cancer Registry. Increased understanding of meningioma predisposition is of great clinical importance in providing tools for better diagnosis and management of the patients.

3426W

***DICER1* Mutations in Pituitary Blastoma: new gene, new disease.**
L. de Kock[1, 2], N. Sabbaghian[2], F. Plourde[2], A. Srivastava[2], D. Bouron-Dal Soglio[3], N. Hamel[4], J.H. Choi[5], S.H. Park[6], C.L. Deal[7], M. Dishop[8], A. Esbenshade[9], T. Jacques[10, 11], A. Perry[12], H. Leichter[13], P. Maeder[14], M.A. Brundler[15, 16], J. Neal[17], M. Zacharin[18], M. Korbonits[19, 20], T. Cole[21], S. Albrecht[22], E. Horvath[23], K. Kovacs[23], J.R. Priest[24], W.D. Foulkes[4]. 1) McGill University, Montreal, QC, Canada; 2) Lady Davis Institute, Segal Cancer Centre, Jewish General Hospital, Montreal, QC, Canada; 3) Department of Pathology, CHU-Sainte Justine, Montréal, QC, Canada; 4) Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montreal, QC, Canada; 5) Department of Pathology, Yeungnam University College of Medicine, Daegu City, South Korea; 6) Seoul National University, College of Medicine, Seoul, Republic of Korea; 7) Department of Endocrinology, CHU Ste-Justine, Montreal, QC, Canada; 8) Department of Pathology and Laboratory Medicine, Children's Hospital Colorado, USA; 9) Department of Pediatrics, Monroe Carell Jr. Children's Hospital, Nashville, TN, USA; 10) Neural Development Unit, UCL Institute of Child Health, London, UK; 11) Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; 12) Departments of Pathology and Neurological Surgery, UCSF Medical Centre, San Francisco, CA, USA; 13) Department of Pediatrics, Olgahospital, Dresden, Germany; 14) Department of Neuroradiology, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 15) University of Calgary, Alberta, Canada; 16) Alberta Children's Hospital, Alberta, Canada; 17) Department of Histopathology, UHW, Cardiff, UK; 18) Department of Endocrinology and Diabetes, Royal Children's Hospital, Flemington Road, Parkville, Australia; 19) Department of Endocrinology, Barts and the London School of Medicine, London, UK; 20) Queen Mary University of London, London, UK; 21) Birmingham Women's NHS Foundation Trust, Edgbaston, Birmingham, West Midlands, UK; 22) Department of Pathology, Montreal Children's Hospital, McGill University Health Centre, Montreal, QC, Canada; 23) Department of Laboratory Medicine and Pathobiology, St. Michael's Hospital, University of Toronto, Toronto, Canada; 24) Minneapolis, Minnesota, USA.

BACKGROUND: DICER1, a non-coding small RNA processing enzyme, cleaves micro RNA (miRNA) precursors into mature miRNAs, which are known to regulate mRNA expression. Germ-line mutations in *DICER1* have recently been found to predispose to a rare cancer syndrome consisting of primarily childhood tumours. The main manifestations of the syndrome include pleuropulmonary blastoma (PPB), cystic nephroma (CN), Sertoli-Leydig cell tumours (SLCT), multinodular goiter and other rare childhood sarcomas and dysplasias. We and others have shown that germ-line *DICER1* mutations are often accompanied by specific somatic mutations in the *DICER1* RNase III catalytic domain in several tumour types, and we hypothesized that both germ-line and somatic mutations might also occur in children with pituitary blastoma (PitB). PitB, identified as a distinct entity in 2008, is a very rare, potentially lethal childhood brain tumour. Since the discovery by our team of an inherited mutation in *DICER1* in a child with PitB in 2011, we have identified a further 12 pathologically-confirmed PitB cases. Histological studies suggested perturbation of pituitary stem cells could underlie the pathogenesis of PitBs. OBJECTIVES: We aim to determine the contribution of germ-line and somatic *DICER1* mutations to PitB and to analyse the effect of the mutations on the mi/mRNA landscape. We hypothesize that targets involved in the regulation of pituitary stem cells will be dysregulated in these tumours. RESULTS: To date, 6 of 6 PitB cases analysed genomically harbour a germ-line *DICER1* mutation, suggesting that mutation of DICER1 is a key predisposing genetic event. Furthermore, we identified likely contributory "second hits" in *DICER1* in 6 of the 7 PitBs tested thus far for somatic mutations. Five of these 6 mutations occur within the RNase IIIb catalytic domain, a domain essential to the generation of miRNAs from the 5' arm (5p) of precursor miRNAs. Work is underway to complete the analysis of the remaining samples and to determine the effect of these *DICER1* mutations on the generation of miRNAs. CONCLUSION: Germ-line *DICER1* mutations are a major contributor to PitB. Second somatic 'hits' in *DICER1*, within the RNase IIIb domain, also appear to be critical steps in their pathogenesis. Further studies will focus on mRNA, miRNA and gDNA profiling of these rare tumours.

3427T

Targeted sequencing of *MLH1*, *MLH2*, and *MSH6* in defective mismatch repair colorectal cancer cases with no identified mutation. *M. DeRycke[1], S. Gunawardena[2], S. McDonnell[1], S. Middha[1], S. Riska[1], Z. Fogarty[1], B. Eckloff[3], D. Schaid[1], E. Goode[1], N. Lindor[4], S. Thibodeau[1], Colon Cancer Family Registry.* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Advanced Genomics Technology Center, Mayo Clinic, Rochester, MN; 4) Health Sciences Research, Mayo Clinic, Scottsdale, AZ.

Approximately 25% of colorectal cancers (CRC) are clustered in families expected to harbor predisposing genetic mutations. Of all CRC cases, ~3-5% are caused by Lynch Syndrome (LS), an autosomal dominant condition resulting from mutations in the mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*. Currently, when LS is suspected, germline testing of the proband can be done to determine the underlying mutation and facilitate testing relatives who may be at increased risk. Typically, the entire coding region and exon/intron boundaries of the gene are targeted for sequencing, and identified variants found are classified as deleterious, variants of uncertain significance (VUS), or common polymorphisms. A significant fraction of cases show loss of protein expression by immunohistochemistry but no detected variant or deletion in the affected gene, presenting a problem for families with suspected LS: no mechanisms exist to determine which family members share the variant and are at increased risk for developing CRC. In this study, we completed targeted sequencing of *MLH1*, *MLH2*, and *MSH6* in 130 individuals with defective MMR CRC in which previous sequencing studies failed to detect any disruptive germline MMR variants. We targeted exonic, intronic, and nearby intergenic regions of *MLH1*, *MSH2*, *MSH6*, as well as the 3' end of *EPCAM*. Agilent's SureSelect Custom Capture system was used to target the genes and samples were indexed then pooled for sequencing on a HiSeq 2000. Samples underwent extensive quality control, which resulted in the exclusion of 10 samples due to low coverage (n=7) or high read duplication (n=3). Mean coverage for the remaining 120 samples was 633x (range 164 - 3,114x; median 367x). Transition-Transversion ratios for known and novel SNPs were 2.34 and 2.29, respectively. Samples had an average of 3,124 SNPs (range 2,639 - 4,168) and 386 (range 249 - 464) INDELs in the capture region. Previously completed genotyping results were compared with sequencing results for 118 samples, resulting in 99.6% concordance. Preliminary analysis of the 85 cases with loss of MLH1 protein expression identified the -93G>A transition in 45 cases, with six individuals being homozygous recessive. This variant has been reported to reduce expression of MLH1 and has been considered a low penetrance mutation contributing to CRC development. Further results from the analysis will be presented.

3428F

Annotation of Rare Variants from Exome Sequencing in Families with Lymphoid Malignancies. *L.R. Goldin[1], M.L. McMaster[1], M. Rotunno[1], J. He[2], L. Burdette[2], A. Hutchinson[2], J. Boland[2], M. Yeager[2], M.A. Tucker[3], S.J. Chanock[2], N.E. Caporaso[1].* 1) Genetic Epidemiology Branch, DCEG/NCI, Bethesda, MD; 2) Cancer Genomics Research Laboratory, DCEG/NCI, Bethesda, MD; 3) Human Genetics Program, DCEG/NCI, Bethesda, MD.

Several variants with relatively small effects have been identified from GWAS of lymphoid malignancies but single high penetrance genes have not been identified. We have conducted exome sequencing in 178 individuals from 49 families at high risk for lymphoid malignancies including chronic lymphocytic leukemia, Hodgkin lymphoma, and non-Hodgkin lymphomas. We used Nimblegen v2.0 and v3.0 for exome capture followed by sequencing on the Illumina HiSeq2000. We required that 80% of coding sequences had at least 15x coverage. Novoalign v.2.07.14 was used for alignment and GATK was used for local re-alignment and variant calling. We sequenced 3 or more patients or obligate carriers from the families. A large number (~1000) of rare (<1% frequency in European populations) non-synonymous variants were shared among patients/carriers in each family. In order to eliminate other possible sequencing artifacts, variants found in more than 1% of samples sequenced in our laboratory from other studies were excluded, leaving ~20-100 variants per family. We prioritized variants for further validation and follow-up by looking for genes with the same shared variant in more than one family that were predicted to be damaging by at least one program (or were in a regulatory region), were highly conserved, and had a literature link to cancer or a cancer-related pathway. The program, Ingenuity Variant Analysis, was used to facilitate filtering and annotation. Variants in 7 genes (*BORA, FAF1, ITGB2, PRDM2, TXNDC17, ZNF189, and ELMO3*) met these conditions. We will conduct targeted sequencing of these genes in 366 patients from an additional 170 multiplex families. There were an additional 40 genes where different variants from the same gene were found in more than 1 family and selected genes from this group will also be sequenced in the new sample. Results of our validation/follow-up studies will be presented.

3429W

Constitutional mismatch repair deficiency syndrome caused my MMR gene founder mutations with a high prevalence in Israel. *I. Kedar-Barnes[1], Z. Levi[2], H. Toledano[3], M. Halpern[4], I. Lerer[5], T. Peretz[6], R. Kariv[7], E. Half[8], N. Magal[1], K. Wimmer[9], Y. Goldberg[6], D. Berkovich[10], H.N. Baris[1].* 1) The Raphael Recanati Genetics Inst, Rabin Med Ctr, Petah Tikva, Israel; 2) Gastroenterology Division, Rabin Medical Center, Beilinson Hospital , Petach Tikva, Israel; 3) Pediatric Oncology, Schneider Children's Medical Center, Petach Tikva, Israel; 4) Department of Pathology, Rabin Medical Center, HaSharon Hospital, Petach Tikva, Israel; 5) Deprtment of Human Genetics, Hadassa-Hebrew University Medical Center, Jerusalem, Israel; 6) Sharett Institute of Oncology, Hadassa-Hebrew University Medical Center, Jerusalem, Israel; 7) Deaprtment of Gastroenterology& Liver Disease, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; 8) Gastroenterology Department, Rambam Health Care Campus, Haifa, Israel; 9) Division of Human Genetics, Medical University Innsbruck, Austria; 10) GGA-Galil Genetic Analysis, Kazerin, Israel.

Heterozygous germline mutations in one of the 4 mismatch repair (MMR) genes, *MLH1, MSH2, MSH6* and *PMS2* cause Lynch syndrome (LS), an autosomal dominant cancer predisposition syndrome conferring a high risk of colorectal, endometrial and other cancers in adulthood. Offsprings of couples, both having LS, have a high risk to inherit biallelic MMR gene mutations. The cause constitutional MMR deficiency (CMMRD), a severe recessively inhereted childhood cancer syndrome with a very broad tumor spectrum including mostly hematological malignancies, brain tumors and childhood colon cancer. Most of CMMRD children also present with café au lait spots and axillary freckling mimicking Neurofibromatosis type I. We describe our experience in Israel with 5 CMMRD families. The clinical presentation included: brain tumors at age 2-19 years, colon cancer at ages 9-20 years, and one patient with lymphoma at age 12. In two non-consangineous Ashkenazi families, the common founder Ashkenazi mutation were detected: one family was homozygous for c.1906G>C in *MSH2* and the 2nd family heterozygous for c.3984_3987dupGTCA and c.3959_3962delCAAG in *MSH6*. In the 3 other consanguineous families, different homozygous mutations were identified: c.2192T>G (p.L731X) in *PMS2*, a recurrent mutation among Iranian Jews, *PMS2* c.686_687delCT and *MSH6* 3603_3606delAGTG were found in large pedigrees of Bedouin and Arab ancestries, respetively. Given the prevalence of these mutations among Israeli population, we want to raise the awareness of CMMRD syndrome and open a discussion regarding screening for MMR founder mutations among spouses of LS patients for the purpose of prenatal diagnosis.

3430T

Magnetic Resonance Imaging screening In Li Fraumeni Syndrome: An exploratory whole body MRI study (the SIGNIFY study). *E. Killick[1, 2], E. Bancroft[3], N. Taylor[3], DG. Evans[4], M. Leach[1, 3], R. Eeles[1, 3], The SIGNIFY collaborators and steering committee.* 1) Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) Cancer Care, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom; 3) Royal Marsden NHS Foundation Trust, Chelsea, London, United Kingdom; 4) Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdon.

Background Li Fraumeni Syndrome predisposes individuals to a range of different malignancies with a lifetime cancer risk of up to 90% in women and 70% in men. Current national screening recommendations are for dual modality breast screening with mammography and MRI in women. Some centres employ family specific screening tailored to malignancies found at increased frequency within certain families with most centres having an 'open-door' policy. Recent evidence suggests there may be a survival benefit for more intensive screening, including whole body MRI; there are no published data on the psychological impact of such screening programmes. Aims The primary end-point is to assess the incidence of malignancies diagnosed in asymptomatic TP53 mutation carriers using whole body MRI, against general population controls. The secondary end-points are to assess the incidence of non-malignant relevant disease, to assess the incidence of irrelevant findings and the investigations required to determine relevance of MRI findings and finally to assess the psychological impact of whole body MRI screening in TP53 mutation carriers. Methodology We will recruit 44 TP53 mutation carriers and 44 population controls aged 18 - 60 years, who will undergo conventional and diffusion weighted whole body MRI. The MRI will be reported independently by two radiologists blinded to the mutation status of the individual, with a consensus read with a third radiologist in cases of discordance. The results will be relayed to the individual and any necessary further investigations arranged. An intersite virtual MDT will be held at regular intervals to discuss the diagnosis and management of incidental findings. A series of questionnaires will be used from time of recruitment up to a year after the MRI scan to assess the psychological impact of screening. Results The study recently opened to recruitment; thus far 4 individuals have been enrolled. Results from these and further recruits will be shown.

3431F

Characterization of RB1 mutations and incidence of undetected defects in retinoblastoma. *P.S. Lai[1], A.P. Alcasabas[1], G. Sundar[2], B.L. Quah[3].* 1) Dept Pediatrics, National Univ Singapore, Singapore, Singapore; 2) Dept Ophthalmology, National University Hospital, Singapore; 3) Singapore National Eye Centre, Singapore.

Retinoblastoma (RB) is a retinal tumor associated with biallelic loss of RB1 gene. Most cases are diagnosed by five years of age and occur in both hereditary and sporadic forms. Bilateral RB cases are usually associated with constitutional or germline mutations while approximately 15% of sporadic unilateral cases may carry heritable constitutional mutations. Recently, there has been speculation that some rare RB tumors may arise by mechanisms other than RB1 mutations. In this study, we characterized RB1 mutations in 40 retinoblastoma cases in order to determine the incidence of non-detectable RB1 mutations. To detect large deletions, segregation analysis using polymorphism makers, quantitative multiplex PCR or MLPA assays were carried out while small point mutations were detected by sequencing analysis. RB1 promoter methylation was analysed by methylation-specific PCR. RESULTS: Large deletions were present in 27 out of 40 (67.5%) tumors. Germline mutations were detected in 11 out of 13 bilateral cases (84.6%). Among the 27 unilateral RB cases, germline mutations were detected in 4 cases (14.8%). The mutation spectrum shows RB1 mutations (7 nonsense and 2 frameshift mutations) in 16 cases as distributed among exons 12-22 which codes for the RB protein pocket domain known to be critical for transcriptional repression. The most frequent mutations were nonsense (24/35; 68.6%) followed by frameshifts (7/35; 20%) with splice site and missense mutations contributing towards the remaining genetic defects. Most of the mutations identified in this study correspond to C to T transitions in 7 CGA-arginine codons in exons 8, 10, 11, 14, 15, 17 and 23 of the RB1 gene. No mutations were found in four other CGA codons located in exons 1, 18 and 27. These findings correspond with previously published data analysed by meta-analysis. RB1 promoter hypermethylation was rare and observed only in 1 tumor. CONCLUSIONS: Mutation in only one RB1 allele was detected in 5 cases while no RB1 mutations on both alleles were found in 2 unilateral cases. This could be due to (1) undetected translocations, deep intronic mutations, or alterations in unknown RB1 regulatory regions, (2) presence of low level RB1 mosaic mutations not detectable by Sanger sequencing, or (3) involvement of other genetic mechanisms or genes. Further investigations are warranted in these 5 cases.

3432W

Exome Sequencing of Family with Carcinoid Cancers. *D.W. Neklason[1, 2, 3], G.W. Gilcrease[1,2], N.R. Sargent[1], A. Snow[1], C. Teerlink[2,3], M. Bailey[3], R.W. Burt[1,2].* 1) Huntsman Cancer Inst, Univ Utah, Salt Lake City, UT, 84112; 2) Department of Medicine, University of Utah, Salt Lake City, UT 84112; 3) Division of Genetic Epidemiology, University of Utah, Salt lake City, UT 84112.

We describe here genetic study of a family with carcinoid tumors of the gastrointestinal tract. Two siblings and a first cousin were diagnosed with carcinoid tumors at ages 41, 63, and 59. Carcinoids are rare neuroendocrine tumors and most neuroendocrine tumors are not familial. Incidence of neuroendocrine tumors is 2.47 per 100,000 with the rate steadily increasing over the past 20 years. The majority of neuroendocrine tumors occur in the GI tract (67%), primarily in the small intestine as represented in this family. Carcinoid tumors derive from enterochromaffin cells of the neuroendocrine system in the gut. These cells contain a large amount of the body's store of serotonin and in response to stimuli in the lumen (chemical, mechanical, pathological), the serotonin release signals nausea to the brain. The two more advanced cases in the family were associated with carcinoid syndrome, which includes flushing, nausea and diarrhea from secretion of serotonin from the tumor. We performed whole exome sequencing using Agilent SureSelect capture on all three individuals with mean coverage of >60X on Illumina HiSeq. Genomes were aligned and variants called using standard best practices. Variants were filtered and prioritized using the VAAST (the Variant Annotation, Analysis & Search Tool) algorithm (Yandell et al, 2011) within the Omicia Opal 1.7.0 tool. Each individual had ~18,500 exonic mutations with a protein impact. No deleterious variants were identified in genes known to cause familial neuroendocrine cancers (MEN1, MEN2, VHL, NF1, TSC1, TSC2, and PPKAR1A). When scored as deleterious by a metaclassifier (Omicia Score >0.5), filtered for frequency <5% in the population, shared by all 3 family members, and not shared by a control family sequenced at the same time, 38 variants were identified. Three variants reside in large regions of shared genomic segments identified through SNP genotyping. Interesting genes include ones associated with metastasis (LRRC59), telomerase (TEP1), cell cycle (WDR35), and a neurotensin receptor (NTSR1). The identified variants will be verified by Sanger sequencing and examined in an unaffected sibling. Likely candidate genes will be discussed.

3433T

Frequencies of *BRCA1*, *BRCA2*, *PALB2*, and *CDKN2A* germline mutations in familial pancreatic cancer (FPC): A PACGENE study. *D.B. Zhen[1], K.G. Rabe[1], S. Gallinger[2], S. Syngal[3], A.G. Schwartz[4], M.G. Goggins[5], R.H. Hruban[5], M.L. Cote[4], K. Moyes[6], R.J. Wenstrup[6], A.R. Hartman[6], D. Seminara[7], A.P. Klein[5], G.M. Petersen[1].* 1) Mayo Clinic, Rochester, MN; 2) Toronto General Hospital, Toronto, Ontario, Canada; 3) Dana-Farber Cancer Institute, Boston, MA; 4) Karmanos Cancer Institute, Detroit, MI; 5) Johns Hopkins University, Baltimore, MD; 6) Myriad Genetic Laboratories, Inc, Salt Lake City, UT; 7) National Cancer Institute, Bethesda, MD.

Background: FPC (defined as at least two affected first degree relatives (FDR) in a kindred) accounts for 5-10% of pancreatic adenocarcinoma and is thought to be genetically heterogeneous. Genetic testing is available for germline mutations in several genes known to predispose to FPC. However, our knowledge is limited because published studies to date have been small and have been non-systematic.

Methods: The Pancreatic Cancer Genetic Epidemiology Consortium (PACGENE) has assembled a registry of 1747 kindreds containing two or more relatives affected with pancreatic cancer. Five PACGENE sites tested germline DNA samples from 234 unrelated cases affected with pancreatic adenocarcinoma (80 met criteria for FPC) for four genes: *BRCA1* and *BRCA2* (including analysis of deletions and rearrangements), *PALB2*, and *CDKN2A*. The cases who were tested were not previously known to carry germline mutations in these genes. We also studied mutation frequencies between FPC versus familial non-FPC cases (at least two affected blood relatives, but not FDR).

Results: Frequencies for deleterious/suspected deleterious (D/S) mutations and variants of uncertain significance (VUS) among **FPC cases** were: *BRCA1* 2/68 (2.9%), no VUS; *BRCA2* 3/68 (4.4%), no VUS; *PALB2* 1/77 (1.3%), no VUS; *CDKN2A* 4/80 (5.2%), 3 VUS. Among **Familial non-FPC** cases, frequencies were: *BRCA1* 0/123, no VUS; *BRCA2* 1/123 (0.8%), 1 VUS; *PALB2* 1/147 (0.7%), 4 VUS; *CDKN2A* 0/153, 3 VUS. FPC cases carry more mutations than familial non-FPC cases. Because we had excluded known mutations in our registry from this study, these frequencies may be a lower estimate. No case carried more than one mutation.

Conclusions: This is the first comparative study of four susceptibility genes in a large sample of FPC cases. The data support genetic heterogeneity of FPC. Genetic testing for cases meeting FPC criteria may be warranted, but multiple genes must be evaluated. The genetic basis for the majority of FPC remains to be determined.

3434F

Common Somatic Variations Identified in Maffucci Syndrome. *M. Amyere[1], V. Wouters[1], A. Dompmartin[2], O. Enjolras[3], I. Kaitila[4], P.L. Docquier[5], J.B Mulliken[6], L.M Boon[1,7], V. Vikkula[1,8].* 1) Human Molecular Genetics, de Duve Institute, Brussels, Belgium; 2) Université de Caen Basse Normandie, CHU Caen, Department of Dermatology; 3) Consultation des Angiomes, Hôpital Lariboisière, Paris, France; 4) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland; 5) Division of Orthopedic Surgery, Cliniques Universitaires St-Luc, Université catholique de Louvain, Brussels, Belgium; 6) Department of Plastic Surgery, Children's Hospital and Harvard Medical School, Boston, MA, USA; 7) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques Universitaires St-Luc, Université catholique de Louvain, Brussels, Belgium; 8) Walloon Excellence in Lifesciences and Biotechnology (WELBIO), Université catholique de Louvain, Brussels, B-1200, Belgium.

Maffucci syndrome (MS) is a rare congenital disorder characterized by multiple central cartilaginous tumors (enchondromas) that are accompanied by cutaneous spindle-cell hemangiomas. These patients have a high incidence of malignant transformation. No familial case is known and the etiopathogenic cause remains unknown. In enchondromatosis (Ollier's disease, OD), which is comprised of enchondromas only, four mutations in the PTHR1 gene have been identified in four patients; three were somatic and one was germline. No PTHR1 mutation has been detected in MS. On the other hand, in 77% of patients with MS and 81% of patients with OD, somatic IDH1 and more rarely IDH2 mutations have been observed. These changes are shared with other tumors, including glioblastomas, leukemias and thyroid cancers, and seem to be markers of cellular transformation. To search for underlying somatic genomic causes, we screened MS tissues using Affymetrix SNP-chips. We looked for copy number variations (CNV), loss of heterozygosity (LOH) and uniparental (iso)disomy (UPD) by performing pairwise analyses between allele intensities in tumoral DNA versus the corresponding blood-extracted DNA. While common chromosomal anomalies were absent in constitutional DNA, several shared CNVs were identified in tumors. The most frequently observed somatic alterations are localized in 2p22.3, 2q24.3 and 14q11.2. In the single chondrosarcoma studied, large chromosomal amplifications and/or deletions were observed in chromosomes 3, 6, 9, 10, 12, 13 and 19. Some of these have been reported in other chondrosarcomas underscoring their etiopathogenic role. No LOH/UPD was observed in any Maffucci tissue. Our findings identify frequent somatic chromosomal rearrangements, implicating 2p22.3, 2q24.3 and 14q11.2 in the formation of enchondromas and spindle cell hemangiomas in Maffucci syndrome. WES are performed on MS tissues and blood paired and variations are in validation process to identify Maffucci causative mutations.

3435W

A survey of *HRAS* mutations in a large cohort of patients with Costello Syndrome reveals a predisposition for embryonal rhabdomyosarcoma with paternal uniparental disomy. *K.M. Robbins[1,3], D.L. Stabley[1], A. Sadreameli[1], J. Holbrook[1], S. McCahan[1], R. Sahraoui[1,4], K.W. Gripp[2], K. Sol-Church[1].* 1) Center for Pediatric Research, AI duPont Hospital for Children Wilmington , DE; 2) Medical Genetics, AI duPont Hospital for Children, Wilmington, DE; 3) University of Delaware Biological Sciences, Newark, DE; 4) Delaware State University Biological Sciences, Dover, DE.

Costello Syndrome (CS) is a rare condition resulting in failure to thrive, intellectual disabilities, short stature, coarse facial features, skeletal abnormalities, and congenital heart disease. CS is a rasopathy, caused by heterozygous mutation in the *HRAS* oncogene. Though most mutations are *de novo*, originating in the paternal germline, we identified a few cases of somatic mosaicism. CS patients have an increased risk for malignancies, specifically embryonal rhabdomyosarcoma (ERMS). In the general population, non-syndromic (NS) ERMS, is a relatively common pediatric cancer. It is rarely associated with *HRAS* mutations, but is characterized at the molecular level by loss of heterozygosity (LOH) at 11p15.5. The goal of this study was to molecularly evaluate our cohort, identify dysregulated pathways in CS skin fibroblasts and identify genetic alterations associated with ERMS in CS patients and in NS ERMS. Parental origin of *HRAS* germline mutations was established using either restriction enzyme digestion or allele specific PCR amplification and sequencing. Differential gene expression analysis was performed with Affymetrix Human Gene 1.0 ST arrays, 6 CS patient derived fibroblast lines and 6 control fibroblast lines. Characterization of the genetic lesions in ERMS was performed using chromosome 11 STR markers, fluorescence *in situ* hybridization using centromere 11 probe, as well as Affymetrix CytoScan arrays on CS and NS ERMS tumors and cell lines. We molecularly identified *HRAS* mutations in 104 CS patients. There were 63 cases of paternally derived mutations, 3 maternally derived and the remainder was non-informative. Eleven unrelated patients developed ERMS, and a few relapsed due to residual disease or *de novo* tumor formation. Nine ERMS tumors from 6 unrelated CS patients were available for molecular evaluation. All patients carried germline paternally derived p.G12S (4) or p.G12A (2) mutations in HRAS. We established a pure ERMS primary cell line using a fresh tumor sample from a patient with a p.G12A germline mutation. Loss of imprinting of *H19* and **CDKN1C** at 11p15.5 was identified, and additional functional analysis and drug screens are currently ongoing. Using CS and NS ERMS tumors, complete LOH was observed along chromosome 11. In CS samples, paternal uniparental disomy is observed in all but one recurrent primary tumor. This molecular characterization will aid in developing new treatments for ERMS.

3436T

Clinical characteristics and genotype-phenotype correlations in a large cohort of Brazilian Li-Fraumeni syndrome patients. *M. Achatz[1, 2], C. Sagne[3], P. Ashton-Prolla[4, 5], V. Marcel[3, 6], A. Nobrega[1], J. Hall[3, 6], P. Hainaut[7].* 1) Oncogenetics, A.C. Camargo Cancer Center, Sao Paulo, Sao Paulo, Brazil; 2) National Institute of Science and Technology in Oncogeneomics (INCITO), São Paulo, Brazil; 3) INSERM U612, Centre Universitaire, Orsay, France; 4) Department of Genetics and Molecular Biology, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil; 5) National Institute of Science and Technology in in Populational Medical Genetics (INAGEMP), Porto Alegre, Brazil; 6) Institut Curie, Centre de Recherche, Centre Universitaire, Orsay, France; 7) International Prevention Research Institute, Ecully, France.

Li-Fraumeni syndrome (LFS), an inherited cancer predisposition syndrome, is associated with germ line mutations in TP53. It is characterized by high risk of multiple, early cancers in children and young adults. Carriers of a mutation in DNA-binding domain have a lifetime risk of 90% to develop a cancer. In Brazil, a variant form of LFS is exceedingly frequent due to a widespread founder TP53 mutation, p.R337H, detected in 0.3% of the general population in South and Southeastern Brazil. This mutation occurs in p53 oligomerization domain and its effect is supposed to be dependent upon pH conditions. Individuals with the TP53 p.R337H mutation have lower cumulative lifetime cancer risk (50-65%), with a number of mutation carriers being cancer-free. To identify genetic modifiers of penetrance in TP53 mutation carriers, we have analyzed a panel of 87 single-nucleotide-polymorphisms in the TP53 locus in 402 members of Brazilian families with LF traits, with or without TP53 mutations. We have identified two insertion variants associated with a 12-18 years delay in age at first cancer diagnosis, PIN3 (16bp duplication in intron 3; rs17878362 p=0.082) and DUP3'UTR (6bp duplication in 3'UTR; rs17880560, p =0.067). These SNP are known to regulate p53 pre-mRNA, suggesting that variations in pre-mRNA dynamics may accelerate cancer onset in TP53 mutation carriers. Genotyping these SNP may help in defining personalized follow-up strategies in Brazilian carriers of p.R337H, who have more heterogeneous phenotypes than carriers of 'classic' TP53 mutations.

3437F

Identification of *BRCA1* and *BRCA2* mutations in mexican patients with inherited breast and ovarian cancer by full-exon pyrosequencing. *R.M. Alvarez[1], F. Vaca[2], V. Fragoso[2], S. Vidal[3], L.A. Herrera[4,5], D. Cantu[6], J.E. Bargallo[7], A. Mohar[8], C.G. Perez[1,2].* 1) Unidad de Genómica y Secuenciación Masiva, Instituto Nacional de Cancerología, Mexico , Distrito Federal, Mexico; 2) Laboratorio Oncogenómica, Instituto Nacional de Cancerología, Mexico; 3) Laboratorio de Diagnóstico molecular, Instituto Nacional de Cancerología, Mexico; 4) Unidad de Investigación Biomédica en Cáncer, Instituto Nacional de Cancerología; 5) Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México (UNAM), México; 6) Subdirección de Investigación Clínica, Instituto Nacional de Cancerología, México; 7) Departamento de Tumores Mamarios, Instituto Nacional de Cancerología, México; 8) Dirección General, Instituto Nacional de Cancerología, México.

It has been estimated that 10% of women diagnosed with breast cancer have a hereditary form of the disease, primarily caused by mutations in *BRCA1* or *BRCA2*, the most prevalent and high susceptibility genes for hereditary breast and ovarian cancer. Worldwide, it is common practice to offer high-risk patients genetic counseling and DNA testing. Even though, molecular approach of these genes is not a common practice in Latin American public institutions because the cost and time consuming of the methods used. We proposed the implementation of massive pyrosequencing for screening of BRCA mutations in patients with suspected hereditary breast/ovarian cancer. Therefore, we used this technology for evaluation of germline mutations in the entire exonic and splice site regions of *BRCA1* and *BRCA2* in 163 patients with breast and ovary cancer and with familial history of breast cancer or with clinical features suggestive for BRCA mutations. In a first step, the method was evaluated with positive and negative controls for BRCA1 and BRCA2 germline mutations. We found 30-550 reads per sequence and identified all the BRCA pathogenic mutations; the negative controls did not showed deleterious variants, confirming the suitability of their use. Then, we used in the 163 patients previously described. We found twenty-eight deleterious mutations (eighteen in BRCA1; ten in BRCA2), six of them, had not been reported previously. All the mutations identified were verified by Sanger sequencing. The frequency of BRCA mutations in these patients was 17%. Direct molecular study (PCR and Sanger sequencing) was extended to 55 relatives of index cases, where pathogenic mutations were found. Twenty percent of them were carriers of the familial mutation studied. Patients and their families were enrolled in a clinical follow-up with regular screening and the recommended risk reduction measures. Moreover, we discovered a 40% of missense variants; most of them had been reported as variants of unknown clinical significance (VUS). We analyzed VUS according to a multifactorial model. Our results support the suitability of the use of this method in other public institutions, in order to achieve a personalized medicine.

3438W

Exome sequencing identifies potential new candidate genes for unexplained colorectal adenomatous polyposis. *S. Aretz[1], D. Drichel[2], M. Kerick[3], J. Altmueller[4], A. Laner[5,6], S. Horpaopan[1], S. Vogt[7], T. Becker[2,8], P. Nuernberg[4], S. Perner[9], E. Holinski-Feder[5,6], M.M. Noethen[1,10], P. Hofmann[1,10,11], B. Timmermann[12], M. Schweiger[3], I. Spier[1].* 1) Institute of Human Genetics, University Hospital Bonn, Bonn, Germany; 2) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 3) Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) Cologne Center for Genomics (CCG), University of Cologne, Germany; 5) Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, Germany; 6) Medizinisch Genetisches Zentrum, München, Germany; 7) MVZ Dr. Eberhard & Partner, Dortmund, Germany; 8) Institute of Medical Biometry, Informatics, and Epidemiology, University of Bonn, Germany; 9) Department of Prostate Cancer Research, Institute of Pathology, University Hospital Bonn, Germany; 10) Department of Genomics, Life & Brain Center, University of Bonn, Germany; 11) Division of Medical Genetics, University Hospital Basel, Department of Biomedicine, University of Basel, Switzerland; 12) Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Background: In up to 50 % of families with colorectal adenomatous polyposis no germline mutation in the currently known genes APC causing familial adenomatous polyposis (FAP) or MUTYH causing MUTYH-associated polyposis (MAP) can be identified. Methods: To uncover new causative genes, the exomes of eleven unrelated APC and MUTYH mutation negative polyposis patients were sequenced (Illumina HiSeq platform). For data analysis and variant filtering the GATK software 2.1-8, ANNOVAR, and in-house tools were applied. The variants were filtered for truncating variants and frequent alterations were excluded. Results: In one patient, an APC nonsense mutation in mosaic state (10% of reads, coverage 249) was recognized, which could be confirmed by target sequencing with high coverage (read depths >1000). By Sanger sequencing only a very discreet peak was detectable. In the remaining 10 cases, 66 genes were affected by biallelic truncating variants (recessive model) in at least one patient and 63 genes were affected by truncating heterozygous variants (dominant model) in at least two patients. After detailed visual inspection of the variants in a read browser (Integrative Genomics Viewer) to exclude obvious sequencing artifacts, and data mining according to functions and pathways, six genes of high interest remained (two for the recessive, four for the dominant approach), some of which are involved in cell adhesion, proliferation, or recombination repair. By Sanger sequencing one of the recessive genes (ZSWIM7) and three of the dominant genes could be confirmed (DSC2, HEATR5A and PDE4DIP). Two mutations were artifacts which could not be validated (heterozygous intronic mutation, dominant model) or presented in heterozygous rather than homozygous state (recessive model; the coverage was only 7). Subsequently, one of the two different HEATR5A mutations was also identified in the unaffected mother (no symptoms at 79 years of age, but no colonoscopy). From the other patients no parental blood samples were available. Conclusions: Using exome sequencing we identified new potentially causative genes for adenomatous polyposis. The clinical relevance of the genes is presently clarified in a validation sample of 200 polyposis patients. The study was supported by the German Cancer Aid and BONFOR programme of the University of Bonn.

3439T

Prevalence of BRCA1 mutations in hereditary breast/ovarian cancer families and sporadic triple negative breast cancer patients from Algeria. *F. Cherbal[1], R. Bakour[1], W. Abdou[1], C. Mehemmai[1], K. Gassi[1], H. Gaceb[1], K. Boualga[2], N. Kanoun-Zitouni[3], W. Benbrahim[4].* 1) Unit of Genetics, LBCM, Faculty of Biological Sciences, USTHB, Algiers, Algeria; 2) Anti Cancer Center, Blida, Algeria; 3) Central Hospital of Algiers, Algiers, Algeria; 4) Medical Oncology Service, Anti Cancer Center, Batna, Algeria.

Bakground: Breast cancer is the leading cause of cancer death in Algerian women. The present work aimed to establish the frequency of the three most common *BRCA1* mutations in Algerian cohort of 96 hereditary breast/ovarian cancer families and sporadic TNBC patients. The identification of common mutations in *BRCA* genes may facilitate genetic testing and counseling. The three *BRCA1* mutations screened have previously been reported in hereditary breast/ovarian cancer families and sporadic breast cancer patients from Algeria. Methods: 66 hereditary breast/ovarian cancer families and 30 sporadic breast cancer patients (most of them had triple negative status) were screened for the recurrent *BRCA1* mutations c.181T>G and c.798_799delTT. The *BRCA1* mutation c.83_84delTG has been screened in 51 hereditary breast/ovarian cancer families. The approach used is based in PCR-direct sequencing. Results: The *BRCA1* mutation c.83_84delTG detected previously in two unrelated Algerian breast cancer patients has been identified here in a young breast cancer patient with a strong hereditary breast/ovarian cancer history with the frequency of 1.9% (1/51). Interestingly, the *BRCA1* mutation c.83_84delTG has been reported one time in BIC database in Caucasian family, could be specific of Algerian population. The c.181T>G/p.Cys61Gly mutation, has been detected here in a young bilateral breast cancer patient with a strong breast cancer family history with the frequency of 1.5% (1/66). To date, the Cys61Gly pathogenic variant is one of the most frequent founder mutation identified in Central European populations, has been previously reported for the first time in Maghrebian population in two Algerian and Moroccan families.In addition, haplotype analysis of Maghrebian and central European carriers of the *BRCA1* mutation c.181T>G will establish if the origin of this mutation in Maghrebian populations is linked to the Vandals, an East Germanic tribe, who invaded and established a kingdom in North Africa during the antiquity.The *BRCA1* mutation c.798_799delTT detected previously with a frequency of 4.93% in 4 families from Algeria,has not been detected in the present study.The sporadic TNBC patients were negative for the three *BRCA1* mutations. Conclusions: the screening for *BRCA1* and *BRCA2* germline mutations in large series of Algerian breast/ovarian cancer patients/families will allow to know about the frequency, the spectrum and the contribution of the prevalence of *BRCA* genes mutations.

3440F

Combination of founder mutation screening and genomic capture using BROCA yield high rate of loss of function mutations in early onset and familial breast and ovarian cancer in Greece. *F. Fostira[1], T. Walsh[2], S. Casadei[2], M.K. Lee[2], A. Vratimos[1], G. Fountzilas[3], I. Konstantopoulou[1], M.-C. King[2], D. Yannoukakos[1].* 1) NSCR Demokritos, Athens, Greece; 2) Dept of Medicine (Medical Genetics), University of Washington, Seattle, WA, USA; 3) Aristotle University of Thessaloniki, Thessaloniki, Greece.

The antiquity of the Greeks as a population defined by language and culture, and the complexity of Greek historical demography, present challenges to genetic testing, including genetic testing for predisposition to cancer. On the one hand, the Greek population harbors ancient founder mutations in many genes, including five damaging alleles of BRCA1, of which three are multiple kb genomic deletions. We screened for these founder alleles in patients with breast and ovarian cancer at a young age (<35 years) or with severe family history of breast, ovarian, and/or pancreatic cancer and discovered that 15% (109/754) carry one of the five mutations. On the other hand, many of the patients who were wildtype at these sites have equally young and familial disease. In order to determine the frequency and range of mutations beyond the Greek BRCA1 founder alleles, we carried out additional Sanger sequencing of BRCA1 and BRCA2 and undertook multiplex massively parallel sequencing using BROCA, which captures the entire loci of 30 breast and ovarian cancer genes, detecting all classes of mutations in each gene. Point mutations were validated by diagnostic PCR and Sanger sequencing. Genomic deletions were validated by MLPA. Of 376 patients with breast or ovarian cancer evaluated thus far, 51 patients (14%) carry loss-of-function mutations, other than the BRCA1 founder mutations, in one of the known breast and ovarian cancer genes. Frequencies are: 25 in BRCA1, 10 in BRCA2, 9 in CHEK2, 2 in PALB2, 2 in ATM, 2 in MSH2 (both patients had endometrioid ovarian cancer), and 1 in MLH1 (in a patient with both colon cancer and early onset bilateral breast cancer). Two missense mutations, CHEK2 p.I160M and CHEK2 p.G167R, which are known to abrogate DNA repair activity, and one genomic deletion, MSH2 del exons 1-7, have been observed in multiple Greek families and appear to be founder alleles. We conclude that among Greek patients with familial and early onset breast cancer, where the mutation rate in founder alleles is rather high, the mutational spectrum is nonetheless highly heterogeneous with respect to both loci and alleles. Half of mutation carriers are missed by screening only for founder alleles. A population that has a significant number of founder mutations in breast cancer susceptibility genes can still benefit by an approach that detects all classes of mutations in all known breast cancer genes.

3441W

A constitutional translocation disrupting NUP98 associated with bilateral renal angiomyolipomas. *A. Lehman, M. Steinraths, Z. Zong, L. Brown, K. Mungall, K.M. Nip, I. Birol.* University of British Columbia, Vancouver, Canada.

A 41 year old woman was incidentally found from amniocentesis to have a balanced reciprocal translocation: 46,XX,t(11;12)(p15.3;q14.2). This woman has significant bilateral renal angiomyolipomas, and no other features to suggest a diagnosis of tuberous sclerosis on extensive investigation. The woman's 2-year-old daughter is a carrier of the same translocation. Her parents are not available for investigation. Whole genome sequencing was performed on genomic DNA isolated from blood using the Illumina Hi-Seq 2000. Three lanes of data were generated and a de novo assembly was conducted with ABySS for chromosomes 11 and 12. The breakpoint was identified in 2 reads and 5 reads from each translocated chromosome. It disrupted NUP98, located at 11p15.4, in intron 1, prior to the translation start site in exon 2. The breakpoint on chromosome 12 was located a gene desert. Nuclear pore complex protein Nup98-Nup96 is encoded by NUP98, a gene recurrently involved in translocations, being fused to an activating partner, associated with blood cancer. Somatic loss of function mutations appear in NUP98 in a variety of cancers. Mice haplo-insufficient for NUP98 have decreased nuclear pore density. This is the first report of a germline translocation involving NUP98 to our knowledge.

3442T

An unusual case of Lynch syndrome - should chromosome analysis be offered routinely in the investigation of this condition? *A. Murray, E. Kirk, M. Prothero, D. Barrell, S. Rolleston, S. Palmer-Smith, P. Thompson.* All Wales Medical Genetics Service, Institute of Medical Genetics, University Hospital of Wales, Cardiff, United Kingdom.

Lynch syndrome is an inherited cancer predisposition syndrome, usually caused by germline mutations in a component of the DNA mismatch repair (MMR) system. We present an interesting Lynch syndrome family with an unusual mechanism of loss of MMR function. Our consultand, an asymptomatic, 47-year-old woman, was referred with a family history of colorectal, small bowel, and stomach cancer. We initiated tumour testing on an affected relative, whilst referring our consultand for regular colorectal screening. On her first colonoscopy, an adenocarcinoma was found in the tranverse colon. Local hospital immunohistochemical testing of this tumour, for MLH1 and MSH2 only, showed loss of MSH2 expression; molecular genetic testing in our laboratory showed microsatellite instability (MSI), with absence of the BRAF V600E mutation. More comprehensive immunohistochemical testing showed loss of expression of both MSH2 and MSH6, but sequence and dosage analysis of the MSH2 and MSH6 genes did not reveal any mutations. Karyotyping showed the presence of a small paracentric inversion of the short arm of chromosome 2, with breakpoints at p21.1 and p22.2. There is one report in the literature of a similar inversion in a patient with Lynch syndrome although breakpoints were not assigned. This result appears to confirm the clinical diagnosis of Lynch syndrome in our family. We are currently undertaking karyotyping of other affected family members, with a view to offering predictive testing to unaffected relatives, to clarify their risks and screening needs.

3443F

Germline *BAP1* mutations in uveal melanoma patients with a personal and/or family history of renal cell carcinoma. *R. Pilarski[1], C.M. Cebulla[2], O. Saqr[2], K. Rai[1], J.B. Massengill[2], F.H. Davidorf[2], M.H. Abdel-Rahman[2].* 1) Dept. of Internal Medicine and James Cancer Hospital,; 2) Dept. of Ophthalmology, Ohio State Univ, Columbus, OH.

Background: Germline mutations in the BAP1 gene have recently been shown by us and others to cause uveal melanoma (UM), mesothelioma, cutaneous melanoma, and possibly other cancers. Noting the occurrence of renal cell carcinoma (RCC) in several families, we investigated the frequency of germline mutations in BAP1 in patients with a personal history of UM and also a personal and/or family history of RCC. **Methodology:** Six patients were included in the study, including five females and one male. Average age was 52.6 years (range 41 to 72 years). Germline mutations in BAP1 and VHL were assessed in peripheral blood DNA by direct sequencing. In one patient with both UM and RCC somatic mutation in VHL was assessed in both the ocular and renal tumors. **Results:** We identified germline pathogenic mutations in BAP1 in 2/6 patients. Family histories of mesothelioma and other cancers including lung, colon, breast and pancreatic carcinomas were reported in these two families. No germline VHL mutation was identified; however, a somatic mutation in VHL was detected in the RCC tumor tissue but not in the UM tumor tissue from a patient with both primary tumors. **Conclusion:** BAP1 is a candidate gene in patients with a personal history of UM and a personal or family history of renal cell carcinoma.

3444W

Combined effect of germline mutations in *MUTYH* and DNA mismatch repair genes on risk of colorectal cancer. *A.K. Win[1], J.P. Young[2], D.D. Buchanan[2], S.P. Cleary[3,4], H. Kim[3,4], J.G. Dowty[1], R.J. MacInnis[1,5], T. Burnett[6], L. Le Marchand[6], P.A. Newcomb[7], R.W. Haile[8], N.M. Lindor[9], J.L. Hopper[1], S. Gallinger[3,4], M.A. Jenkins[1].* 1) Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The Univ Melbourne, Carlton, VIC, Australia; 2) Cancer and Population Studies Group, Queensland Institute of Medical Research, Clive Berghofer Cancer Research Centre, Herston, Queensland, Australia; 3) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Cancer Care Ontario, Toronto, Ontario, Canada; 5) Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, Victoria, Australia; 6) University of Hawaii Cancer Center, Honolulu, Hawaii, USA; 7) Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 8) Department of Medicine, Division of Oncology, Stanford University, California, USA; 9) Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona, USA.

Background Identifying modifiers of cancer risks is important for understanding carcinogenesis in carriers of *MUTYH* mutations and for genetic counselling, screening and risk-reduction strategies. As the MUTYH protein interacts with the DNA mismatch repair (MMR) system, we hypothesized that the combination of a monoallelic *MUTYH* mutation with an MMR gene mutation increases colorectal cancer (CRC) risk. It has been controversial that MMR gene mutations, especially *MSH6*, are more frequent in CRC-affected carriers of monoallelic *MUTYH* mutation than CRC-unaffected carriers of monoallelic *MUTYH* mutation. **Methods** Using the data from the Colon Cancer Family Registry Cohort and Cox proportional hazards regression weighted to correct for the method of ascertainment, we investigated (i) effect of MMR gene mutations on CRC risk for monoallelic *MUTYH* mutation carriers, and (ii) effect of monoallelic *MUTYH* mutations on CRC risk for MMR gene mutation carriers. **Results** In a cohort of monoallelic *MUTYH* mutation carriers, we observed that frequency of MMR gene mutations was higher in CRC-affected carriers (8/158, 5.1%) compared with CRC-unaffected carriers (2/225, 0.9%) (p=0.02). CRC risk for monoallelic *MUTYH* mutation carriers with an MMR gene mutation was higher than those without an MMR gene mutation (hazard ratio [HR] 13.2, 95% confidence interval [CI] 5.42-32.11, p<0.001). In a cohort of MMR gene mutation carriers, we observed that frequency of monoallelic *MUTYH* mutation was not different between CRC-affected (8/411, 2.0%) and unaffected carriers (2/68, 2.9%) (p=0.64). There was no evidence of difference in CRC risk for MMR mutation carriers with or without a monoallelic *MUTYH* mutation (HR 0.65, 95% CI 0.24-1.76, p=0.40). **Conclusion** In this study, there was no evidence of modifying effect by monoallelic *MUTYH* mutation on CRC risk for MMR gene mutation carriers whereas MMR gene mutation increased risk of CRC for monoallelic *MUTYH* mutation carriers.

3445T

Towards scoring all 35,397 possible missense variants of BRCA1 for activity. *L.M. Starita[1], J.O. Kitzman[1], J.G. Gullingsrud[1], J.D. Parvin[2], J. Shendure[1], S. Fields[1,3].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Department of Biomedical Informatics, The Ohio State University Comprehensive Cancer, Columbus, OH; 3) Howard Hughes Medical Institute.

BRCA1 is a breast and ovarian cancer-specific tumor suppressor gene that has been subject to much diagnostic sequencing. Multiple cancer-predisposing mutations have been identified along with >500 missense variants classified as Variants of Uncertain Significance or VUS. The BRCA1 protein consists of 1863 amino acids. The N-terminus contains a RING domain that is part of an active ubiquitin ligase, and the C-terminus contains tandem BRCT (BRCA1 C-Terminus) repeats that bind to phosphorylated peptides and activate transcription. BRCA1 is required for double-strand DNA break repair via homologous recombination; mutations found throughout the protein have deleterious effects on this function. We are devising assays to score all 35,397 possible missense mutations of BRCA1 for their effects on the protein's biochemical and cellular functions.

Our approach is to employ deep mutational scanning, a method that uses high throughput DNA sequencing to analyze protein variants that are expressed in a coupled genotype-phenotype system. To assess the effect of mutations on the function of the RING domain, we displayed residues 2-304 of BRCA1 in a phage-based assay that allows us to select variants active for *in vitro* auto-ubiquitination. This assay allowed us to score 2,413 (40%) of the possible 5,757 missense variants of this region of BRCA1. Among the variants scored, 57 have been identified in patients as VUS. In our assay, these variants ranged in ubiquitin ligase activity from totally inactive to fully functional, suggesting that some of the variants of BRCA1 that are classified as VUS are nonfunctional ubiquitin ligases.

We are currently developing three additional assays to score the effects of mutations in BRCA1. First, we are developing a phage display assay to select for BRCT variants that are able to bind to phosphorylated peptides. Second, we are using a yeast one-hybrid assay to analyze transcriptional activation mediated by the BRCT domain. Finally, we are working on a cell-based assay to score the effect of mutations in full-length BRCA1 on the cell's ability to carry out DNA repair via homologous recombination.

3446F

Clinical presentation of patients with mutations in the *APC* regions associated with AFAP. *P. Kaushik, K. Moyes, C. Arnell, M. Landon, R. Wenstrup.* Medical Services, Myriad Genetic Laboratories, Inc., Salt Lake City, UT.

Familial Adenomatous Polyposis (FAP) has been characterized by the development of 100s to 1000s of colorectal adenomas, while Attenuated Familial Adenomatous Polyposis (AFAP) has been characterized by the development of less than a 100 colorectal adenomas. Both FAP and AFAP are associated with mutations in *APC*. Studies looking into the phenotype/genotype correlation associated with the *APC* gene have highlighted specific areas in which mutations are more likely to lead to AFAP. Three significant areas are the 3' end, 5' end, and exon 9. This study aims to describe the polyp history of patients with mutations in these specific areas compared to mutations in the remainder of the *APC* gene. A retrospective analysis was performed on patients who underwent full *APC* gene analysis or targeted *APC* mutation analysis between January 1st 2006 and May 22nd 2013. Demographics and clinical characteristics were reviewed based on data reported on test request forms submitted to a commercial testing laboratory. Mutational testing results were also reviewed. The regions of the *APC* gene considered to be associated with AFAP were codons 1-157 (5' end), 312-438 (exon 9), and 1595-2855 (3' end). A total of 1534 individuals with mutations in the *APC* genes were included in this study. *APC* mutations in the three gene regions associated with AFAP were identified in 461 of 1534 patients (30.1%). Of these, 303 (65.7%) were reported to have developed less than 100 colorectal adenomas. A total of 1073 patients (69.9%) were found to carry an *APC* mutation in the remainder of the gene not associated with AFAP. Of these, 268 (25%) were reported to have developed less than 100 polyps. The majority of patients with an *APC* mutation in the three regions associated with AFAP were over the age of 40 at the time of testing (72.7%), while the majority of patients with mutations in the remainder of the *APC* gene were aged 20-40 at the time of testing (57.1%). This data shows that there appears to be a trend of less severe polyposis and later age of testing among patients with a mutation in the three main regions of *APC* associated with AFAP. However, a substantial number of patients were found to have over 100 colorectal adenomas regardless of the region the mutation was found in, suggesting that there is no definitive genotype/phenotype correlation between *APC* mutations and FAP/AFAP.

3447W

High sensitive detection of colorectal cancer mutations using third generation sequencing. *G. Russo[1], A. Patrignani[1], L. Poveda[1], F. Hoehn[2], R. Schlapbach[1], A. Garvin[2].* 1) Functional Genomics Center Zurich - ETH/UZH, Zurich, Switzerland; 2) Droplet Diagnostic SAS, France.

Colorectal cancer (CRC) is a major cause of cancer mortality. Unlike methylation, which occurs to some extent in normal tissue, mutations in the oncogenes and tumor suppressor genes that drive CRC are highly specific for tumor tissue. Mutations in patient samples can be found at many locations in these genes, therefore a scanning method for mutation detection is needed for high sensitivity. **Objectives:** To determine the sensitivity and specificity of single molecule, third generation sequencing for an assay that detects rare mutations in the genes that drive CRC. **Methods:** We performed three experiments in which mutant DNA from CRC derived cell lines and mixtures of mutant and wildtype DNA were PCR amplified and sequenced on the PacBio RS. High accuracy reads on short amplified fragments were obtained using circular consensus sequencing (CCS) to correct random errors. The assay consists of 15 amplicons covering regions of five genes mutated in CRC: KRAS, CTNNB1, BRAF, APC and P53. The amplicons were optimized at the same size to ensure uniform coverage and therefore uniform power of detecting mutations. The length of the total test sequence is about 5,000 bp. The results from the PacBio RS instrument were directly compared with Illumina MiSeq on a sample of amplified DNA containing three mutations (two substitutions and one, one base, deletion) present at the 1.5% level. **Results:** The quality of the CCS-called bases is orders of magnitudes higher than those of the Illumina sequences. Using DNA with mutations present at 1.5%, PacBio RS shows 100 % sensitivity and specificity while Illumina MiSeq, due to its systematic bias and higher background noise, missed one of the two substitutions and reported 10 false positives. **Conclusions:** We performed third generation, single molecule sequencing to detect low frequency CRC mutations in 15 short amplicons. A variety of sane and mutated cell lines and their mixtures at different concentrations was employed to simulate the real case of stool DNA. With no systematic bias and a much higher CCS base-calling quality, PacBio RS is superior than Illumina MiSeq. **Future work:** We generated mixtures with lower levels of mutant DNA (1 and 0.5 %) so to investigate the lowest possible frequency at which the detections occur; we have also sequenced two real samples, obtained from excised polyps, so to have our first test on a real tissue where the exact mutations are not known a priori. The data are currently being analysed.

3448T

SNV calls using RNA-Seq complements Exome-Seq variant calling in tumor cell lines. *S. Wong, P. Ebert, S. Bray, J. Calley, T. Barber.* Tailored Therapeutics, Eli Lilly and Company, Indianapolis, IN.

Whole-exome sequencing (exome-seq) and RNA sequencing (RNA-seq) have been widely utilized to uncover critical mutations and expression changes that drive tumorigenesis, drug resistance, or tumor evolution in cancer. Typical genomic approaches have relied on exome-seq to call single nucleotide variants (SNVs), while RNA-seq is primarily used for differential gene expression analysis or identifying tumor-specific translocation events. Though not routinely done, calling SNVs with RNA-seq data could provide additional benefit, including confirmation of exonic variants, however, RNA-seq variants are limited exclusively to expressed genes, which could also affect the detection of variants that result in reduced expression, such as some tumor suppressor mutations. To explore the concordance between exome-seq variants and RNA-seq called variants, we performed whole-exome sequencing and RNA sequencing from cancer cell lines and assessed the variant calls in genes expressed at differing levels. To address whether deleterious variants in tumor suppressor genes could be identified in RNA-seq data, we analyzed all nonsense and frame-shift mutations identified in exome sequencing for their presence in RNA-seq variant calls. In conclusion, RNA-seq variant calling can complement the variant calling from exome sequencing and can be a valued tool for calling somatic variants in expressed tumor genes.

3449F

Novel genetic and epigenetic alterations in the von Hippel-Lindau gene in a cohort of sporadic renal cell carcinoma patients from Pakistan. *A. Abid[1], S. Ajaz[1], S. Firasat[1], S. Shahid[1], A. Shehzad[2], G. Sultan[2], R. Mohsin[2], A. Hashmi[2], M. Mubarak[3], S.A.A. Naqvi[4], S.A.H. Rizvi[4], S.Q. Mehdi[1], S. Khaliq[1,5].* 1) Centre for Human Genetics, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 2) Department of Oncology, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 3) Department of Histopathology, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 4) Department of Urology, Sindh Institute of Urology and Transplantation, Karachi, Pakistan; 5) Department of Human Genetics, University of Health Sciences (UHS), Lahore, Pakistan.

Abstract: Background: Renal cell carcinoma (RCC) is the most frequent form of kidney cancer in adults and somatic mutations that inactivate the von Hippel-Lindau (VHL) gene are the most common cause of RCC. VHL gene alterations have been associated with clear cell subtype of RCC (ccRCCs), though there have also been reports on other types of RCCs, albeit at lower frequencies. Huge variations have been reported in the frequencies of mutations in the VHL gene in sporadic RCCs from different parts of the world ranging from 45% to 71%. However, there is still no data available on the involvement of the VHL locus in RCC cases from South Asian region. This report presents, for the first time, the genetic and epigenetic changes in a cohort of 300 renal cell carcinoma patients from Pakistan. Methods: To identify mutations in the VHL gene, direct DNA sequencing was carried out and epigenetic silencing was investigated by using methylation-specific polymerase chain reaction. Results: Our data showed molecular alterations in the VHL gene in 144 (48%) renal cell carcinoma patients. Somatic mutations were found in 87 (29%) patients and 35 novel mutations were identified. VHL promoter hyper-methylation analysis showed epigenetic changes in 83 (39.7%) out of 209 patients. A total of 105 ccRCC patients (52%) and 22 papillary RCC patients (41%) had a molecular abnormality in the VHL gene. A total of 26 RCC cases (12%) carried both types of molecular alterations i.e. somatic mutations as well as the VHL promoter hypermethylation. Patients who had no evidence of molecular alterations in the VHL gene were significantly younger than patients who carried some molecular change. However, there was no significant correlation of a genetic and/or epigenetic change in the VHL gene with other prognostic markers of renal cell carcinomas including tumour grade, size and stage. Molecular alterations in the VHL gene were not restricted to clear-cell renal cell carcinomas (ccRCC). Conclusions: The comparatively lower frequency of somatic mutations and higher frequency of epigenetic changes in the Pakistani cohort should be considered for developing specifically tailored preventive and therapeutic regimens against non-familial renal cell carcinomas.

3450W

Exome sequencing of cancer cell lines uncovers mutations not reported in commonly used COSMIC and CCLE databases. *S. Bray, K. Yu, Y. Webster, S. Wong, Y. Yue, I. Wulur, T. Barber.* Tailored Therapeutics - Genetics, Eli Lilly and Company, Indianapolis, IN.

Genomic characterization of cancer cell lines has become a critical tool for preclinical modeling of cancer. This is especially true for the interpretation of drug response data of targeted therapies, which relies on the accurate annotation of putative somatic mutations in the screened cell lines. Publicly available databases of likely somatic variants, such as the COSMIC Cell Line Project and the Cancer Cell Line Encyclopedia (CCLE), are valuable resources for annotating genetic variants in commonly utilized cell lines. These public databases, however, only provide sequence data for a relatively small panel of candidate cancer genes and the availability and affordability of next-generation sequencing have now made it more feasible to sequence the whole genomes or exomes of cell lines for additional genetic coverage. We used whole-exome sequencing to characterize the likely somatic variants across cancer cell lines and compared our variant calls to the public data for the mutually covered genes. We found very good sensitivity of our exome data to detect variants reported in COSMIC and CCLE. We also identified 'novel' mutations in candidate cancer genes that were not identified in either COSMIC or CCLE. We chose a subset of these novel variants to validate by Sanger sequencing and all were confirmed to be true variants. While we cannot exclude the role of cell-line divergence, our data suggest there may be many false negatives in the frequently used public databases that could ultimately impact interpretation of cell line behavior in functional studies.

3451T

Hierarchical models of mutational recurrence and allelic burden in prostate cancer. *R. Cowper[1,2], N. Sinnott-Armstrong[1,2], M. Lupien[3,4,5], M. Kellis[1,2].* 1) MIT Computer Science and Artificial Intelligence Laboratory, Cambridge MA USA; 2) The Broad Institute of MIT and Harvard, Cambridge MA USA; 3) Ontario Cancer Institute, Princess Margaret Cancer Centre-University Health Network, Toronto ON Canada; 4) Ontario Institute for Cancer Research, Toronto ON Canada; 5) Department of Medical Biophysics, University of Toronto, Toronto ON Canada.

Genome wide association studies (GWAS) have identified hundreds of prostate cancer risk-associated alleles. Similarly, whole genome sequencing (WGS) projects have identified thousands of mutations found in prostate tumors. Although the fields of post-GWAS and post-WGS aim to characterize the variants found thus far, many remain to be discovered. Indeed, all of these variants are still only able to account for a fraction of the heritability and a portion of tumor mechanisms. The main confounder in variant discovery is heterogeneity at the level of the population and of tumors. This heterogeneity precludes discovery through epistasis, whereby sets of variants manifest phenotypically as the same trait. Because the majority of GWAS and WGS studies have focused on individual alleles or mutations, heterogenous sets of causal variants have remained hidden. The notions of burden and recurrence have emerged as mechanisms for the integration of alleles and mutations in GWAS and WGS data respectively. However, both of these approaches are still founded on a model where contiguous genomic elements constitute the units being tested. But the set of genomic elements that can affect a gene is large and scattered across the entire genome. In order to overcome this limitation, we have studied non-contiguous sets of interacting variants that impinge on the same gene. We therefore use chromatin conformation and open chromatin data to identify the sets of elements that affect each active gene in prostate tissue. Through this process we build a hierarchical structure where smaller genomic elements are nested in larger functional aggregates. We then traverse this hierarchical structure looking for nodes that are recurrently mutated in tumors or that carry a high allelic burden in patients. We use WGS data from ten prostate adenocarcinomas and the entire set of prostate cancer risk-associated variants identified through GWAS. Furthermore, we characterize the mechanisms through which these variants disrupt gene expression through the analysis of affinity modulation and chromatin states.

3452F

Genomic Characterization 0f Invasive Cervical Cancer in Guatemala and Venezuela: Common Activation of the PIK3CA Pathway. *M. Dean[1], E. Gharzouzi[2], E. Alvirez-Freites[3], G. Villagran[2], G. Calabrese[3], J. Sawitzke[4], H. Lou[4], J. Boland[5], S. Chanock[5], M. Yeager[5].* 1) Lab Experimental Immunology, NCI-FCRDC, Frederick, MD; 2) Instituto Cancerologia, Guatemala City, Guatemala; 3) Hospital Central Universitario 'Dr. Antonio M Pineda', Barquisimeto, Lara State. Venezuela; 4) SAIC-Frederick; 5) DCEG, NCI, NIH.

Cervical cancer represents the most dramatic cancer health disparity of women in the world, with over 270,000 deaths annually, concentrated in poor, rural, and indigenous populations. Guatemala and Venezuela are illustrative of this disparity as cervical cancer is the predominant cause of cancer cases and deaths in women. The average age of the subjects is 50, and smoking and oral contraceptive use is low. The affected women have an average of 6 children and adeno- or adenosquamous pathologies represent 25% of the tumors. To determine the genomic changes in invasive tumors, we initiated a prospective collection of invasive cervical cancer tissue. To determine the most frequently mutated genes we sequenced the exomes of 23 tumors in the Ion Torrent Proton and followed up with targeted sequencing of genes by Ampliseq in an additional 80 tumors. The exome sequence revealed potential driver mutations in several known cancer genes such as PTEN, PIK3CA, TSC1, SETD2, HRAS, PTCH, ARID1A, ARID4A and RB1. Targeted resequencing in a larger set of tumors of commonly mutated genes in the COSMIC database (PIK3CA, TP53, STK11, PTEN, KRAS, HRAS, and CDKN2A) revealed 40% mutation in PIK3CA and 11% in PTEN (a negative regulator of the PI kinase pathway). TP53 and STK11 were significantly mutated (10 and 8% of tumors, respectively) but HRAS, KRAS and CDKN2A at a very low level (0-2%). The PIK3CA mutations were concentrated in the known activating sites in the helical domain (E542K and E545K). To determine HPV stain types we amplified the HPV L1 gene found detectable in HPV in 96% of samples with HPV16 in 53%, HPV18 in 12%, HPV45 in 10% and HPV 26, 31, 33, 35, 39, 52, 56, 58, 67, 68, 69, and 73 in 0.8-4.9% of samples. In conclusion, invasive cervical cancers display frequent mutation of the helical domain of PIK3CA and activation of the PI kinase pathway, suggesting that targeted inhibition of this pathway may provide an avenue for therapy in this largely untreatable, common cancer.

3453W

Whole Genome Sequencing of High-Grade Treatment-Naïve Prostate Tumors. *B.J. Decker[1,2], D.M. Karyadi[1], E. Karlins[1], L.S. Tillmans[3], S.N. Thibodeau[3], E.A. Ostrander[1].* 1) Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States; 2) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, United States.

Prostate cancer (PCa) will impact one in six American men and many low-risk men endure therapy, while other men die of PCa. Optimal treatment selection depends on distinguishing indolent tumors from those that will cause prostate cancer specific mortality (PCSM). To date, sequencing of prostate tumor DNA has highlighted the fact that prostate tumors develop differently than other cancers. Instead of recurrent protein-altering mutations, many prostate tumors harbor 'closed chain' structural rearrangements that may mediate tumor evolution. However, we do not yet understand which somatic mutations are involved in aggressive disease and PCSM. To address this crucial question, we are using whole genome sequencing to pinpoint somatic genetic features in tumor/normal paired samples from four PCSM and six non-PCSM patients with treatment-naïve, Gleason 8+; tumors. Tumor and normal DNA were sequenced to a mean coverage of 95.3X and 48.3X, respectively. Tumor DNA purity ranged from 49%-79%; and tumor ploidy ranged from 1.96 to 2.22. To define the genomic landscape of aggressive PCa, we will compare our findings with published studies and genotype high-priority candidates in at least 50 additional low- and high-grade tumors. We will also compare somatic mutations in PCSM versus non-PCSM cases and use network analysis to identify pathways that may be involved in progression and metastasis. Finally, we will characterize the clonal evolution of PCa tumors using deep digital sequencing to distinguish mutations that were present in the original tumor clone from subclonal mutations that arose after oncogenesis. In this way, we aim to identify potential oncogenic drivers and candidate mediators of progression and metastasis. Ultimately, we aim to 1) Uncover somatic mutations that are specific to aggressive tumors 2) Determine whether PCSM tumors have unique genetic features, and 3) Understand the role of clonal evolution in oncogenesis and progression. As more tumors are sequenced, this approach may yield valuable markers for diagnosis, prognosis and treatment selection.

3454T

Genomic landscape of 2 gastric cancer cases. *D. Esser[1], N. Holze[2], J. Haag[2], S. Krüger[2], V. Warneke[2], S. Schreiber[1,3], C. Röcken[2], P. Rosenstiel[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany; 2) Institute of Pathology, Christian-Albrechts-University, Kiel Germany; 3) Department of General Internal Medicine, University Hospital Schleswig-Holstein.

Gastric cancer is the fourth most common cancer type. With a 5-year survival rate of 26% gastric cancer causes 10% of all cancer deaths worldwide. In order to understand the genetic background we sequenced the whole exome and the whole genome of one microsatellite stable as well as one microsatellite unstable tumor and the matched healthy tissue on two different next generation sequencing (NGS) platforms (Illumina HiSeq and Solid 4). The samples were from two female patients at the same age with gastric cancer of the same tumor stage without histological evidence of EBV-infection. The average coverage in the whole exome sequencing was between 41- and 66-fold and in the whole genome sequencing between 14- and 78-fold. The study combines the analysis of single nucleotide variations (SNV) as well as structural variants like small and large insertions/deletions, translocations, inversions and tandem duplications and can be used as a blueprint for further gastric cancer genome analysis. Relevant mutations, pathways (KEGG) and functional terms (GO) were filtered out using population-based whole genome resources. The databases OMIM, HGMD and GWAS were accessed to detect known cancer associated variations. We further analyzed the mutation pattern and could show differences in the relative contributions of each of the six base substitution classes between tumor and healthy control samples. Additionally, we investigated the sequence context of each SNV by considering the bases immediately 5' and 3' to each base substitution and observed a tumor specific signature. Rainfall-plots were applied to identify chromosomal regions of high SNV density, however, in the preliminary results no obvious intragenic SNV clusters were detected. The results highlight a strategy of an exemplary tumor genome analysis that combines both exome and whole genome sequence information with two different NGS platforms and is one of the first gastric cancer whole genome studies. Using that comprehensive strategy we were able to identify a multitude of novel potentially damaging mutations, which were partially validated in an independent gastric cancer sample cohort of 452 independent samples.

3455F

The Application of Next Generation Sequencing (NGS) for Mutation Analysis of MyeloProliferative Neoplasms (MPNs) Patients in the State of Qatar. *Q. Fernandes[2], N. Al-Dewik[1,2], YK. Naidu[2], B. Cassinat[3], JJ. Kiladjian[4], H. El Ayoubi[2], N. Hammadi[2], G. Perkins[2], M. Yassin[2].* 1) Qatar Medical Genetics Center, Hamad Medical Corporation, Doha, Qatar; 2) Molecular Genetics of Cancer - Research Lab, NCCCR-HMC, Doha, Qatar; 3) Unité de Biologie Cellulaire, France; 4) Centre d'Investigations Cliniques, Hôpital Saint-Louis, Paris, France.

Since the last two decades, there has been major advances in studying the genetic basis of MPNs. Research in cancer genetics has enabled the identification of several genes in MPNs. However, there is still a significant number of MPN cases that are negative to most common genetic anomalies and many mutations are still unknown. Advanced genomic technologies like NGS play an important role in identifying and elucidating genetic defects causing polygenic disorders such as MPNs. 7 MPN patients and 7 healthy individuals from 3 consanguineous families were recruited into the study. DNA was extracted and sequenced using Ion Torrent PGM with 318 Chip. The hg19 human genome was used as the reference sequence. 190 amplicons targeting 739 cancer associated mutations in 604 loci from 46 key cancer genes were amplified. The NGS data analysis was done using Ion Reporter, Ingenuity Variant Analysis and Catalog Of Somatic Mutations in Cancer (COSMIC) Out of four PV patients, three were positive for the JAK2 V617F mutation. One patient had mutations in PDGFRA 838-840delGLA, APC Q1285fs, NPM1 Splice Site Loss and PTEN S10N. The second patient had mutations in KIT M541L, KDR Q472H, TP53 P72R and the third patient had PI3KCA I391M, KIT M541L, KDR Q472H, P53 P72R mutations. The fourth patient negative to the JAK2 V617F mutation had PDGFR A838-840delGLA and SMARCB1 T72K mutations. Out of three ET patients, one patient was positive for JAK2 V617F, SMARCB1 T72K, IDH1 K115Q and PDGFRA 838-840delGLA. two patients were negative for the JAK2 V617F and MPL mutations, out of which one patient had mutations in PDGFRA 838-840delGLA, APC Q1285fs and SMAD4 198-200delPAL. The other patient had mutations in CDKN2A S73R, APC Q1285fs and PDGFRA 838-840delGLA. It was identified that PDGFRA 838-840delGLA and APC Q1285fs mutations were also found in healthy individuals. This study was able to identify a list of deleterious somatic mutations such as missense mutations, frame shift mutations, in-frame deletions, splice loss sites and Single Nucleotide Variation (SNV) in MPNs patients and healthy individuals. Our preliminary results suggest that the MPNs patients in Qatar have complex mutations. Evidences show that there exists a possibility of the disease arising out of the accumulation of genetic alteration and not as the consequence of a single genetic event. This could possibly be due to high rate of consanguineous marriages in Qatar ie. the 'Founder Effect'.

3456W

Massive parallel sequencing of BRCA1 and BRCA2: detection of deleterious mutations and variant of unknown significance in breast cancer patients from Colombia - South America. *L. Galeano Petro[1,2], G. Guevara Pardo[1], H. Groot De Restrepo[2], D. Restrepo Montoya[1].* 1) Instituto Colombiano de Genética y Oncología Molecular, Bogotá - Colombia; 2) Universidad de los Andes, Bogotá - Colombia.

Introduction: This study describes how the high -throughput or massive parallel sequencing technology can be implemented in a diagnostic setting for the breast cancer susceptibility genes, BRCA1 and BRCA2. At the present, genetic testing is offered in many centers in North America, Europe, Australia and Israel, but is not generally available in developing countries as Colombia-South America. Genetic testing is gaining acceptance worldwide because of the increasing numbers of preventive options available to women with a mutation, and because of the development of novel, individualized, cancer therapies. Purpose: We performed deleterious mutations and variant of unknown significance analysis of BRCA1 and BRCA2 on patients with breast cancer from Colombia using massively parallel sequencing. Methods: Of 120 patients, initially sequenced by the method of Sanger, were subselected 12 patients for the massive parallel sequencing of BRCA1 and BRCA2. The throughput was maximized by increasing uniformity in coverage, obtained by target enrichment of BRCA1 and BRCA2 in 21 amplicons ranging from 1.2-5.9kb and 6 reactions by long PCR multiplex approach. The preparation and sequence of fragment library was performed using NanoDrop quantification of PCR products, shearing of PCR products using the Covaris S220 System and verification using the Agilent Bioanalyzer, end-polish and size-select the DNA by Agencourt AMPure XP reagent, verification on the Agilent Bioanalyzer, add a dA-tail to the size-selected DNA, ligation of adaptors to the DNA, quantification the ligated DNA, check the size distribution of the library, pool equal molar barcoded libraries and quantitate the library, emulsify and amplify the beads, enrich the template beads, modify the 3' ends of the DNA, deposit the beads and sequence the library in 5500xl SOLiD Sequencer, post-run data export and analysis by LifeScope 2.5.1 Software and finally, confirmation of all sequence variants by Sanger sequencing. Results: deleterious mutations and variants of unknown significance were identified and confirmed by Sanger sequencing with 100% concordance. Conclusion: Our workflow illustrates the potential of massive parallel sequencing of large genes in a diagnostic setting which is of great importance to meet the increasing expectations of genetic testing. A wider spectrum of at risk women in Colombia will be able to benefit from therapeutic and prophylactic interventions.

3457T

Understanding the Significance of Individual Tumor Genetic Heterogeneity by Developing Next Generation Genetic Databases as Advanced Analysis Tools. *B. Gottlieb[1, 2, 3], L.K. Beitel[1, 2, 3, 4], M. Trifiro[1, 2, 3. 4].* 1) Dept Cell Gen, Lady Davis Inst Med Res, Montreal, QC, Canada; 2) Segal Cancer Centre, Jewish General Hospital, Montreal, QC, Canada; 3) Dept Human Genetics, McGill University, Montreal, QC, Canada; 4) Dept Medicine, McGill University, Montreal, QC, Canada.

The discovery of intra-tumor genetic heterogeneity in cancer tissues has had a significant effect on cancer ontogeny by making it even more difficult to assess which gene alterations are drivers and which are passengers within individual cancer patients. To resolve this issue, we have developed a new approach to sequencing that has allowed us to create a distribution profile of mutant variants within tumors. This has allowed us to measure the frequency of distribution of an androgen receptor (AR) repeat length variant associated with breast cancer. Initial data indicate that the relationship of AR mutations and breast cancer is complex and cannot be explained solely by the presence of specific mutations within breast cancer tissues. Understanding intra-tumor genetic heterogeneity and its possible relationship to cancer ontogeny, is further complicated by the presence of genetic heterogeneity in normal breast tissues as well. Thus, it is becoming clear that most individual cancers will have their own unique genetic profiles. In addition to intra-tissue genetic heterogeneity, the discovery that post-genomic events such as epigenetics can play an important role in determining cancer phenotype, have combined to challenge a number of classic genetic paradigms with regards to the genotype to phenotype relationship. It is the exact nature of this relationship that will need to be resolved to eventually understand carcinogenesis, and we are proposing the creation of next generation genetic databases (NGDBs) as a solution. These NGDBs will incorporate the new genotype to phenotype paradigms, as well as include powerful systems biology analysis tools to actively process and evaluate the vast amounts of both genomic and post-genomic information that will undoubtedly be revealed. The Human Variome Project (HVP) will be an essential element to ensure the effectiveness of NGDBs with its goals of setting up a worldwide network of nation-based genome collection nodes and creating standards for next generation sequencing. Thus, NGDBs, and the HVP can become the instruments for achieving real breakthroughs in understanding the genetic basis of individual cancers as well as their prevention and treatment.

3458F

Utilizing Publically Available NGS Tumor Data to Identify Novel Oncology Targeted Therapies. *D. Greenawalt[1], J. Bradford[2], M. Wappett[2], A. Dulak[1], K. Vasudevan[3], K. Jacques[3], S. Guichard[3], J. Dry[1].* 1) Oncology Bioinformatics, AstraZeneca, Waltham, MA; 2) Oncology Bioinformatics, AstraZeneca, Alderley Park, UK; 3) Cancer Biosciences, AstraZeneca, Waltham MA.

Public tumor consortiums, including the The Cancer Genome Atlas (TCGA1) and International Cancer Genome Consortium (ICGC2) are generating and making publically available a large amount of molecular data on a diverse set of tumor types. Through tumor type specific publications the TCGA is releasing intensive, integrative analysis of specific tumor types and many groups are clamoring to extract novel findings from the data through deep mining and novel integrative analysis. These analyses will lead to an abundance of novel targeted oncology therapies emerging over the next few years. We will present our integrative analysis methods for identifying novel genetic targets from publically available molecular datasets, integrating mutations, gene expression, copy number and fusions from the TCGA data; how these targets are filtered and screened to identify novel drug targets with the highest probability of success, and the pitfalls identified mining primary, treatment naïve datasets. These methods can then be translated to data released from the ICGC and additional datasets to learn how the disease segmentation for targets identified translate across tumors from diverse populations. 1.http://cancergenome.nih.gov/ 2.http://icgc.org/.

3459W

Mutation Status of P53 in Head and Neck Squamous Cell Carcinoma. *E. GUNDUZ[1], G. NAS[1], M. ACAR[1], S. DEDE[1], K. ERDOAN[1], C. MOROSKI ERKUL[1], M. GUNDUZ[1,2].* 1) MEDICAL GENETICS, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey; 2) DEPARTMENT OF OTOLARYNGOLOGY, HEAD AND NECK SURGERY, TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey.

Head and Neck Squamous Cell Carcinoma (HNSCC) occurs in the Oral Cavity, oropharynx, larynx or hypopharynx and is the sixth most frequent cancer worldwide. Although molecular mechanisms of HNSCC have not yet been revealed completely, p53 alterations are the most frequent event that occurs during carcinogenesis. TP53 is a tumor suppressor protein encoded by the TP53 gene in human and responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence or DNA repair. Mutations in the p53 gene result in loss of function or decreased levels of gene expression and is associated with a variety of human cancers. Also, p53 inactivation is clinically important because of a strong correlation to an acquired or innate resistance to chemotherapy. Hence, p53 mutation status is one of the most important biomarkers for carcinogenesis and therapy. Our laboratory aimed to identify the mutation status of p53 in HNSCC by full exon sequencing from exon 4 to exon 9 where the most mutations of p53 are located. P53 mutation status was investigated in 31 Head and Neck Squamous cell carcinoma patients and 12 different HNSCC cell lines and these mutations were compared with those in the literature. Our results not only verified previously identified mutations, but also uncovered new mutations that may have functional and clinical importance. Consequently, newly discovered p53 mutations in HNSCC suggest new targets may be important in carcinogenesis, drug resistance and gene therapy.

3460T

Next-generation sequencing of paired drug-sensitive and resistant cell lines identifies spectrum of DNA changes associated with drug resistance. *P. Jia[1], H. Jin[2], C.B. Meador[2], J. Xia[1], K. Ohashi[2], L. Liu[3,4], V. Pirazzoli[5,6], K.B. Dahlman[7], K. Politi[5,6,8], F. Michor[3,4], Z. Zhao[1,7], W. Pao[2,7].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Departments of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA; 4) Department of Biostatistics, Harvard School of Public Health, Boston, MA; 5) Department of Pathology, Yale University School of Medicine, New Haven, CT; 6) Yale Cancer Center, Yale University School of Medicine, New Haven, CT; 7) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN; 8) Department of Medicine, Yale University School of Medicine, New Haven, CT.

Somatic mutations in genes encoding kinases are associated with increased sensitivity of some solid tumors to kinase inhibitors, but patients with metastatic cancer eventually develop disease progression. So far, known resistance mechanisms have been identified such as the second-site mutation, *EGFR* T790M, amplification of the gene encoding an alternative kinase, *MET*, and epithelial-mesenchymal transition (EMT). However, the full spectrum of DNA changes associated with EGFR TKI acquired resistance remains unknown. Here, we used next-generation sequencing and bioinformatics analysis to characterize mutational profiles associated with 4 populations of *EGFR* mutant drug-sensitive cell lines and 5 matched drug-resistant cell lines. We developed a data analysis pipeline to detect SNVs, small insertions and deletions (indels), and CNV changes. By comparing resistant cells with their parental counterparts, we identified 16-89 coding SNVs/indels that were acquired and 1-27 that were lost; few SNVs/indels were shared across resistant lines. Comparison of two related parental lines revealed no unique coding SNVs/indels, suggesting that the changes in the resistant lines were due to drug selection. We further analyzed the genomic features of the SNVs identified through whole genome sequencing, and found that SNVs tend to occur more frequently in 'constant late' replication timing zones as compared to 'constant early' replication timing zones (chi-squared p-value $< 10^{-5}$). An enrichment of SNV frequencies was observed in genomic regions harboring lamina-associated domains compared to the remainder of the nucleus (chi-squared p-value $< 10^{-5}$). Surprisingly, we observed a higher burden of CNV changes across all resistant lines, and the one line that had an EMT phenotype displayed significantly higher levels of CNV changes than the other lines with acquired resistance. These results demonstrate a framework for studying the evolution of drug-related genetic variants over time and provide the first genome-wide spectrum of mutations associated with the development of cellular drug resistance in an oncogene-addicted cancer. Collectively, the data suggest that CNV changes may play a larger role than previously appreciated in the acquisition of drug resistance and highlight that resistance may be heterogeneous in the context of different tumor cell backgrounds.

3461F

Somatic Mutation Profiles of Non-Syndromic Early-Onset Colorectal Cancer. *N. Jinawath[1], T. Pongrujikorn[1], S. Ngernna[2], C. Songpattanasilp[1], J-S. Su[3], A. Jinawath[4], A. Tunteeratam[5], B. Suktitipat[6].* 1) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Department of Biology, Faculty of Science, Mahidol University, Bangkok, Thailand; 3) Illumina, Taiwan; 4) Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 5) Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 6) Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

Colorectal cancer (CRC) is the third most common cancer in the world and the fourth most common cause of cancer death worldwide. As a result of dietary westernization, the incidence rates of CRC in Asian countries such as Thailand, where rates were previously low, are rising. Less than 10% of CRC are diagnosed in patients younger than 50 years old. Among these, most cases have no known family history of cancer or genetic predisposition, and are considered to be sporadic despite the fact that high genetic burden is likely involved. Although the sequential genetic alterations underlying CRC have been extensively studied, the somatic changes of non-syndromic early-onset CRC have been largely unexplored. Using a set of Formalin-Fixed-Paraffin-Embedded (FFPE) samples from 119 Thai CRC patients, we have analyzed the differential somatic mutation profiles between the 63 early-onset non-syndromic CRCs (< 50 years) and 56 average-age-of-onset sporadic CRCs (> 50 years). Illumina's TruSeq Amplicon Cancer Panel which comprises of 212 amplicons containing somatic mutation hotspots from 48 known cancer-related genes were utilized in our study, and pooled amplicon sequencing was carried out on a MiSeq platform. Variants were called by Somatic Variant Caller, and filtered according to default settings with sequence coverage of at least 100X. Furthermore, immunohistochemical analyses of MLH1, MSH2, MSH6, and PMS2 expression to rule out Hereditary Nonpolyposis Colorectal Cancer (HNPCC), the most common inherited CRC syndrome, and to indirectly assess microsatellite instability (MSI) status were performed. Among the 119 CRCs, the commonly mutated genes *KRAS*, *BRAF*, and *TP53* were mutated in 44%, 40%, and 76%, respectively. Their mutation frequencies in Thai CRCs were in line with those previously reported. Single variant analyses did not identify any variant associated with early-onset CRC. However, using gene-based burden test for rare variants (MAF < 0.02), we identified 10 genes associated with early-onset CRC, namely *BRAF*, *PDGFRA*, *ATM*, *ABL1*, *EGFR*, *RB1*, *APC*, *KDR*, *GNAQ*, and *IDH1* ($p < 0.0002$ after Bonferroni correction). Our results indicate that the development of CRC in non-syndromic early-onset patients may have different underlying genetic alterations than those of average-age-of-onset patients.

3462W

Comprehensive molecular analysis of basal breast cancer treated with neoadjuvant paclitaxel.
KK. Kalari[1], X. Tang[1], KJ. Thompson[2], PT. Vedell[1], H. Sicotte[1], SN. Hart[1], A. Moyer[3], D. Visscher[3], AA. Nair[1], JP. Sinnwell[1], DW. Mahoney[1], P. Barman[1], TJ. Dockter[1], KN. Jones[4], AL. Conners[4], VJ. Suman[1], JE. Eckel-Passow[1], CG. Schultz[8], JP. Kocher[1], ED. Wieben[9], RM. Weinshilboum[7], L. Wang[7], JC. Boughey[5], MP. Goetz[6,7]. 1) Department of Health Science Research, Mayo Clinic, Rochester, MN,55905 USA; 2) Cancer Research, Mayo Clinic, Florida 32224 USA; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester,MN 55905 USA; 4) Department of Radiology, Mayo Clinic, Rochester, MN 55905 USA; 5) 5Department of Surgery, Mayo Clinic, Rochester, MN 55905 USA; 6) Medical Oncology, Mayo Clinic, Rochester, MN 55905 USA; 7) Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905 USA; 8) Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905 USA; 9) Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN 55905 USA.

Basal breast cancers (BBC) are typically negative for ER, PR and HER2. It is known that the underlying genomic alterations and drug response phenotypes in BBC are different compared to other breast cancer subtypes. There are no validated markers associated with paclitaxel response. We performed next generation sequencing to investigate biomarkers of paclitaxel drug response in women who enrolled in the prospective Breast Cancer Genome Guided Therapy (BEAUTY) neo-adjuvant clinical study. In this study, women undergo tumor biopsies for NGS at baseline, after 12 weeks of paclitaxel and following anthracyline-based chemotherapy. Subsets of women with BBC in which both sequencing data along with paclitaxel drug response as measured by MRI were considered in this study. From a cohort of 54 patients, we identified the extremes of phenotype response to paclitaxel in 12 patients with BBC, 5 with evidence for a complete response by MRI and 7 with stable disease (no tumor shrinkage). As expected, BBC showed a high frequency of somatic mutations. 11/12 patients have at least one or more somatic mutations in TP53, PTEN, RB1 and PIK3CA. Several TP53, BRCA1 and MYO3A germline mutations were also identified. In addition, we identified tumor amplification of MYC, PIK3CA, and MYO3A in 7/12 samples. We obtained basal gene expression data for these 12 patients and performed differential gene expression (DGE) analysis based upon these MRI response phenotypes. We identified 805 genes differentially expressed between two groups with a p-value < 0.05. Our analyses of 4 different pathway tools (cytoscape, IPA, SPIA, metacore) demonstrated that tight junction, Jak-STAT, ErbB, and Wnt signaling pathways are activated and maturity onset diabetes pathway is inhibited in stable group compared to the complete response patients. Significant pathways obtained from DGE analysis were integrated to build a basal transcriptomic landscape (T1) for paclitaxel response. Key nodes in T1 will be further investigated with a variety of genomic features such as splicing, expressed single nucleotide variants, fusion transcripts, and copy number alterations. For this same cohort of BBC, we have also obtained RNA-Seq data after paclitaxel treatment. Hence, we plan to compare and contrast the transcriptomic landscape (T1) with the genomic features obtained from second time point. Integration of such comprehensive OMICS data will help us identify novel molecular targets for BBC treatment.

3463T

Next Generation sequencing of osteosarcoma identifies the PI3K/mTOR pathway as a unifying vulnerability to be exploited for targeted therapy.
A. Kiezun[1], K. Janeway[2], P. Tonzi[2], J. Mora[3], S. Aguiar[5], G. Mercado[4], J. Melendez[4], L. Garraway[2], C. Rodriguez-Galindo[2], S. Orkin[2], T. Golub[1], G. Getz[1], JA. Yunes[5]. 1) Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Dana Farber Cancer Institute, Boston, MA, USA; 3) Sant Joan de Déu Hospital, Barcelona, Spain; 4) Instituto Nacional de Medicina Genómica, Mexico City, Mexico; 5) Centro Infantil Boldrini, Sao Paulo, Brazil.

Background: Osteosarcoma is an aggressive bone cancer often presenting in childhood or adolescence. Due to the complexity of the osteosarcoma genome, identification of tractable therapeutic targets has been challenging and there have been no substantial therapeutic advances in osteosarcoma in over 2 decades. Methods: We examined 59 matched tumor/normal sample pairs from osteosarcoma patients by whole-genome sequencing (WGS; 12 cases), whole-exome sequencing (WES; 59 cases), and RNA-Sequencing (RNASeq; 35 cases). Statistical approaches were used to identify genes and pathways likely to be involved in oncogenesis. We applied a heuristic algorithm (PHIAL) to identify the spectrum of clinically actionable alterations in osteosarcoma. In addition, comparative oncology was employed by whole exome sequencing of 9 TN pairs from a murine model of osteosarcoma. Results: The average non-silent somatic nucleotide variant frequency was to 1/Mb. The MutSigCV algorithm, identified two genes mutated at significant frequency across the 59 samples (at q-value threshold of 0.1): *TP53* and *PRB2*. Additionally, *PTEN* had qvalue of 0.13. Analysis performed using the MutSig2.0 algorithm identified 37 pathways as being significantly ($q<0.1$) affected by mutations across the 59 samples. The 26 most significant pathways contained *TP53*. Of the remaining 11 significant pathways, 5 implicate the PI3K/AKT/mTOR signaling pathway. *PTEN* and *PIK3R1* were affected by somatic mutations in both mouse and human tumors, including a nonsense mutation in Pten in one mouse. Osteosarcoma cell lines treated with PI3K/mTOR inhibitors BEZ235 and PIK75 had decreased pAKT and pS6. IC50s of these inhibitors in osteosarcoma cell lines were 8 to 82nM. *TP53* germline mutations were present in 7 (12%) patients. Somatic rearrangements were extremely common with an average rate of 10/Mb and there was chromoplexy in 11/12 cases with WGS data. **Conclusions:** This study identifies the PI3K/mTOR pathway as a unifying vulnerability to be exploited for targeted therapy in osteosarcoma with supporting evidence from patient-derived samples and comparative oncology, genomic and functional approaches. These results support further study of dual PI3K/mTOR inhibition in osteosarcoma.

3464F

RNA-seq revealed alternative splicing and fusion transcripts in non-small cell lung cancer.
W. Kim, S. Park, M. Cheon. Kangwon National University, Chuncheon, South Korea.

Lung cancer is the most common cause of cancer related death. Alterations of sequence or structure of genes and their expression have important role in the pathogenesis of lung cancer. Fusion genes and alternative splicing of cancer-related genes can have the potential oncogenic activity. In the current study, we performed RNA-seq to investigate potential fusion genes and alternative splicing in non-small cell lung cancer. RNA was isolated from lung tissues obtained from 86 subjects with lung cancer. The RNA samples from lung cancer and normal tissues were processed with RNA-seq using the HiSeq 2000 system. Fusion genes were evaluated using deFuse and Chimerascan. Candidate fusion transcripts were validated by Sanger sequencing. Alternative splicing was analyzed using multivariate analysis of transcript sequencing (MATS) and validated using RT-PCR. From RNA-seq data of 86 paired samples, 3 known fusion genes were identified. Nine cnadidate fusion transcripts were selected using DeFuse or Chimerascan. Among them, 5 fusion genes including *HNRNPA2B1-SKAP2* were validated by Sanger sequencing. MATS revealed twenty-nine tumor specific skipped exon events and six mutually exclusive exon events. Among skipped exon variants, 8 were known splice variants and 21 were novel variants. *ITGB4* and *PYCR1* were top genes that showed significant tumor specific splice variant. In conclusion, novel potential fusion transcripts and splice variants were identified. Their functional significance in pathogenesis of lung cancer should be evaluated.

3465W

Exome sequencing of mammospheres and primary tumors indicates dynamic transitions between stem-like and differentiated states in breast cancer. *D. Klevebring[1,2], G. Rosin[3,4], R. Ma[4], J. Lindberg[1,2], K. Czene[1], J. Kere[2,3,5], I. Fredriksson[6,7], J. Bergh[4,8], J. Hartman[4,9].* 1) Department of Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Stockholm, Sweden; 2) Science For Life Laboratory, Stockholm, Sweden; 3) Department of Biosciences and Nutrition, Center for Biotechnology, Karolinska Institutet, Stockholm, Sweden; 4) Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden; 5) Research Programs Unit, University of Helsinki, and Folkhälsan Institutet of Genetics, Helsinki, Finland; 6) Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 7) Department of Breast and Endocrine Surgery, Karolinska University Hospital, Stockholm, Sweden; 8) Radiumhemmet - Karolinska Oncology, Karolinska University Hospital, Stockholm Sweden; 9) Department of Clinical Pathology, Karolinska University Hospital, Stockholm, Sweden.

Breast cancer stem-like cells (CSCs) are a slow-proliferating although highly tumorigenic subpopulation of cells defined by expression of embryonic stem cell markers and tumor- initiating capacity. Although a minority in tumors, CSCs have been shown to resist chemotherapy, and are consequently considered as a cause of therapy resistance and late tumor relapse. In order to understand the origin of CSCs, we performed exome sequencing of CSCs from ten breast cancer patients, along with paired primary tumor samples. Our analysis revealed that the majority of somatic mutations are shared between CSCs and bulk primary tumor. Furthermore, we show that the allele frequencies of these shared mutations do not differ significantly between the two components. The results indicate that CSCs are a population of cells that can dynamically switch from differentiated tumor cells, and vice versa. This finding increases our understanding of the central role of CSCs in tumor heterogeneity and the importance to identify drugs to counter dedifferentiation rather than targeting CSCs.

3466T

Genomic alterations during disease evolution in CML patients displaying an isochromosome 17q. *N. Larsson[1], H. Liljebjorn[1], M. Rissler[1], C. Högberg[1], J. Richter[2], t. Fioretos[1].* 1) Clinical Genetics, Lund University, Lund, Sweden; 2) Department of Hematology, Lund University Hospital, Lund, Sweden.

Chronic myeloid leukemia (CML) is a clonal disease of hematopoietic stem cell that is characterized by a t(9;22) (q34;q11), also known as the Philadelphia chromosome. The translocation gives rise to the BCR/ABL1 fusion gene that encodes a constitutionally active tyrosine kinase. However, further genetic changes are needed for CML to progress from chronic phase to the more advanced disease stages; accelerated phase and/or blast crises. Isochromosome for the long arm of chromosome 17, i(17q), is one of the most common cytogenetic abnormalities detected at the disease progression of CML, but the changes at the molecular level still remain elusive. We have analyzed the clinical features and mutation patterns in three CML patients that acquired an i(17)(q10) during disease progression. Samples representing different stages of CML were subjected to SNP-array analysis and exome sequencing. Mutations in genes involved in DNA repair, cell cycle control and epigenetic modification such as *TP53, ASXL1* and *EZH2* were detected, but no recurrent mutations were identified. In one case, a *TP53* mutation was detected before the occurrence of i(17q), suggesting that the loss of 17p-material partly served to unmask the effect of the *TP53* mutation. However, no *TP53* or other mutations on chromosome 17 were detected in the other two patients, suggesting that i(17q)-formation leads to disease progression by causing dosage imbalances of several genes encoded by chromosome 17 rather than via specific gene mutations.

3467F

Whole exome sequencing identifies different sets of gene mutations in various subtypes of early-onset breast cancers. *C.Y. Lee[1], W.H. Kuo[2], C.H. Lin[3], Y.S. Lu[3], H.I. Yang[1,4], K.Y. Lo[1], C.K. Liu[5], C.H. Chang[6], C.N. Hsiung[5], N. Leng[7], K. Nobuta[7], C.D. Haudenschild[7], C.H. Chen[5], J.Y. Wu[5], K.P. Chiu[1], A.L. Cheng[3], K.J. Chang[2,8], C.Y. Shen[5,9], C.J. Chen[1].* 1) Genomics Research Center, Academia Sinica, Taipei, Taiwan; 2) Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Oncology, National Taiwan University Hospital, Taipei, Taiwan; 4) Molecular and Genomic Epidemiology Research Center, China Medical University Hospital, Taichung, Taiwan; 5) Institute of BioMedical Sciences, Academia Sinica, Taipei, Taiwan; 6) Department of Medical Research, National Taiwan University Hospital, Taipei, Taiwan; 7) Illumina Inc., Hayward, CA, USA; 8) Department of Surgery, Cheng Ching General Hospital, Taichung, Taiwan; 9) School of Public Health, China Medical University Hospital, Taichung, Taiwan.

Breast cancers in young women manifest differently from those diagnosed at later ages. They are typically aggressive and most patients are candidates for adjuvant therapy. Knowledge of genomic changes is therefore valuable for optimizing treatment strategy. This study identified single nucleotide variations (SNVs) using whole-exome sequencing to provide a comprehensive view of genetic alterations in these patients. Somatic and germline SNVs were investigated in 81 patients diagnosed at 40-years-old or younger. Based on the expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67, tumors were classified into five subtypes: luminal A (ER- and/or PR-positive, HER2-negative, Ki67 <14%, 34 cases), luminal B-HER2-negative (ER- and/or PR-positive, HER2-negative, Ki67 >=14%, 13 cases), luminal B-HER2-positive (ER- and/or PR-positive, HER2-positive, any Ki67, 14 cases), HER2-positive-non-luminal (ER- and PR-negative, HER2-positive, 9 cases), and triple-negative (ER- and PR-negative, HER2-negative, 11 cases). To define which SNVs were functional or related to functional changes, we only included non-synonymous/synonymous changes in regions evolutionally conserved across 44 species. To focus on mutation and not polymorphism, any genetic changes reported in the 1000 Genomes Project and dbSNP were excluded. Nine hundred and thirty-six (877 somatic and 59 germline) SNVs were detected. In the highest-mutation-frequency genes, only PIK3CA and TP53 have been previously reported. All triple-negative and five non-triple-negative patients had a high-mutation-frequency phenotype involving mutations in genes maintaining genomic stability. In the remaining non-triple-negative patients, 32.4% harbored mutation in the PI3K pathway-related genes, but almost all of these (tyrosine kinase receptor genes, G protein-coupled receptor genes and PIK3CA) play an upstream role in the pathway. Mutated genes involved in diverse and unrelated tumorigenic mechanisms were identified in cancers of the same subtype, and tumors of different subtypes were found to harbor genes in the same functional pathways. These findings not only reveal the complexity of the etiology, but also provide valuable clues for developing therapeutic targets and predictive biomarker for early-onset breast cancer.

3468W

Exome sequencing of cell-free plasma DNA in prostate cancer patients. *J. Lindberg[1], M. Neiman[1], D. Klevebring[1], P. Wiklund[2], L. Egevad[3], H. Grönberg[1].* 1) MEB, Karolinska Institutet, Stockholm, Sweden; 2) Department of Urology, Karolinska Hospital, Stockholm, Sweden; 3) Department of Pathology and Cytology, Karolinska Hospital, Stockholm, Sweden.

Circulating nucleic acids have previously been detected in serum and plasma of both healthy and diseased individuals. In cancer patients, somatic events, such as translocations and point mutations can be identified in the circulation. The main bulk of circulating tumor DNA (ctDNA) are short (<300 bp) and have been proposed to be of apoptotic origin. Various methods have been used to track ctDNA. The detection of structural rearrangements is highly specific, which allows for the detection of single copy cancer genomes in milliliters of plasma. Sequencing, based on amplification, of a selected set of point mutations has demonstrated robust detection of ctDNA down to 0.02%. Therefore, whole-genome sequencing and exome sequencing, by using all point mutations in a tumor collectively to increase signal, has the potential to allow for effective monitoring of systemic tumor load. Since whole-genome sequencing remains ten times as expensive relative whole-exome sequencing, it is still not applicable in large patient cohorts where serial plasma samples are available over time. Here we report on a careful technical evaluation of various approaches to perform whole-exome sequencing performed directly on cell-free DNA. Also, the chosen methodology was applied on paired primary tumor material and plasma from ten prostate cancer patients to estimate the proportion of ctDNA.

3469T

Mutational enrichment of cancer-related gene sets in 11 aggressive prostate cancers. *K.J. Lindquist[1], R. Kazma[1], TJ. Hoffmann[1], BA. Rybicki[2], A. Levin[2], PL. Paris[1], JS. Witte[1].* 1) UCSF, San Francisco, CA; 2) Henry Ford Health System, Detroit, MI.

Prostate cancer is a leading cause of cancer mortality. Many prostate tumors are benign, but some are aggressive and lethal. The mutation profile of aggressive prostate tumors may differ from that of other tumors. To investigate whether somatic mutations in aggressive prostate tumors are more frequent in gene sets previously identified as functionally involved in benign prostate and other cancers, we sequenced the tumor and matched normal tissues of 11 aggressive prostate cancer patients using Complete Genomics' platform. After removing low-confidence calls, we selected mutations within any gene or gene regulatory region (using data from the Encyclopedia of DNA Elements) in the human genome. Then, we determined if 22 gene sets were associated with mutation rates using a mixed-effects Poisson regression model. The gene sets included 18 cancer-related pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG), 2 differential expression indicators from the Prostate Expression Database, and 2 indicators of cancer drug sensitivity or resistance from the Genomics of Drug Sensitivity in Cancer database. Membership in the KEGG prostate cancer pathway was independently associated with higher mutation rates in our samples ($p=0.024$), but stronger predictors of high mutation rates were the KEGG transcriptional misregulation pathway ($p<0.001$), differential expression in response to androgen ($p=0.001$), and two other KEGG cancer pathways (non-small cell lung and endometrial cancers, $p=0.001$ for both). Our work paves the way for future studies to link the mutational enrichment of cancer-related gene sets with the aggressiveness of prostate cancers.

3470F

Retrospective analysis of genomic and transcriptional changes in a case of Ewing's sarcoma tumor progression determined by whole transcriptome and exome semiconductor-based sequencing. *G. Meredith[1], Y. Sun[1], N.S. Hernandez[2], M. Taylor[1], B. Sanderson[2], K. Giorda[1], T. Woodburn[6], G. Bee[3], J. Bishop[3], S. Ghosh[1], P. Kapranov[4], J. Buckley[5], K. Bramlett[2], C.P. Reynolds[6], T. Triche[7].* 1) Life Technologies, South San Francisco, CA; 2) Life Technologies, Austin, TX; 3) Life Technologies, Carlsbad, CA; 4) St. Laurent Institute, Providence, RI; 5) Keck School of Medicine of USC, Los Angeles, CA; 6) Texas Tech University Health Sciences Center, Lubbock, TX; 7) The Saban Research Institute Childrens Hospital Los Angeles, Los Angeles, CA.

Ewing's sarcoma is a pediatric cancer that often presents in the second decade of life and is usually associated with a chromosomal translocation t(11;22)(q24;q12) that results in a EWS/FLI1 gene fusion. Long term survival rates for patients with metastases can be less than 10%. Four independent cell-lines have been established from a patient who succumbed to metastatic disease following relapse after myeloablative chemotherapy. Whole-transcriptome and exome sequencing of normal primary bone marrow-derived stromal fibroblasts (cell-line COG-FB-425), Epstein-Barr Virus (EBV) transformed normal lymphoblasts (cell-line COG-V-455), a pre-therapy primary tumor-derived cell-line (CHLA-9), and a post-chemotherapy metastatic tumor-derived cell-line (CHLA-10), was conducted on an Ion Torrent Proton™ system to profile the differences in gene expression and differences in exonic DNA sequence to characterize the molecular changes associated with primary tumorigenesis and disease persistence after treatment. All cell lines matched by short-tandem repeat analysis. The presence of the EWS/FLI1 fusion gene in the tumor cells was confirmed and the breakpoint determined from both observation of chimeric reads in the RNA-seq data and exome sequence analysis. Exome datasets, collected to >140X average depth of coverage, indicate apparent loss of heterozygosity genome-wide in CHLA10 consistent with cytogenetic analysis that shows tetraploidy in this cell-line. Results from RNA-seq also indicate numerous instances, genome-wide, of differing transcript isoform expression and exon usage between normal, primary tumor, and metastatic tumor cells suggesting an increasing genomic mutational burden in the evolution of the disease, and pointing in particular toward aberrant regulation of RNA-splicing components. One co-expressed, first exon-sharing pair of sense/antisense transcripts corresponding to the gene FEZF1 and transcript FEZF1-AS1 that is unique to the tumor lineage is the subject of further investigation. Taken together, the combination of RNA-seq and exome-sequencing on normal cells and primary vs. post-chemotherapy tumor is providing a broad and deep view of molecular signatures in tumor progression and indicating that a significant role is played by changes in non-coding RNA expression.

3471W

Immunohistochemical analysis of uterine leiomyomas, histopathological uterine leiomyoma subtypes, and uterine leiomyosarcomas. *N. Mäkinen[1], K. Kämpjärvi[1], R. Bützow[2], P. Vahteristo[1].* 1) Department of Medical Genetics, Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 2) Department of Pathology, The Laboratory of Helsinki University Central Hospital (HUSLAB), Helsinki University Central Hospital and Haartman Institute, University of Helsinki, Helsinki, Finland.

Uterine leiomyomas (ULMs) are benign tumors that arise from smooth muscle cells of the myometrium. Regardless of their benign nature, leiomyomas can cause difficult symptoms, such as excessive bleeding, abdominal pain and discomfort, pregnancy complications, and infertility. Based on histopathology, ULMs can be divided, in addition to common leiomyomas, into various relatively rare subtypes that mimic malignancy in one or more aspects. Cellular, atypical, and mitotically active leiomyomas are examples of these subtypes. Rarely, ULMs may undergo malignant transformation and develop into a leiomyosarcoma. We have previously shown by exome sequencing that MED12 exon 2 is mutated in approximately 70% of ULMs. All the mutations affect an evolutionary conserved region of the MED12 protein. Our subsequent studies have revealed that mitotically active leiomyomas display a MED12 mutation frequency that is not statistically different from common leiomyomas, whereas the rest of the histopathological ULM variants and uterine leiomyosarcomas harbor significantly less MED12 exon 2 mutations than common leiomyomas. The aim of this study is firstly to examine the frequency and respective proportions of known mutations of ULMs, such as mutations in FH and HMGA2, and secondly to scrutinize the possibility that one of the histopathological ULM subtypes would undergo malignant transformation and develop into a leiomyosarcoma by immunohistochemistry. The study material consists of 199 formalin-fixed paraffin-embedded samples including 66 common leiomyomas, 103 histopathological ULM variants (cellular, atypical, and mitotically active leiomyomas), and 30 uterine leiomyosarcomas. We are constructing tissue microarrays of these different sample groups for immunohistochemical stainings. Antibodies for proteins, such as Ki67 and HIF1α, are utilized to analyze the similarities and differences between the histopathological variants and uterine leiomyosarcomas. It is clinically relevant to recognize possible tumor subtypes with different molecular genetic background to provide better tools for more accurate diagnosis and drug development. It is also important to understand the mechanisms behind the progression to malignancy, because uterine leiomyosarcomas are usually diagnosed as late as in the surgery.

3472T

LOH Analysis of the ING 3 and 5 genes in Breast Cancer. *G. NAS[1], E. GUNDUZ[1], M. ACAR[1], E. UCTEPE[1], M. BOZER[3], C. DENER[3], S. YENIDUNYA[4], M. GUNDUZ[1,2].* 1) TURGUT OZAL UNIVERSITY MEDICAL FACULTY, ANKARA, Turkey; 2) DEPARTMENT OF OTOLARYNGOLOGY HEAD AND NECK SURGERY, TURGUT OZAL UNIVERSTY MEDICAL FACULTY, ANKARA, Turkey; 3) DEPARTMENT OF GENERAL SURGERY, TURGUT OZAL UNIVERSTY MEDICAL FACULTY, ANKARA, Turkey; 4) DEPARTMENT OF PATHOLOGY, TURGUT OZAL UNIVERSTY MEDICAL FACULTY, ANKARA, Turkey.

The tumor suppressor genes (TSG) ING3 and ING5, the members of inhibitor of growth gene family, are effective in inhibition of cell growth and induction of apoptosis to eliminate cancerous cells. However, in many cancer types, one of the alleles of a TSG is lost through carcinogenesis, while the rest allele is usually inactivated through a process called as loss of heterozygosity (LOH). Previous studies in head and neck cancer revealed that allelic loss and reduced expression is a common pattern of ING gene family members but our study showed that despite the high proportion of LOH in HNSCC, allelic loss of ING3 and ING5 does not commonly occur in breast cancer. 50 paraffin embedded breast cancer tissues were analyzed and the allelic deletion frequency of ING3 and ING5 gene were detected as 12% and 16% respectively. We used to Kaplan-Meier survival analysis to evaluate overall(OA) and disease-free survivals (DFS) in the groups of breast cancer patients with LOH negative and LOH positive ING genes. Both ING 3 and ING 5 LOH status have no significantly difference in OA and DFS of breast cancer patients. These results provide a rational explanation and relative contribution for complexity of tumor formation that allelic loss of ING3 and ING5 gene are not major factors for breast cancer but a part of big complex mechanism.

3473F

Characterization of molecular alterations in urologic cancers. *M.L. Nickerson[1], K.M. Im[1], S. Turan[1], T. Andresson[2], L.E. Moore[3], M. Dean[1].* 1) Cancer & Inflammation Program, National Cancer Institute, Frederick, MD; 2) Laboratory of Proteomics and Analytical Technologies, Advanced Technology Program, SAIC-Frederick, National Cancer Institute, National Institutes of Health, Frederick, MD; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD.

Metastatic cancer is the primary cause of death among patients with urologic and other cancers. Metastatic prostate cancer (mPC) alone is responsible for over 258,000 deaths worldwide each year and is poorly controlled using existing therapies. The molecular chronology from primary to metastatic disease is not known in sufficient detail and detailed molecular markers that differentiate indolent from aggressive subtypes of cancer are needed. We employed exome and targeted sequencing strategies to examine mutated genes in prostate, bladder, and kidney cancers. We present detailed analysis of an index patient with metastatic, castration-resistant prostate cancer where we sequenced the exomes of five metastatic tumors and non-cancer kidney tissue. We observed sequential acquisition of somatic alterations in the primary tumor, including a deletion associated with formation of a TMPRSS2-ERG fusion transcript and a missense alteration in the acetylated lysine binding pocket of a bromodomain in polybromo 1. The majority of somatic alterations, including an alteration of the CpG demethylase, TET2, were not observed in the primary tumor but were observed in all metastatic tumors, indicating mutations likely contributing to metastatic disease may not be detected by analysis of primary tumor tissue. Alterations observed in individual metastatic tumors allowed us to anatomically map the spread of disease. We show TET2 is altered in additional mPC tumors and identify signaling pathways likely affected by TET2 loss. This data is used to discuss observations in urologic cancers that indicate common features of these cancers. We show combinations of rare germline and somatic alterations, and combinations of sequence alterations and loss of heterozygosity differentially target individual cancer genes. Sequential accrual of somatic alterations during disease progression represents a challenge to identify therapies targeting individual cancer genes.

3474W

Exome sequencing identifies putative drivers of progression of transient myeloproliferative disorder to AMKL in infants with Down Syndrome. *S.I. Nikolaev[1], F. Santoni[1], A. Vannier[1], E. Falconnet[1], E. Giarin[2], G. Basso[2], A. Hoischen[3], J.A. Veltman[3], J. Groet[4], D. Nizetic[4], S.E. Antonarakis[1].* 1) GeDev, University of geneva, geneva, geneva, Switzerland; 2) Department of Pediatrics, University of Padua, Padua, Italy; 3) Department of Human Genetics, Nijmegen Center for Molecular Life Sciences, Institute for Genetic and Metabolic Disease, Radboud University Medical Center, Nijmegen, Netherlands; 4) Queen Mary University of London, Blizard Institute of Cell and Molecular Science, Barts and The London School of Medicine, London, United Kingdom.

Some neonates with Down Syndrome are diagnosed with self-regressing TMD and 20-30% of those progress to AMKL. To elucidate the molecular mechanisms of these neoplasias we have performed exome sequencing in 7 cases and copy number analysis in these and 10 additional cases. All TMD/AMKL samples contained GATA1 mutations. None of 7 exome-sequenced TMD/AMKL samples had any other recurrently mutated genes. However 2 of 5 TMD cases, and all sequenced AMKL cases, showed mutations/deletions other than GATA1, in genes proven as transformation-drivers in non-DS leukaemia (EZH2, APC, FLT3, JAK1, PARK2-PACRG, EXT1, DLEC1 and SMC3). One patient at the TMD stage revealed two clonal expansions with different GATA1 mutations, of which one clone had an additional driver-mutation. Interestingly, it was the other clone (with just a GATA1-mutation) that gave rise to AMKL after accumulating mutations in 7 other genes. Our data suggest that GATA1-mutations alone are sufficient for clonal expansions, and additional driver mutations at the neonatal TMD stage do not necessarily predict AMKL progression. Later in infancy, leukaemic progression requires 'third-hit-driver' mutations/SCNAs found in non-DS leukaemias. Putative driver-mutations affecting WNT, JAK-STAT or MAPK/PI3K pathways were found in all cases, aberrant activation of which converges on overexpression of MYC. Thus the study of DS-AMKL established the 'three-hit' hypothesis for this neoplasia: trisomy 21, GATA1 and variable third driver on the MYC pathway.

3475T

Characterization of somatic alterations in the novel tumor suppressor DEAR1 using ultra-deep targeted next generation sequencing. *J. Reuther[1,2], N. Chen[2], A. Sahin[3], S. Lott[4], A. Killary[1,2].* 1) Human and Molecular Genetics Program, University of Texas Graduate School for Biomedical Science, Houston, TX; 2) Genetics, MD Anderson Cancer Center, Houston, TX; 3) Pathology, MD Anderson Cancer Center, Houston, TX; 4) Clinical Implementation, Life Technologies, Carlsbad, CA.

Ductal Carcinoma In Situ (DCIS) accounts for 12-24% of all diagnosed breast cancers and is one of the earliest pre-invasive forms of breast cancer. Without treatment, DCIS can progress to invasive disease with recorded frequencies from 14-60%. There is an urgent need to identify prognostic markers for women with a heightened risk of progression from DCIS to invasive breast cancer (IBC) for which more aggressive surveillance and treatment might be warranted, as well as individuals with favorable prognosis, who might be spared rigorous therapeutic regimens and for whom breast conservation therapy might be the preferred surgical option. We previously identified the novel tumor suppressor Ductal Epithelium Associated Ring Chromosome 1 (DEAR1) as a gene, located at 1p35.1, the expression of which is downregulated or lost at the DCIS stage. Our previous work has also shown that loss of expression of DEAR1 by immunohistochemistry significantly predicted local recurrence in early onset breast cancer. Our sequencing efforts, in addition to analyses of the TCGA cohorts, have identified over 30 non-synonomous mutations in DEAR1 in a variety of epithelial cancers including breast, pancreatic and colon cancer, many of which undergo copy number losses within chromosome 1p35. In order to better understand DEAR1's role in cancer and specifically, its potential role in the progression of DCIS lesions to IBC, our lab is currently undergoing efforts to determine if DEAR1 undergoes mutation in DCIS utilizing the Ion Torrent PGM platform. We are performing targeted ultra-deep sequencing of a 52kb locus containing DEAR1 in both pure DCIS samples and DCIS with associated invasive components, as well as IBC lesions in order to determine the earliest stage in breast cancer in which mutations in DEAR1 can be observed and whether mutations correlate with DCIS that progress to IBC. Targeted deep sequencing has yielded a mean depth of coverage of 9,000x, allowing us to identify both common as well as rare variants. Preliminary results indicate that DEAR1 is mutated at the earliest stage of breast cancer and include the identification of novel variants in DCIS as well as missense mutations that have been found previously in IBC. To our knowledge, only few other genes are known to be mutated in DCIS, which suggests the importance of the characterization of this novel tumor suppressor and the determination of its role as a tumor suppressor in breast cancer.

3476F

Individualized analysis of somatic mutations and CNV from exomes obtained from breast cancer core needle biopsies in women with newly diagnosed locally advanced breast cancer. *H. Sicotte[1], S. Hart[1], J.P. Sinwell[1], S. Baheti[1], P.T. Vedell[1], K.R. Kalari[1], X. Tang[1], K.J. Thompson[1], D.W. Mahoney[1], P. Barman[1], J.M. Evans[1], C. Wang[1], Y.W. Asmann[2], J.P.A. Kocher[1], T.J. Dockter[1], K.N. Jones[3], A.L. Conners[3], A.M. Moyer[3], D.W. Visscher[3], V.J. Suman[1], J. Eckel Passow[1], R. Weinshilboum[1], L. Wang[1], J.C. Boughey[4], M.P. Goetz[5].* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic, Jacksonville, FL; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Department of Surgery, Mayo Clinic, Rochester, MN; 5) Department of Oncology, Mayo Clinic, Rochester, MN.

Introduction: Individualization of cancer therapies based on tumor sequence data is increasing and guidance is needed on what platforms should be used to detect actionable genomic variations or alterations. In the Breast Cancer Genome Guided Therapy Study (BEAUTY), women with locally advanced breast cancer undergo core needle breast tumor biopsies to study genome variations prior to initiating neoadjuvant chemotherapy. Those biopsies are profiled using exome sequencing (tumor and germline) and RNASeq.We will present our methods of analyzing the exome for mutations and somatic copy number variations as well as some cross-platform comparisons. Exome sequencing is rapidly becoming the platform of choice for tumor sequencing because it affords much higher coverage of actionable variants (in coding region and around splice sites of genes) and is cheaper than whole genome sequencing, but it suffers from the lack of accurate tools for calling somatic CNV.

Results: We have developed a somatic CNV calling tool based on the pattern CNV method (submitted), which uses reference samples to learn the pattern and variance of the exome sequencing coverage to better enable somatic CNV calling. This tool is able to detect with high sensitivity EGFR amplifications (FISH validated) in low clonality tumors from core needle biopsies. We will also present the comparison of the efficiency of CNV detection with changes detected using RNASeq. Somatic variant calling in exomes, and in particular low tumor purity samples, is a challenging problem. We are calling somatic variants using multiple somatic calling tools:, Joint SNVMix2, SomaticSnipper, Mutect, and GATK Somatic indel detector. These tools call sets of variants with very little overlap. We will present our protocol for integrating multiple callers as well as for our tiering system for judging the technical validity, functional significance, and clinical significance of variants.

3477W

Subclonal evolution and genomic drivers of relapse in childhood acute lymphoblastic leukemia. *JF. Spinella[1], R. Vidal[1], J. Healy[1], V. Saillour[1], C. Richer[1], P. Cassart[1], M. Ouimet[1], S. Busche[2], B. Ge[3], T. Pastinen[2,3], D. Sinnett[1,4].* 1) Sainte-Justine UHC Research Center, University of Montreal, Montreal, Qc, Canada; 2) Department of Human Genetics, McGill University, Montreal, QC, Canada; 3) McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada; 4) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Qc, Canada.

Precursor B-cell acute lymphoblastic leukemia (pre-B ALL) is the most frequent pediatric cancer. Increased understanding of the pathobiology of this disease has led to risk-targeted treatment regimens and increased survival rates. However, close to 20% of patients still do not respond to current treatment protocols. Refractory ALL is the leading cause of death by disease among children. Genetic intratumoral heterogeneity is one of the major mechanisms leading to treatment resistance; whereby clonal evolution and interclonal competition give rise to the accumulation and maintenance of subclonal mutations with a relapse driving potential. However, the clinical significance of these relapse driver mutations and their underlying impact on childhood ALL outcome remain elusive. We are using high throughput sequencing technology to better understand the genomic landscape of primary and relapsed childhood ALL. Using a unique experimental design that consists of ten matched tumor (at diagnosis), normal (remission) and relapse samples, we performed whole exome and whole genome sequencing, as well as genome-wide genotyping, to identify copy number aberrations (CNAs) and somatic single nucleotide variants (SNVs) potentially driving primary leukemic transformation and relapse. We found gain of function mutations in the RAS-MAPK signaling pathway, as well as SNVs in the HAT domain of CREBBP that were associated with both ALL onset and relapse. SNV clustering analysis will allow us to describe clonal evolution and analyze the dynamics of interclonal competition within individual ALL tumors. Ultimately, this work will provide invaluable insight of the genetic mechanisms underlying pediatric ALL which could lead to the development of powerful clinical tools to improve detection, diagnosis and treatment of this childhood cancer.

3478T

Ultra-high quality sequencing assay for comprehensive genetic panel analysis of rare tumor-derived circulating cell-free DNA. *A. Talasaz, D. Sebisanovic, G. Mei, L. Siew, H. Eltoukhy.* Guardant Health, 2686 Middlefield Rd, Suite D, Redwood city, CA.

Background: Current approaches based on invasive biopsy genetic analysis can fail to capture an accurate picture of the real-time tumor genetic profile due to high tumor heterogeneity and cannot be in practice for serial monitoring of disease progression or acquired resistance. Analysis of circulating tumor nucleic acids (ctDNA), on the other hand, presents a new tool for the monitoring and treatment of cancer. However, due to high-quality false positives in current NGS assays, the majority of studies on ctDNA have been limited to hotspot analyses, amplicon sequencing approaches and typically only involve patients where ctDNA fractions are high (>1-5%). **Methods:** We have developed a differentiated sequencing assay, Digital Sequencing Technology (DST) that enables ultra-sensitive and ultra-specific detection of rare genomic abnormalities. Standard NGS workflows are plagued by extremely high noise and distortion in sample-prep and sequencing. DST is able to eliminate the error and distortion created by these processes and produce near-perfect representations of all rare variants. Moreover, our DST workflow enables the vast majority of DNA molecules (even in <10ng input samples) to be converted to sequencing libraries, enabling ultra-sensitive detection. **Results:** We first compared the sensitivity and specificity of conventional Illumina SBS sequencing versus DST. We have shown that in sequencing a comprehensive cancer panel of 70kbp in 0.1% cancer cell line titration samples, standard Illumina SBS generates many high-quality false positive variant calls in the range of 0.05-5% (even with a rigorous methodology of overlapped paired-end read bases with Qscore>30 filtering) while DST resulted in highly sensitive and completely error-free variant calls across the entire panel. We then applied DST to rare tumor-derived cell-free DNA on more than 25 cancer patient plasma samples across different cancer types. We investigated the concordance of tumor mutation profiles derived from ctDNA with those derived from matched tumor biopsies. We found higher than 90% concordance between tumor and ctDNA somatic mutation profiles in colorectal, breast, esophagus and melanoma cancers across all patients studied and multiple stages (II-IV). **Conclusion:** This work indicates the potential of using DST in deep analysis of ctDNA, thereby allowing researchers and clinicians to comprehensively and non-invasively monitor the genetic dimension of cancer throughout the body.

3479F

Next-generation sequencing of BRCA1 and BRCA2 in Bulgarian breast cancer patients and controls. *D.I. Toncheva[1,3], L. Balabanski[1], G. Antov[2], I. Dimova[3], S. Ivanov[1], M. Nacheva[4], I. Gavrilov[4], D. Nesheva[3], B. Rukova[3], S. Hadzhidekova[3], M. Malinov[1].* 1) Medical Genetics, Medical University-Sofia, Sofia, Sofia, Bulgaria; 2) nstitute of Genetics, Bulgarian Academy of Sciences, "Acad. Doncho Kostov". Sofia-1113; 3) Department of Medical Genetics, Medical University of Sofia, Zdrave str. 2, 1431 Sofia, Bulgaria; 4) Specialized Hospital for Active Treatment in Oncology, Sofia, "Plovdivsko Pole" str. 6, Bulgaria.

Breast cancer is the most common type of cancer in women. Most of its hereditary forms are caused by mutations in the BRCA1 and BRCA2 genes whose main function is DNA repair of the double-stranded ruptures. Genetic testing of women with family history is therefore recommended to determine whether they have hereditary predisposition for this type of cancer. The variants with no clear clinical significance are problematic to make a diagnosis when performing target re-sequencing. After taking the informed consent, DNA samples from 24 patients (average age of 35 ±10) diagnosed with breast cancer and having family history and 71 age matched not affected women were collected. Sequence-targeted BRCA1 and BRCA2 libraries were then prepared using the TruSeq Custom Amplicon method, and these were then sequenced on Illumina MiSeq system. A wide range of variants was found in BRCA1 and BRCA2 genes. In patients' group were found two pathological/presumably pathological variants: mutation in BRCA2 at position chr13:32890665 that affect the first position of the 5' splice region following exon 2 and mutation in BRCA1 at position chr17:41219635 - in-frame triple nucleotide deletion of Valine 1688 (8,3%). In the both groups (patients/control) we have detected 7 likely polymorphic variants and 13 common variants in BRCA1 and BRCA2 genes. For the first time 3 common polymorphisms in BRCA2, characteristic only for Bulgarian population, were detected: chr13:32973737, T/-, SNP within 3'-UTR of exon 27; chr13:32973280, A/-, mononucleotide deletion within 5'-UTR of exon 27; and chr13:32973924, T/-, mononucleotide deletion, downstream of gene sequence. This is the first next generation targeted sequencing of BRCA1 and BRCA2 genes in Bulgarian population, and it paves the way for finding local founder mutations as well as variants typical for this region.

3480W

Recurrent Somatic Mutations in Loss of Heterozygosity Regions of Hepatocellular Carcinoma. *S. Tsai, Y. Lin.* Molecular and Genomic Medicine, Nat. Health Research Institutes, Zhunan, Miaoli, Taiwan.

BACKGROUND Hepatocellular carcinoma (HCC), the most common liver malignancy, is characterized by frequent loss of heterozygosity (LOH) at multiple chromosomal locations. Genetic alterations in these LOH regions have not been systematically examined. METHODS We investigated 58 target genes in the LOH regions by using Agilent SureSelect enrichment technology in 12 HCC paired samples. Eight candidate genes identified by capture sequencing were further examined by TruSeq custom amplicon sequencing technology to uncover mutation hotspots in 95 HCC tumors. DNA mass spectrometry was used to validate the findings in independent cohorts. RESULTS Initially we discovered 15 nonsynonymous mutations in 7 genes (CSMD1, DOCK5, KCNH1, HHAT, PCCA, RYR2 and ZNF423) by capture sequencing. Then 113 variants were identified by amplicon sequencing, including 5 nonsense mutations. Among them, only 11 were reported in the dbSNP137. Eight missense mutations in 4 genes (CSMD1, DOCK5, PCCA, and PCDH7) and one nonsense mutation in RYR2 gene occurred in at least two patients. CSMD1 is a known tumor suppressor gene and 9/95 HCCs contained a CSMD1 mutation at sites that are critical for protein-protein interactions. Furthermore, a single mutation in PCCA gene encoding propionyl CoA carboxylase occurred in 8/95 HCCs. CONCLUSIONS Focused investigation of selected genes in the LOH regions revealed frequent and recurrent somatic mutations. Genetic alterations discovered by this approach can be used to identify distinct HCC subtypes for developing patient-specific management.

3481T

Importance of genetic analysis in the prediction of retinoblastoma in South Indian patients. *A. Vanniarajan[1], G. Namrata[2], K. Usha[2], RS. Akram[1], K. Thirumalairaj[1], I. Jeyaram[3], R. Santhi[4], R. Kim[5], VR. Muthukkaruppan[6].* 1) Department of Molecular Genetics, Aravind Medical Research Foundation, Madurai, Tamil Nadu, India; 2) Department of Orbit, Oculoplasty and Oncology, Aravind Eye Hospital, Madurai, Tamil Nadu, India; 3) Statistics Division, Lions Aravind Institute of Community Ophthalmology, Madurai, Tamil Nadu, India; 4) Department of Pathology, Aravind Eye Hospital, Madurai, Tamil Nadu, India; 5) Retina Clinic, Aravind Eye Hospital, Madurai, Tamil Nadu, India; 6) Department of Immunology and Stem Cells Biology, Aravind Medical research Foundation, Madurai, Tamil Nadu, India.

Purpose: India has the highest incidence of retinoblastoma (RB) among the developing countries. Genetic analysis was performed in south Indian RB patients to understand the mutation pattern of RB1 gene in order to predict the risk of RB in siblings and offspring. Methods: The study population included 73 patients of RB who attended Orbit, Oculoplasty and Oncology Department of Aravind Eye Hospital during January to October 2012. A semi log plot was derived to understand the significance of age at diagnosis of RB. Quantitative Multiplex PCR and Exonic sequencing were carried out in tumor samples, wherever available and blood samples of patients and their families. Results: Out of 73 patients, 31 had bilateral and 42 had unilateral RB. Pedigree showed positive family history in 14 patients (19%) that include 12 bilateral and 2 unilateral RB. Bilateral RB occurred much earlier (9.82 ±11.52 months) than unilateral RB (24.02± 15.11 months), in accordance with Knudson's hypothesis. Further genetic analysis of tumor samples of 2 unilateral cases showed somatic, nonsense mutations that were present only in tumor but not in blood samples of probands and parents and hence the risk of RB was predicted to be less than 1% in siblings. Among 3 bilateral cases without family history, two patients had deletions of RB1 and one patient had a nonsense mutation in blood, which is more likely to be inherited to next generation, although parents did not have those mutations. Out of 2 cases with family history the deletions of exons affecting the Pocket domain was observed in both the proband and father. In another family, proband and mother were having a splice site mutation and hence predicted the next sibling had 50% risk of getting RB. When the child was born, clinical examination and genetic testing confirmed RB with same mutation as that of proband. Because of early detection, the child was treated successfully and vision was preserved. Conclusion: Genetic analysis confirmed the Knudson's two-hit hypothesis in Indian population for the first time. Detection of somatic mutations in sporadic unilateral cases predicted a lesser chance of RB in future generation. The identification of heritable mutation in bilateral cases predicted an increased risk of RB in siblings and next generation, thereby providing early diagnosis and treatment.

3482F

Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *M. Zeschnigk[1], M. Martin[2], L. Maβhöfer[1], P. Temming[3], S. Rahmann[2], C. Metz[4], N. Bornfeld[4], J. van de Nes[5], L. Klein-Hitpass[6], A.G. Hinnebusch[7], B. Horsthemke[1], D.R. Lohmann[1].* 1) Institute of Human Genetics, Faculty of Medicine, University of Duisburg-Essen, Essen, Germany; 2) Genome Informatics, Faculty of Medicine, Institute of Human Genetics, University of Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; 3) Department of Paediatric Haematology and Oncology, University Hospital Essen, Hufelandstrasse 55, 45122 Essen, Germany; 4) Department of Ophthalmology, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; 5) Institute of Pathology and Neuropathology, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; 6) Biochip Laboratory, Institute for Cell Biology, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; 7) Laboratory of Gene Regulation and Development, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Building 6, Room 230, Bethesda, MD 20892, USA.

Uveal melanoma (UM), a rare eye cancer, can be divided into two major classes based on molecular features. Tumors of either class are distinct by metastatic potential and patients' prognosis. Specifically, tumors with monosomy 3 (UM-M3), present in about half of patients, lead to metastases and poor survival. Tumors with disomy 3 (UM-D3) rarely metastasize. To capture the mutational differences underlying the diverse classes of UM we performed exome sequencing on 10 UM-D3 and 12 UM-M3. We identified six genes -GNAQ, GNA11, BAP1, SF3B1, EPB41L3,and EIF1AX- as targets of recurrent somatic protein-altering mutations. GNAQ and GNA11 are known to be mutated in both UM classes. Mutations in BAP1 and SF3B1 were reported in some UM-M3 and UM-D3 tumors, respectively. We sequenced EPB41L3 in 89 UM and found an overall low mutational frequency at this locus (<2%) thus suggesting a minor role of this gene in UM biology. Mutations in EIF1AX, an X-chromosomal gene which codes for a translation initiation factor, have not been reported in cancer so far. We performed re-sequencing of EIF1AX and SF3B1 in 31 UM-D3 and found somatic missense and in-frame mutations at a frequency of 48% (15) and 29% (9), respectively. Of note, in UM-D3 mutations of SF3B1, which encodes a subunit of the U2 snRNP splice complex, mostly affected codon 625. This mutational pattern is distinct from that in other tumors, specifically CLL and MDS, which show more diverse patterns of SF3B1 mutations. In 35 UM-M3 re-sequencing revealed that mutations of EIF1AX and SF3B1 are infrequent (2 tumors, 5.7%). Notably, EIF1AX and SF3B1 mutations are mutually exclusive in UM suggesting a complementary role of both genes in UM tumorigenesis. Re-sequencing of 13 UM with loss of parts of chromosome 3 (partM3), which are rare, showed SF3B1 mutations in 7 (53%) and an EIF1AX mutation in one other (8%) sample. The presence and type of EIF1AX and SF3B1 mutations in UM-partM3 hints at a common tumor biology of at least the mutant UM-partM3 and UM-D3. We also sequenced SF3B1 and EIF1AX in 10 rare UM-D3 tumors that developed metastases (UM-D3met) and found SF3B1 mutations in 3 samples. However, all 3 mutations were outside of codon 625 thus suggesting that the rare cases of UM-D3met are molecularly distinct. Interestingly, SF3B1 and EIF1AX mutations were biased towards UM from male patients. This lead us to hypothesize that a Y linked gene(s) is part of the pathway targeted by altered EIF1AX and SF3B1 function.

3483W

Exome sequencing characterize the somatic mutation spectrum of early serrated lesions in a patient with *BRAF* negative hyperplastic polyposis syndrome. *P. Hoffmann[1,2,3], S. Horpaopan[1], J. Altmueller[4], R. Hueneburg[5], P. Kahl[6], D. Drichel[7], H. Thiele[4], I. Spier[1], M. Odenthal[6], M.M. Noethen[1,2], C. Strassburg[5], H. Froehlich[8], P. Nuernberg[4], R. Buettner[6], S. Aretz[1].* 1) Institute of Human Genetics, University of Bonn, Germany; 2) Department of Genomics, Life & Brain Center, University of Bonn, Germany; 3) Division of Medical Genetics, University Hospital and Department of Biomedicine, University of Basel, Switzerland; 4) Cologne Center for Genomics (CCG), University of Cologne, Germany; 5) Department of Internal Medicine I, University of Bonn, Germany; 6) Institute of Pathology, University of Cologne, Germany; 7) German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 8) Department of Bonn-Aachen International Center for IT (B-IT), Algorithmic Bioinformatics, University of Bonn, Germany.

Background. Hyperplastic polyposis syndrome (HPS) is a yet poorly defined colorectal cancer (CRC) predisposition characterized by the occurrence of multiple and/or large serrated lesions throughout the colon. A serrated polyp-CRC sequence (serrated pathway) of CRC formation has been postulated, however, to date only few molecular signatures of serrated neoplasia (*BRAF*, *KRAS* mutations, CpG Island Methylation, microsatellite instability) were described in a subset of HPS patients and neither the etiology of the syndrome nor the distinct genetic alterations during tumorigenesis have been identified.

Methods. To describe the mutational landscape of serrated polyps and the involved pathways we sequenced the exomes (Illumina HiSeq platform) of 11 early hyperplastic polyps without the *BRAF* V600E mutation obtained from a 41 year-old female patient with clinically confirmed HPS. For data analysis the VARBANK pipeline of the Cologne Center for Genomics was used. Somatic mutations were identified by comparison with leukocyte DNA and validated by Sanger sequencing.

Results. By analyzing the exome data we initially identified 25 unique somatic alterations in 8/11 serrated tumors. All variants are single basepair substitutions. The predominant mutation type seemed to be missense mutations caused by G>T transversions. However, all but one of the G>T transversions could not be validated by Sanger sequencing, pointing to technical artifacts. The remaining validated seven somatic mutations in seven genes (*ABI3BP, CATSPERB, CCBP2, COL8A1, VGLL2, CALD1, DNAI1*) are each present in one polyp only. Four mutations occurred in the same single polyp, the other three were identified each in another polyp. No known cancer genes are among the seven candidates, however, some of them are described to be involved in cell adhesion, proliferation, or cell invasion.

Conclusions. Somatic mutations seem to be rare events in early hyperplastic lesions of HPS patients without a *BRAF* mutation. No frequently affected genes and no enrichment of specific pathways have been observed. Thus, other alterations such as epigenetic changes or variants in regulatory regions might be the major driving force of tumor progression.

The study was supported by the German Cancer Aid (project no. 108421).

3484T

First identification of frequent somatic *VHL* gene alterations in head and neck paragangliomas. *A. Merlo, S. Bernaldo de Quirós, I. Saenz de Santamaría, C. Suarez, M.D. Chiara.* Hospital Central Asturias, Oviedo, Spain.

Head and neck paragangliomas (HNPGLs) are rare tumors arising from parasympathetic paraganglia. 30-45% of HNPGLs are hereditary caused by germline mutations in succinate dehydrogenase genes (*SDHs*). The molecular connection between SDH dysfunction and tumor development is still unclear. The most accepted hypothesis proposes a central role for the pseudohypoxic pathway activated by the hypoxia inducible factor (HIF). Paradoxically, we recently showed that activation of HIF in HNPGLs is restricted to a subset of non-*SDHx*-mutant-HNPGLs. This consisted in accumulation of HIF-1α protein, and over-expression of both, HIF target genes and the HIF-inducible microRNA, miR-210. The present study aimed at unraveling the *SDH*-independent mechanisms involved in the activation of HIF in HNPGLs. To this end, we analyzed *VHL* gene in 53 tumors from patients with HNPGLs by direct gene sequencing, Multiplex-Ligation-dependent Probe amplification, qPCR and array CGH. *VHL* mRNA and protein levels were analyzed by RT-qPCR and immunohistochemistry, respectively. Meta-analysis of the gene expression signature of pseudohypoxic-HNPGLs was performed using Oncomine platform. Meta-analysis of the gene expression signature of pseudohypoxic-HNPGLs revealed that these tumors are highly related to clear cell carcinomas suggesting that HIF/miR-210 pathway is activated via pVHL deregulation in HNPGLs. Accordingly, we identified, for the first time, somatic *VHL* inactivating mutations (c.482G>A and c.227_229delTCT) in two of four pseudohypoxic-HNPGLs with concomitant LOH in one of them. *VHL* gene mutations were not found in HNPLGs that lack activation of HIF. However, partial or complete deletion of *VHL* was found in 32% of non-pseudohypoxic-HNPGLs regardless of the presence or absence of germline *SDHx* mutations. In addition, low VHL protein levels were detected in 64% tumors in association with decreased *VHL* mRNA levels. Taken together, our results suggest that the *VHL* gene is the most frequent target of somatic gene alterations in HNPGLs; and may be an important player in the development of a subset of sporadic HNPGLs via activation of HIF-1α-dependent pathways. Moreover, *VHL* may also have a previously unrecognized role in the pathogenesis of the vast majority of hereditary and sporadic HNPGLs by activating HIF-independent pathways yet to be identified. This provides the foundation for the development and use of therapeutic approaches that target the pVHL and/or HIF pathway in HNPGLs.

3485F

Sensitive, highly multiplexed somatic mutation analysis of FFPE tissues by deep amplicon sequencing on MiSeq®. *C. Lin, K. Chang, E. Upsall, I. Lewis, A. Tian, A. Iyer, G. Costa, R. Shen.* Illumina, Inc, San Diego, CA.

DNA extraction from formalin-fixed, paraffin-embedded (FFPE) tumor tissues typically results in low quantities of poor-quality, damaged DNA, which presents challenges for sequencing. Here we provide a fast, streamlined FFPE sample-to-answer workflow using probe design software and sample preparation methods optimized to overcome these challenges. DesignStudio® allows for the targeting of up to 1,536 user-defined genomic regions and supports amplicon sizes ideal for characterizing genomic hotspot regions or sequencing of whole exons with FFPE DNA (amplicons down to 125-bp, with tiling optimums at 175-bp median or higher). The TruSeq® Custom Amplicon (TSCA) workflow delivers high target enrichment specificity (>80% of reads on target), robust target multiplexing (up to 1,536 amplicons per reaction), and scalable sample indexing (up to 96 libraries). The improved protocol allows lower DNA input (≥100 ng) as compared to previous methods, and supports reduced error rates by means of optimized reagent formulations. Importantly, target representation demonstrates 3x - 47x more read depth in some regions with 80-90% GC. The TSCA assay is highly sensitive, detecting variants down to 2.5% allele frequency using FFPE reference samples verified with digital PCR. As a result of increased MiSeq output, 96 FFPE DNA samples can be targeted with the TruSeq Amplicon Cancer Panel and sequenced at 1,000x mean read depth in a single MiSeq run. For germline variant detection, 80 high-quality DNA samples can be targeted with a 1,536-amplicon custom panel and sequenced in a single MiSeq run. In support of a sample-to-answer workflow, the TSCA software analysis tools significantly reduce false-positive variant calls, and enable annotation and filtering of variants. In conclusion, we have developed an integrated FFPE DNA sequencing solution for highly accurate and reproducible detection of somatic variants implicated in cancer biology.

3486W

Alternatively spliced *DICER1* transcripts arising from genomic point mutations. *M. Wu, L. de Kock, L. Witkowski, M.R. Fabian, W.D. Foulkes.* Medical Genetics, Lady Davis Institute, McGill University, Montreal, Quebec, Canada.

DICER1 is a key endoribonuclease involved in generating small non-coding RNAs, including microRNAs (miRNAs) and short interfering RNAs (siRNAs). Several germ-line *DICER1* mutations have been identified and are associated with a pleiotropic tumour predisposition syndrome. The RNase IIIb domain of DICER1 is responsible for generating 5p miRNAs and is encoded largely by exon 25. We recently identified two tumour-specific somatic single base substitutions (c.5429A>G and c.5438A>G) that result in exon 25 skipping from the *DICER1* transcript. We postulate that these point mutations create novel exonic splicing silencer motifs that lead to exon 25 exclusion. After surveying the literature and documenting novel exon 25 mutations from our own studies, we examined whether another 8 other point mutations in exon 25 could also result in its skipping. In total, we found that 4/10 of the exon 25 mutations resulted in either exon skipping or alternative splicing. Here we demonstrate *in vitro* that the mutation c.5429A>T results in an exon 25 exclusion event while the mutation c.5428G>T creates a cryptic splice site. Importantly, we show that *DICER1* transcripts lacking exon 25 or bearing part of exon 25 can be translated. Since exon 25 encodes much of the DICER1 RNase IIIb domain, we plan to determine the impact of these DICER1 mutants on miRNA generation and to compare the functional significance of different mutations in RNase IIIb. This study highlights the importance of examining the effects of genomic mutations at the level of transcript maturation.

3487T

A personalized mutation network approach to detect putative cancer driver genes from next generation sequencing data. *Z. Zhao[1,2,3,4], P. Jia[1].* 1) Department of Biomedical Informatics, Vanderbilt Univ, Nashville, TN; 2) Department of Psychiatry, Vanderbilt Univ, Nashville, TN; 3) Department of Cancer Biology, Vanderbilt Univ, Nashville, TN; 4) Center for Quantitative Sciences, Vanderbilt Univ, Nashville, TN.

Next generation sequencing (NGS) of an individual genome typically reveals a large volume of point mutations. A major challenge in interpreting such mutation data is to distinguish clinically relevant or driver mutations from neutral passenger mutations, thus, facilitating the identification of actionable mutations and drug's target genes. Current approaches are primarily frequency-based to identify highly recurrent genes in sequenced samples; these approaches lack the power to detect moderately or rarely mutated driver genes and ignore functional interconnection and regulation among cancer genes. To tackle these issues, we proposed a novel, personalized mutation network method, VarWalker. VarWalker first estimates the probability of a gene to be mutated in each sample as a function of its cDNA length through a resampling-based test. It then applies the Random Walk with Restart algorithm for each individual sample to search for neighborhood interactors of the mutation genes in the reference network, which can be defined by the user. Finally, it builds a consensus mutation network based on the recurrent significant interactions at the cohort level. We demonstrated our approach in two NGS datasets: a lung adenocarcinoma (LUAD) sample cohort including 183 matched tumor/normal patients and a melanoma patient cohort including 121 matched pairs. For both datasets, the resultant mutation networks were significantly enriched with known cancer genes from Cancer Gene Census (CGC): the LUAD network recruited 70 CGC genes (p-value $< 2.2\times10^{-16}$, Fisher's Exact Test) and the melanoma network recruited 65 CGC genes (p-value $< 2.2\times10^{-16}$). Specifically for the LUAD network, for which we extracted a total of 57 known LUAD genes from previous studies, the mutation network recruited 31 of them (p-value $< 2.2\times10^{-16}$, Fisher's Exact Test). Importantly, within the major functionally connected, cancer-gene driven subnetworks that VarWalker identified, we detected moderate- and low-frequency cancer genes, demonstrating VarWalker's effectiveness on detecting these categories of driver genes, which are ignored by current frequency-based approaches. We expect VarWalker can be applied broadly to NGS-based disease studies for identification of disease candidate genes and their functional interactions.

3488F

DNA copy number variation and expression of miRNA150 suggestive as prognostic factor for Colorectal cancer. *R.V. Andrade[1], N. Gasparini[1], L. Sakamoto[3], G. Pereira[2], T. Lins[3].* 1) Universidade Católica de Brasília, Brasília, Brazil; 2) Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil; 3) Universidade de Brasília, Brasília, DF, Brazil.

Colorectal cancer (CRC) is the third most common cancer worldwide. It is caused by adenocarcinomas that arise from polyps. The survival rate for CRC is lower as the disease progresses, influencing the prognosis of the patients. Several molecular mechanisms were already described, such as copy number variation (CNV) and miRNA differential expression. MiRNAs can regulate the gene expression by the inhibition of translation or cleaving target mRNAs, and CNVs could regulate gene expression by deletion, translocation, duplication and insertion of these regions. MicroRNA genes embedded in copy number variation can have the expression affected by gene dosage effect. Both molecular mechanisms can be part of cell maintenance, and also modulate the prognosis of cancer. This work aimed to correlate the copy number variation of microRNA genes with clinical features of patients with CRC. Copy number variation and microRNA were evaluated by qPCR using TaqMan assays in normal, tumor and lymph node tissues. The Mantel-Cox test along with the Kaplan-Meier test was used to determine the survival time between the group of diploid (2 copies of microRNA genes) and non-diploid carriers. The expression of miR150 was statistically different between normal and tumoral (p=0.001); and between normal and lymph node tissues (p=0.001) showing that the expression of tumor is downregulated in relation to normal tissue, and the expression of lymph node is upregulated in relation to normal tissue. However, there was no significance between survival rates and up or downregulation (p>0.05). For the CNV, the Wilcoxon test did not indicate statistical difference in the gain of CNV between normal and tumor, and with lymph node tissues (p>0.748). However, when groups were defined by diploid and non-diploid individuals it indicated that the CNV-miR150 diploid genotype is associated with lower survival rates (p=0.022). Data showed that the diploid copy of CNV-miR150 is associated with lower survival rates. This indicates that CNVs containing genes of variable nature may behave differently to the gene dosage altering the functional expression in a tumor sample. For future experiments it is suggested to evaluate whether other genes are inserted within the CNVs in addition to miRNAs and quantify their expression in relation to normal tissue.

3489W

MicroRNA-192 regulates cellular proliferation in medulloblastoma seeding by targeting of dihydrofolate reductase. *S.A. Choi[1,2], S.Y. Yang[3], H.A. Kim[1,2], K.C. Wang[1,2], J.H. Phi[1,2], J.Y. Lee[1,2], J.H. Choi[1,2], D.Y. Um[1,2], Y.J. Moon[1,2], P.A. Kwak[1,2], E.J. Kwon[1,2], S.K. Kim[1,2].* 1) Division of Pediatric Neurosurgery, Pediatric Clinical Neuroscience Center, Seoul National University children's Hospital; 2) Adolescent Cancer Center, Seoul National University Cancer Hospital; 3) Division of Neurosurgery, Dongguk University Hospital.

Medulloblastoma (MB), a most common malignant brain cancer in children, frequently disseminates throughout the cerebrospinal fluid. The treatment of leptomeningeal dissemination (seeding) of the MB is challenging and the prognosis is extremely poor. In this study, we profiled miRNA expression and investigated the effect of miR-192 regulated DHFR (Dihydrofolate reductase) expression to get a better understanding of biology in MB seeding. miRNA profiling was carried out using agilent human miRNA microarray kit and qRT-PCR and western blot were performed using tissues and MB cell lines. For the further studies, MB cell lines were transfected miRNAs and investigated the luciferase activity, viability, proliferation assay and cell cycle analysis. The expression level of miR-192 is down-regulated in seeding group and three different MB cell lines compared to non-seeding MB and/or normal cerebellum. MiR-192 directly binds to the 3'-untranslated region (3'UTR) of DHFR mRNA, resulting in over 50% decrease in DHFR protein level in MB cell lines. The overexpressed miR-192 reduced the cell proliferation and induced cell cycle arrest by controlling DHFR target protein. The miR-192 overexpression considerably inhibited vimentin protein level suggesting that it might control the anchorage-dependent cell growth not migration. Our results demonstrate that DHFR is a direct target of miR-192 regulation and overexpression of miR-192 reduced tumor cell proliferation by inhibiting the expression level of the DHFR. Our findings describe a new mechanism for the regulation of DHFR/miR-192 and vimentin link providing a better understanding of signal pathway of MB seeding.

3490T

Genome Wide miRNA Expression Profile in Oral Cancer: An Exploratory Study. *N. De Sarkar, B. Roy.* Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India.

MicroRNAs (miRNAs) are a family of small non-coding RNA molecules of about 20-23 nucleotides in length, which is known to regulate expression of protein-coding genes at post-transcriptional level. Aberrant miRNA expression contributes to tumorigenesis and cancer progression and new member/s of miRNAs aberrantly expressed in various cancer types is also being reported frequently. Using TLDA array (a real-time-TaqMan Assay, Life technologies) we have checked expression of 762 human miRNAs in 18 pairs of Gingivo-Buccal cancer-adjacent control tissues and identified 7 miRNAs whose expression levels were significantly altered. Known oncogenic miRNAs, e.g. hsa-miR-31and hsa-miR-7, are among the miRNAs whose expression is up-regulated in this study. Besides these, we have also noticed significant up regulation of expression of hsa-miR-31*, which also act as Onco-miR via RHOA mediated pathway. Tumor suppressive miRNAs, such as hsa-miR-206, hsa-miR-204 and hsa-miR-133a, are also significantly down-regulated in this study. For the first time, we are reporting that expression of hsa-miR-204 is being down regulated in oral cancer. Besides we have also noticed significant up-regulation of expression of hsa-miR-1293, for the first time, in 9 of the 18 cancer samples. To explore the possible roles of these miRNAs in the disease process of Gingivo-Buccal carcinoma, we looked into the expression data with the help of different web databases and bioinformatics tools. Several biological pathways that are well characterized in cancer are significantly targeted by these seven miRNAs and they include MAPK and PI3-AKT pathways. These results indicate that these aberrantly expressed miRNAs may regulate, coordinately, several oncogenic pathways in Oral Cancer.

3491F

Germline BAP1 mutations in sporadic and familial mesothelioma. *I. Dianzani[1,2], M. Betti[1], E. Casalone[1], D. Ferrante[3], A. Romanelli[4], F. Grosso[5], S. Guarrera[6,7], L. Righi[8], G. Pelosi[9,10], R. Libener[11], D. Mirabelli[12,2], V. Ascoli[13], M. Papotti[8], G. Matullo[6,7], C. Magnani[2,3,12].* 1) Dept Health Sciences, Univ Piemonte Orientale, Novara, Italy; 2) Interdepartmental Center 'G. Scansetti', University of Turin, Italy; 3) CPO-Piemonte and Unit of Medical Statistics and Epidemiology, Dept. of Translational Medicine, University of Piemonte Orientale, Novara, Italy; 4) Emilia-Romagna Mesothelioma Registry, Department of Public Health, Reggio Emilia, Italy; 5) Division of Medical Oncology, SS. Antonio e Biagio General Hospital, Alessandria, Italy; 6) Human Genetics Foundation, HuGeF, Turin, Italy; 7) Dept. Medical Sciences, University of Turin, Italy; 8) Dept. Of Oncology, University of Turin at San Luigi Hospital, Orbassano, Turin, Italy; 9) Dept. of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy; 10) Dept. of Biomedical and Clinical Sciences "Luigi Sacco", University of Milan, Italy; 11) Pathology Unit, SS. Antonio e Biagio General Hospital, Alessandria, Italy; 12) Unit of Cancer Epidemiology, CPO-Piemonte and University of Turin, Italy; 13) Dept. of Radiological, Oncological and Pathological Sciences, Sapienza University, Rome, Italy.

Malignant mesothelioma (MM) is the only tumor due to exposure to a single carcinogen, asbestos. Only 10-17% of subjects exposed to high levels of asbestos develop MM, suggesting the role of individual genetic susceptibility. In two GWA studies we have reported several SNPs that cause a 2-3 fold increase in MM risk, but a much higher risk is due to asbestos exposure (Matullo et al 2013, Cadby et al 2013). Predisposition to MM may also occur with an autosomal dominant fashion. Germline loss-of-function (LOF) mutations in the BAP1 oncosuppressor gene are responsible for an inherited syndrome with predisposition to MM, uveal and keratinocytic melanoma, renal carcinomas and other malignancies (Testa et al 2011, Popova et al 2013). So far germline mutations in BAP1 were identified in 28 families, 9/28 showed multiple cases of MM and 3/28 showed a single case of MM. Germline mutations have been identified also in sporadic MM cases (2/26, Testa et al 2011). In this study, we report the analysis of BAP1 in four multiplex families and in 103 sporadic MM cases with the two aims of further investigating BAP1 related cancer syndrome and of estimating the role of germline BAP1 mutations in sporadic MM. One family carried a new LOF germline mutation (c.46_47insA). By using immunohistochemistry we show that BAP1 is not expressed in tumor tissue, according with Knudson's two hits hypothesis. Interestingly, whereas the three patients (mutation carriers) that were possibly exposed to asbestos developed MM, a further individual (mutation carrier) who was not exposed developed a different tumor type, i.e.mucoepidermoid carcinoma. This suggests that carcinogen type exposure may be important for the cancer type that is developed by mutation carriers. The other families did not show mutations in BAP1. None of the 103 sporadic patients showed mutations in BAP1, with an estimated confidence interval computed using Poisson distribution from 0 to 3.58% . Our data show that germline BAP1 mutations have a very limited role in sporadic MM. On the other hand, our study focused on familial aggregation of MM identifies a new BAP1 mutation, extends the cancer types associated with these mutations and suggests the existence of other yet unknown genes in the pathogenesis of familial MM.

3492W

The role of genetic ancestry in DNA repair capacity among Puerto Rican women with breast cancer. *H.J. Diaz-Zabala[1], L. Morales[2], J. Matta[2], J. Dutil[1].* 1) Department of Biochemistry, Ponce School of Medicine, Ponce, PR; 2) Department of Pharmacology, Ponce School of Medicine, Ponce, PR.

Introduction: Breast cancer incidence and tumor characteristics have been shown to vary among different ethnic populations. Previous studies in Puerto Rico have shown that breast cancer patients have a reduced DNA repair capacity (DRC) when compared to non-cancerous controls. In the Hispanic population of Puerto Rico, the contribution of genetic ancestry to breast cancer and associated risk factors is currently unknown. Objective: This study aims to determine the role of genetic ancestry in DNA repair in a group of breast cancer cases and controls. Methods: The study population consisted of 317 breast cancer patients and 364 non-cancerous controls that were recruited from private oncology practices in Puerto Rico. A panel of 106 ancestry-informative markers was genotyped using iPLEX sequencing technology. Individual genetic ancestry was estimated using a maximum likelihood approach. DRC levels were measured using a modified host-cell reactivation assay. An epidemiological questionnaire with clinical information from each patient was assessed for statistical associations with genetic ancestry. Results: The distribution of genetic ancestry among the 681 individuals varied from 0 to 100% African, 0 to 100% European, and 0 to 39% Native American. There was no significant difference in the average European (p= 0.163), African (p=0.382) and Native American (p=0.243) ancestry proportions between breast cancer cases and controls. We observed that low DRC levels were significantly associated (p=0.047) with higher proportions of Native American ancestry in non-cancerous controls. In cases, low DRC levels were significantly associated with higher proportions of European ancestry (p=0.011) and lower proportions of African ancestry (p=0.027). Conclusion: Our data indicate that global genetic ancestry is not associated with breast cancer risk in Puerto Rican women. However, we provide evidence that African, European and Native American ancestries are associated with changes in DRC.

3493T

A Non-Smad TGF-β Signaling Pathway Govern Adult Stem Cell State and Breast Cancer Metastasis with the Cooperation of SLUG and SOX9. *H. Fazilaty[1], B. Behnam[2].* 1) Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran; 2) Medical Genetics and Molecular Biology, Iran University of Medical Sciences, Tehran, Iran.

Transforming growth factor (TGF)-β superfamily has shown to be crucial in development, homeostasis and malignancies. TGF-β promotes metastasis by the induction of epithelial to mesenchymal transition (EMT), a process involved in gastrulation, wound healing and induction of cancer stem cells. Interestingly, EMT has been connected to stem cell traits of epithelial tissues by the induction of stemness. Although master regulatory networks are expected in adult stem cells, our knowledge of such network is still poor. Cancer stem cells are the only capable population of tumor cells to progress to overt metastasis, and are known share adult stem-like cells features. On the other hand, occurrence of appropriate microenvironmental conditions would be critical for stem/cancer stem cells. Master regulators not only determine stem/cancer stem cell state, but also may have regulatory roles in niche elements. Meanwhile, both stem/cancer stem cell and niche factors may function like two sides of the same coin. TGF-β superfamily, also, may play crucial roles in regulation of niche elements as well. In this regard, we hypothesize presence of a possible emerging molecular pathway in the biological process of breast stem/cancer stem cells induction and maintenance. In this process, non-Smad TGF-β induced signaling connects stem/cancer stem cell and niche elements by the mediation SLUG and SOX9 master transcription factors, and TNC and POSTN ECM components.

3494F

Copy number variation and the expression of miRNAs in colorectal cancer. *N. Gasparini[1], T. Lins[2], S. Andreoli[1], L. Lemos[1], J. Lima[1], G. Pereira[3], L. Sakamoto[2], R. Pereira[1,2], R. Pogue[1], R. Andrade[1].* 1) Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília, DF, Brazil; 2) Universidade de Brasília, Brasília, DF, Brazil; 3) Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Colorectal cancer (CRC) is a pathology that affects the colon and rectum. It arises from polyps that become adenocarcinomas, and these polyps may occur sporadically or be inherited. It is the third most common cancer worldwide, and accounts for 8% of cancers in Brazil. Several molecular mechanisms such as mutation, copy number variation (CNV) and miRNA expression have been evaluated in CRC. MiRNAs (non-coding RNAs that regulate gene expression by inhibiting translation or cleaving target mRNAs) and CNV (DNA sequences susceptible to deletion, duplication, translocation, inversion and insertion) are of great importance in the regulation of cell growth, differentiation, proliferation and apoptosis. MiRNAs and CNVs are associated with various types of cancer, and show differential expression in tumor tissues, and miRNA sequences frequently occur in CNV regions; thus it is reasonable to speculate that their differential expression in cancer may be due presence of differing numbers of copies of the miRNA genes. In this context, this work aimed to investigate the association between copy number variation of miRNA genes and the respective differential expression of miRNAs in normal, tumor and lymph node tissues of patients diagnosed with CRC. We selected 10 miRNAs that are known to be differentially expressed in CRC, and located within CNV regions. miRNA expression and copy number variation in 30 patients were analyzed by qPCR, using TaqMan assays. Relative quantitation of miRNAs was examined by ddCq, and quantification of copy number was performed using the CopyCaller software, normalized to RNAseP endogenous reference gene. The t-test significance for tumor and lymph node showed that miR570, miR338, let7g, miR1, miR150, miR183, miR650 and miR31 were differently expressed. The Wilcoxon signed-rank test showed that CNV-miR570 and CNV-miR650 had different copy number when comparing normal and cancerous sample tissues. Finally, a non-parametric correlation assessed the existence of a link between variation in copy number and expression levels of miRNAs. In only one CNV-miRNA (CNV-miR16 in tumor samples), was the correlation was significant, which indicates that in general, CNV probably does not influence gene dosage of miRNAs related to CRC. This study was essential to contribute for the understanding of the molecular mechanisms that which elucidates better diagnosis and prognosis of CRC.

3495W

Expression profiling of cofilin-1 in breast cancer cell lines and biopsies. *A. Hadjisavvas[1], C. Sutton[2], S. Shaheed[2], P. Loadman[2], V. Speirs[3], A. Hanby[3], K. Kyriacou[1].* 1) The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) Institute of Cancer Therapeutics University of Bradford, Bradford, UK; 3) Leeds Institute of Molecular Medicine, Leeds, UK.

Breast cancer is a significant cause of death since many tumors do not respond to treatment or acquire resistance. Hence there remains the need to identify novel targets for therapy and biomarkers for prediction of response to treatment. A quantitative proteomics study of matched normal and tumor biopsies identified 63 proteins to be significantly increased or decreased in stage-specific tumors. Some have previously been associated with breast cancer, while others such as cofilin-1 represent new targets for investigation. Cofilin-1 was subject to a range of analyses to determine its association with breast cancer. Western blotting (WB) was performed on protein extracts from breast cancer cell lines and matched normal and tumor biopsies from Cypriot patients with different stages of the disease. Immunohistochemistry (IHC) was carried out on core biopsies. Multiple reaction monitoring (MRM) mass spectrometry was performed on trypsin-digested protein extracts from biopsies.WB indicated that cofilin-1 was ubiquitously expressed in tumor cell lines, representative of Luminal A and B, basal-like, claudin-low and HER2 phenotypes, at higher levels than normal breast cell lines. WB also indicated increased expression of cofilin-1 in invasive carcinoma tissues, compared to matched normal. Patients with ductal carcinoma in situ or fibroadenoma exhibited less clear results, either increasing slightly or remaining unchanged. MRM analysis of three cofilin-1 peptides in tissue extracts of invasive carcinoma patients indicated expression only in tumor. IHC of core biopsies exhibited strong staining for cofilin-1 in ductal invasive carcinoma tissues with no staining in normal breast cells. Genomics and proteomics databases indicate that cofilin-1 is expressed in a diverse range of normal tissues. However, mRNA and protein levels are observed to be increased in skin,liver, pancreatic and breast tumors. The original quantitative proteomics data indicated only relatively small changes in expression (<2-fold) due to dynamic range limitations of the technique. Additional analytical approaches substantiated that cofilin-1 is significantly up-regulated in advanced breast cancer patients. Further patient sets are required to confirm the importance of cofilin-1 levels in early stages of breast cancer compared to benign tissues. Bioinformatics provided further confidence in our findings, highlighting the value of utilising databases for evidence of disease-specific proteins.

3496T

Angiogenesis in Chronic Hepatic Patients. *I.A. Helwa[1], H.A. El Dakhakhny[1], N.M. Kholoussi[1], R.F. Mahmoud[1], A. Hamada[2].* 1) Immunogenetics Department, National Research Centre, Cairo, Egypt; 2) Ain Shams University Hospitals, Cairo, Egypt.

Introduction: Hepatocellular carcinoma (HCC) is one of the most aggressive cancers worldwide. In Egypt, the disease is usually detected in an advanced stage at which no treatment may be effective including surgery. Early detection of the disease is thus an important goal allowing the patient to be treated before the enlargement of the tumor or its metastasis to distant organs. Objectives: To determine the role of Vascular Endothelial Growth Factor (VEGF) and Angiopoitin-2 (Ang-2) levels in prognosis of Hepatocellular Cancer (HCC) and Cirrhosis. Subjects and Methods: Eighty subjects were recruited in this study from Ain Shams University Hospitals; 29 patients suffering from liver cirrhosis, 36 HCC patients in addition to 15 apparently healthy control subjects. Serum levels of VEGF were assayed using a standard sandwich enzyme linked immunosorbent assay (ELISA) (RayBio Human VEGF ELISA Kit) with a minimum detectable dose less than 10 pg/ml. Moreover, serum Angiopoitin-2 (Ang-2) levels using ELISA technique (Invitrogen) with a minimum detectable dose <6 pg/mL. Results obtained were analyzed using Statistical Package for the Social Sciences (SPSS). Results: The serum VEGF levels in the HCC patients were significantly elevated as compared to the liver cirrhosis group and the apparently healthy control group. Nevertheless, serum Ang-2 levels in HCC patients were significantly elevated as compared to the liver cirrhosis group and the apparently healthy control group. Conclusions: Angiogenesis plays a crucial role in carcinogenesis, growth and progression of HCC, which is one of the most vascularized human cancers. Serum VEGF and Ang-2 levels appear to be critical angiogenic factors regulating angiogenesis in HCC and hepatic cirrhosis. Nevertheless, further directions aim at linking the 61*G allele polymorphism to EGF expression to assess as well the association between the EGF polymorphism and HCC risk.

3497F

Cancer Phenotypes of Germline Monoallelic *ATM* Mutation Carriers and Their Families. *H. LaDuca, A. Stuenkel, S. Keiles, T. Pesaran, S. Tandy, E. Chen, V. Speare, C. Radford, C. Gau.* Ambry Genetics, Aliso Viejo, CA.

Monoallelic *ATM* mutations have been associated with an increased risk of both breast and pancreatic cancers. *ATM* sequence and gross deletion/duplication analyses are included in three hereditary cancer panels offered by our laboratory, and to date, 26 *ATM* pathogenic mutation carriers have been identified. The purpose of this study is to describe the clinical characteristics of these *ATM* mutation carriers and their families. Retrospective test requisition form review was used to obtain clinician-reported clinical history information, including patient clinical history, testing history, and family history, for all pathogenic *ATM* mutation carriers. Follow-up contact with referring providers yielded additional clinical data for select patients. Chi-square analysis was used for statistical comparison of the observed mutation frequency in our Caucasian patients with that expected in the European American general population. All 26 *ATM* mutation carriers had a reported personal history of cancer, with 9 (34.6%) having a personal history of more than one primary cancer diagnosis. All but one patient (n=25) also had a family history of cancer. The spectrum of primary cancers observed in our *ATM* probands included breast, pancreatic, colorectal, ovarian, fallopian tube, leukemia, lymphoma, lung, basal cell carcinoma, cervical, and squamous cell carcinoma of the perineum. Twenty (76.9%) *ATM* carriers had a personal history of breast cancer, with an average age of 43 at first breast cancer diagnosis (range 31-69 years). Four (20.0%) of these breast cancer patients had a history of more than one primary breast cancer. Seven (26.9%) carriers had personal and/or family history of pancreatic cancer. *BRCA1/2* testing information was reported for 18 (69.2%) mutation carriers, and all reported results were negative. The observed frequency (1.8%) of pathogenic *ATM* mutations in our Caucasian patients was significantly elevated compared to the frequency in the European American general population (0.4%) (OR= 4.84, $p=1.3\times10^{-7}$, 95% CI= 2.54-9.25). Despite the selection bias of our testing population and the limitations of clinician-reported clinical history details, our data contributes to the growing body of literature on cancer risks in monoallelic *ATM* mutation carriers. Efforts are being made to determine co-segregation of mutations with disease in these families and will increase our knowledge of the penetrance of mutations in families and any additional cancer associations.

3498W

Contralateral Mastectomy and Survival after Breast Cancer in BRCA1 and BRCA2 Mutation Carriers. *K. Metcalfe[1], S. Gershman[1], P. Ghadirian[2], H. Lynch[3], C. Snyder[3], N. Tung[4], C. Kim-Sing[5], A. Eisen[6], W. Foulkes[7], B. Rosen[8], P. Sun[9], S. Narod[9]*. 1) Faculty Nursing, Univ Toronto, Toronto, ON, Canada; 2) Epidemiology Research Unit, Centre Hospitalier de Université de Montreal (CHUM), Montreal, Canada; 3) Department of Preventive Medicine and Public Health, Creighton University School of Medicine, Omaha, NE,; 4) Beth Israel Deaconess Medical Center, Boston, MA; 5) BC Cancer Agency, Vancouver, Canada; 6) Toronto Sunnybrook Regional Cancer Center, Toronto, Canada; 7) Program in Cancer Genetics, McGill University, Montreal, Canada; 8) University Health Network, Toronto, Canada; 9) Women's College Research Institute, Toronto, Canada.

Background: Women who carry a mutation in either the BRCA1 or BRCA2 gene face a lifetime risk of breast cancer and once diagnosed, face a high risk of second primary breast cancer. Contralateral mastectomy is an option to reduce the risk of contralateral breast cancer. However, it is unclear if contralateral mastectomy impacts on survival. The objective of the current longitudinal study was to compare the survival rates of women with BRCA-associated breast cancer who did and who did not undergo a contralateral (bilateral) mastectomy. **Methods:** Patients were 390 women with stage I or II breast cancer and a BRCA1 or BRCA2 mutation who were initially treated with unilateral or bilateral mastectomy. These patients were selected by pedigree review of families who received counselling at one of 12 participating clinical genetics centers. The medical records were reviewed for clinical presentation, medical and surgical treatments and outcome. Patients were followed for up to 20 years from diagnosis. Survival experience was compared for women who did and who did not have a contralateral mastectomy. **Results:** 79 women died of breast cancer in the follow-up period. At twenty years, the survival rate was 88% for women who had a contralateral mastectomy and 66% for women who did not. In a multivariable analysis, controlling for age and year of diagnosis, treatment and other prognostic features, contralateral mastectomy was associated with a 49% reduction in death from breast cancer (HR = 0.51; 95% CI: 0.51 to 0.28; p = 0.03). **Conclusions:** BRCA-positive women who are treated for breast cancer with bilateral mastectomy are less likely to die from breast cancer than women who are treated with unilateral mastectomy.

3499T

Genetic diagnosis of a series of Brazilian patients with pheochromocytomas and paragangliomas. *O. Moraes[1], D. Oliveira[1], B. Araújo[1], C. Meireles[1], L. Vianna[1,2], J. Mazzeu[1], M. Batista[3], A. Lofrano-Porto[1,2]*. 1) Universidade de Brasília, Brasilia, Distrito Federal, Brazil; 2) Hospital Universitário de Brasília, Distrito Federal, Brazil; 3) Hospital Regional de Taguatinga, Brasilia, Brazil.

INTRODUCTION: Pheochromocytomas (PCC) and Paragangliomas (PGL) are neuroendocrine tumors originating from chromafin cells, located in the adrenal medulla and extra-adrenal tissue, respectively. PCC and PGL are usually benign, with morbidity and mortality related to the production of catecholamines. Malignancy is reported in approximately 10% of cases. Recent studies have shown that at least 25% of all PCC/PGL cases may have a genetic basis. METHODS: 15 patients with PCC/PGL were included. Clinical, biochemical and radiological data were obtained from medical records. Histopathological and immunohistochemical studies for neuroendocrine markers were performed. Genomic DNA was extracted and the coding regions of VHL, SDHB and SDHD were amplified and automatically sequenced. Multiplex ligand-probe amplification (MLPA) was used for screening large deletions/insertions. RET gene analysis was also performed in one patient with clinical evidence of multiple endocrine neoplasia type 2A (MEN2A). Genuine consent was obtained from all. RESULTS: Our series consisted of 4 PGLs and 11 PCCs patients; 5 had familial history and 2 were malignant. Immunohistochemistry confirmed the neuroendocrine origin of all tumors. Genetic analysis revealed the (p.Q164R) VHL mutation in a woman with PCC and cerebellar hemangioblastoma and her asymptomatic daughter. A RET mutation was found (p.C618R) in a woman with PCC and thyroid medullary carcinoma (TMCa; MEN2A) and her 2 daughters with TMCa. A large deletion of SDHB exon 1 was found in 3 patients: Two sisters with paraortic PGL, and an apparently sporadic case presenting with relapsing retroperitoneal PGL. DISCUSSION: Molecular diagnosis done by Sanger's sequencing combined with MLPA constitute an adequate strategy to search for both point mutations and large deletions. Genotype-phenotype associations for VHL and RET are well described, but the pathophysiological effects of large deletions on SDHB are still unclear. No phenotype correlations have been characterized for large SDHB deletions due to a highly variable presentation. SDHB may act as a tumor suppressor, and a large deletion may lead to a tumor phenotype. Rational recommendations for genetic studies in PCCs and PGLs are in progress given that mutations may also be found in apparently sporadic cases, and that a positive result might influence clinical monitoring and genetic counseling of patients and their offspring.

3500F

Prevalence of Succinate Dehydrogenase-deficient GIST in adults with GIST. *I.R Rainville[1], E.J. Root[1], A.J. Wagner[2], S. George[2], C.L. Corless[4], J.E. Garber[1], J.A. Hornick[3]*. 1) Center for Cancer Genetics and Prevention, Dana-Farber Cancer Inst, Boston, MA; 2) Center for Sarcoma and Bone Oncology, Dana-Farber Cancer Inst, Boston, MA; 3) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 4) Portland VA Medical Center and Oregon Health and Sciences University Knight Cancer Institute, Portland, OR.

Introduction: GIST is the most common mesenchymal tumor of the GI tract, often associated with somatic gain-of-function mutations in *KIT* or, less commonly, *PDGFRA*. Deficiency of the mitochondrial complex II succinate dehydrogenase distinguishes a class of GISTs wild-type for *KIT/PDGFRA* with distinctive histopathology and clinical behavior. Germline mutations in *SDHX* with somatic loss of heterozygosity in tumors characterize the syndrome of GIST and paraganglioma and a subset of pathologically similar, apparently sporadic GIST, and correlate with loss of expression of SDH subunits by immunohistochemistry (IHC). We are systematically reviewing GIST pathology, family history, and germline status in our multicenter study to estimate the prevalence of SDH-deficiency among adults with GIST. We report the results from our Dana-Farber Cancer Institute (DFCI) cohort thus far. Methods: Individuals ≥ age 18 years with history of confirmed GIST were enrolled through the sarcoma clinic at the DFCI. Gastric GISTs were screened for features of SDH deficiency: epithelioid or mixed morphology, multinodular or plexiform growth pattern, and lymph node metastasis. Somatic *KIT/PDGFRA* genotype was noted when available. Suspected *SDH*-deficiency was confirmed by IHC for SDHA/B. Germline testing, if not done previously, was offered to study participants. Results: From March, 2008 through February, 2010, we enrolled 214 study participants with GIST in the DFCI sarcoma program. To date, IHC has identified 19 SDH-deficient GISTs (8.9%; CI 5.1% to 12.7%). Age of diagnosis ranged from 15 to 55 years (mean 31.9 years). *KIT/PDGFRA* mutation screening was available for 12 tumors, all of which were wild-type. Seven of 17 cases in which the family history was reviewed were apparently sporadic. Three cases had family history of GIST in a first or second-degree relative. Germline *SDHX* analysis of 9 cases performed in our study or elsewhere identified mutations in 7: 5 in *SDHB* and 2 in *SDHA*. Conclusion: SDH-deficient GIST comprises a significant fraction of all GISTs in this cohort. Family history of GIST or paraganglioma, when present, can inform genetic testing. In the absence of family history of other associated tumor, tumor histopathology with SDHA/B IHC are highly sensitive screening tools to identify patients with GIST who are at high risk to carry germline mutations.

3501W

Value of microRNA-based molecular profiling in Cancer of Unknown Primary (CUP) cases for familial cancer risk assessment and genetic counseling. *E.R. Wassman[1], E. Goren[1], J. Hale[2], C. Hogan[1], B. St.Cyr[1], M.O. Sanden[1]*. 1) Rosetta Genomics, Philadelphia, PA; 2) Hunterdon Regional Cancer Center, Flemington, NJ.

Background: Cancer risk assessment and counseling is dependent upon accurate diagnosis of historical malignancies in family members. It is confounded by lack of availability, incompleteness, and inaccuracy due in part to variability in histopathology. Cancer of unknown or uncertain primary (CUP) represents a unique challenge for counseling as they represent failure to reach a diagnosis in late stage cancers. Recently, molecular profiling with microRNAs was shown to effectively identify the underlying tumor type in CUP cases. This series illustrates how this methodology can identify otherwise unrecognized genetic risk. Methods: MicroRNA was isolated from 258 consecutive specimens, analyzed on a custom microarray, and 64 microRNAs interpreted for any of 42 specific tumor types as previously described. We routinely collect follow-up information on correlation of test prediction with other clinical-pathological findings, treatment response, and outcome. Results: 192 cases had sufficient tumor material and were successfully reported. The most common cancers identified were colorectal (12%) and breast (10.4%). Of the breast cancer cases, 5 (3% overall) were in males. Ovarian cancer was found in 6.8% and a total of 32 other tumor types were diagnosed. Referring doctors affirmed confidence and utility in a high percentage of cases. Discussion: MicroRNA resolved diagnoses provide valuable information for family members when no previous specific diagnosis was available. 28.4% of CUP cases analyzed by this method revealed diagnoses (colorectal, breast, ovarian) with particular significance for genetic counseling. The 5 male breast cancer cases described here suggest over a 20-fold enrichment in CUP compared to the observed population rate, where they represent only about 1% of all breast cancers. These are of particular relevance for counseling their sisters and daughters. Depending on the family history, a specific diagnosis of other cancers may also be contributory. MicroRNA has been shown to be particularly stable in FFPE preserved tumor blocks, providing a potential tool for investigation of the etiology of a deceased relative's cancer when CUP or doubtful results confound family history risk assessment. Conclusion: MicroRNA profiling in CUP patients has an additional clinical utility in uncovering important genetic risks associated with specific cancer diagnoses in at least one-third of cases, potentially directing specific genetic testing for relatives.

3502T

Identification of tissue-specific long intergenic non-coding RNAs in normal prostate tissue.
Y. Zhang, A.J. French, A.A. Nair, S. McDonnell, S. Middha, S. Riska, Z. Fogarty, D. Schaid, S.N. Thibodeau. Mayo Clinic, Rochester, MN.

Long intergenic non-coding RNAs (lincRNAs) have been found to perform various functions in a series of important biological processes. Because the transcription of lincRNAs is tissue and temporal specific, it is important to examine a variety of tissue types as part of the discovery process. Recently, genome-wide transcriptome deep sequencing technology and computational approaches have provided an unprecedented opportunity to analyze such transcripts. In this pilot study, we performed RNA-seq on 40 normal prostate tissue samples to investigate and characterize the entire transcriptome. Regions of normal prostate tissue were histologically confirmed from specimens obtained during radical prostatectomy for prostate cancer (n=38) or cystoprostatectomy in patients with bladder cancer (n=2). Specifically, we focused on the identification of prostate-specific lincRNAs. In total, we identified 22,176 candidate lincRNA transcripts at 16,145 loci. Candidate lincRNA were defined as having evidence from assembly for at least two samples. Overall, there was significant overlap between the lincRNAs identified in this study with lincRNAs annotated in GENCODE V16 (75% overlap at loci level). Importantly, 5544 novel lincRNAs were not previously reported. The prostate derived lincRNAs were further examined for evidence of transcriptional activity using the H3K4me3-H3K36mer3 domains generated from nine cell lines in the ENCODE project. Overall, 3659 lincRNAs (~35%) have evidence of a signature consistent with an actively transcribed gene across the entire locus (both H3K4me3 across the promoter region andH3K36me3 along the transcribed region). Of the remaining transcripts, 2797 (27%) overlap a H3K4me3 peak alone (promoter region) and 1221 (12%) overlap a H3K36me3 peak alone (transcribed region). Some future directions to characterize these lincRNA transcripts include differential expression analysis of these lincRNAs between different clinical groups, co-expression analysis of lincRNAs and protein-coding RNAs, and analysis of disease-associated SNPs in these lincRNA regions. Additionally, an expression quantitative trait loci analysis will be investigated between prostate cancer risk-SNPs and their neighboring lincRNAs surrounding the risk-SNPs.

3503F

The novel founder MSH6 mutation c.10C>T is an important cause of Lynch syndrome in the French Canadian population.
E. Castellsague[1], J. Liu[2], A. Volenik[3], R. Gagne[4], B. Maranda[5], S. Giroux[6], J. Latreille[7], L. Palma[3], L. Kasprzak[3], V. Marcus[8], M. Breguet[9], S. Nolet[10], AM. Mes-Masson[11], D. Provencher[11], G. Chong[2], F. Rousseau[12], W. Foulkes[1,3,13,14,15]. 1) Human Genetics, McGill University, Montreal, Canada; 2) Pathology Department, JGH, Canada; 3) Medical Genetics Department, McGill, Canada; 4) Génétique Médicale Departament, CHUL, Canada; 5) Pédiatrie Département, Université-Sherbrooke, Canada; 6) CHU Quebec, Canada; 7) Hematology and Medical Oncology Division, HCL, Canada; 8) Pathology Department, McGill, Canada; 9) Service de Medecine Genique, CHUM, Canada; 10) Pathologie Departement, CHUM, Canada; 11) Centre de recherche CHUM, Hôpital Notre-Dame, Canada; 12) Biologie moléculaire, biochimie médicale et pathologie Département, Université Laval-CHU Quebec, Canada; 13) Cancer Prevention Centre, JGH, Canada; 14) Cancer Genetics Program, McGill, Canada; 15) Medicine Department, McGill, Canada.

Lynch syndrome (LS) is the most common autosomal dominant condition predisposing to colorectal cancer (CRC). It is characterized by early onset cancers of the colorectum and endometrium (EC) among others. LS is caused by germline mutations in mismatch-repair genes MLH1, MSH2, MSH6 and PMS2. Founder mutations among them are an important cause of LS and facilitate genetic testing in specific ethnic populations. We have recently identified a new truncating mutation in MSH6, c.10C>T (p.Gln4X), in 10 Quebec French Canadian families presenting clinical features consistent with LS. Sixteen carriers of the mutation were also identified among 6,433 newborns screened in Quebec City. These data strongly suggest that MSH6 c.10C>T is one of the most common mutations causing Lynch syndrome discovered until now, and point to a common founder origin. We confirmed that MSH6 c.10C>T occurred as a unique event in a single founder individual by the existence of a shared haplotype in carrier individuals (≈3.3 Mb). To determine the age of the mutation, 150 French Canadian controls are currently under analysis. Regarding the clinical features, the mutation co-segregates with LS-spectrum cancers in 13 of the 19 carriers (average age at diagnosis: 44.8y). The remaining 6 non-affected carriers have a mean age of 34.3y. None of the non-carriers have developed a tumor belonging to LS spectrum (average age 50.8y). In accordance with a previously described increased risk of EC in MSH6-mutated LS families, 8 out of 9 affected women developed EC (in addition to CRC in 4 of them). The implication of MSH6 in the tumorogenesis of affected carriers is being assessed at a tumoral level: twelve tumors from different patients have been tested by MSH6 immunohistochemistry, and except for an inconclusive result, all show loss of MSH6 protein expression; similarly, 8 out of 10 analyzed tumors show microsatellite instability; and finally, until now, no loss of heterozygosity from the mutation has been found at tumoral level. To assess the relative risk that this highly frequent mutation (1/402) is conferring to develop LS-tumors, we will genotype around 200 CRC and 500 EC cases from French Canadian patients. Our results will increase the efficiency of molecular genetic testing for Lynch syndrome in the French Canadian population of Quebec. Also, careful clinical characterization of this founder mutation will lead to mutation-specific counseling and improved clinical care.

3504W

Whole-Exome Sequencing of Familial Aggregations of Radiation-Associated Meningiomas (RAM). *A. Pathak[1], R. Bruchim[2], D.R. Stewart[1], S. Sadetzki[2].* 1) NIH NCI Division of Cancer Epidemiology and Genetics, Rockville, MD; 2) Cancer & Radiation Epidemiology Unit, Gertner Institute, Israel.

Sporadic and radiation-associated meningioma are rare diseases in the general population, and familial meningioma occurs in (5-10%) of all cases. Our group has previously demonstrated that 11% of families with RAM have two to four first degree relatives who develop meningioma after exposure to radiation in childhood. However, the key genetic drivers of this ~25-fold increased cancer susceptibility have yet to be fully elucidated. We subjected 18 individuals from 6 RAM families to whole-exome sequencing. All individuals in the study were irradiated for tinea capitis. Five families had 2 affected individuals and 1 unaffected individual. One family had 1 affected individual and 2 unaffected individuals. The study population was derived from the Israeli Tinea Capitis (TC) cohort which included a group of 10,842 individuals treated during the 1950s with radiation therapy for TC. The sequencing revealed 42,860 non-synonymous variants. In our initial bioinformatics analysis, we implemented a custom PERL script to perform automated PubMed queries for the 706 genes of biological interest and the keyword 'meningioma.' This resulted in the identification of 16 variantsin genes associated with meningioma. Further investigation of this literature revealed that a missense mutation in C1D and a stop codon in UBAP1 (a protein involved in the ubiquitinization pathway) may be associated with RAM. Next, we filtered these variants based on disease status and ESP count (<0.5%). In addition to the stop codon in UBAP1, we also identified a stop codon in PTPN7, a protein tyrosine phosphatase. For missense mutations, we further filtered our variants based on having a polyphen-2 score of possibly damaging or probably damaging. We prioritized these variants based on the literature and based on occurrence in multiple families. We found 5 variants in all that occurred in several families. Specifically, we detected two damaging mutations in the RNA polymerase POL2RF and a single damaging variant in two families in a ubiquitinization gene UBQLN3. Further sequencing of these candidate genes in larger validation groups may elucidate the molecular mechanisms underlying the significant genetic predisposition to RAM. The fact that no single locus was found in these families may indicate that the mode of inheritance of familial RAM is based on the combined inheritance of several low-risk variants rather than a single high-penetrance gene.

3505T

Global differential expressions of isoforms of miRNA in retinoblastoma: correlation with level of expressed 3'-5'-exonucleases. *A. Ganguly[1], J. Leipzig[2], J. Richards-Yutz[1], J. Purrazzella[1].* 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Bioinformatics Core Center for Biomedical Informatics The Children's Hospital of Philadelphia.

MiRNAs are short, 21-24nt, oligonucleotides that function as post-transcription regulators of gene expression. Dysregulation of miRNA expression plays an important role in cancer. Retinoblastoma (RB) is the most common intraocular malignancy of childhood. To understand the role of miRNAs in RB, we profiled the miRNA transcriptome of normal neonatal retina, two RB cell lines, WERI and Y79, and three matched retina/RB samples using next-generation sequencing on Illumina platform.

Members of the let-7 family of miRNAs were differentially expressed between normal retina and two RB derived cell lines when sequenced on GAIIx platform. Additionally, pervasive presence of length variations at the 3'-ends of miRNAs, specifically in let-7b, let-7i and miR-181b were observed. The number of reads corresponding to each miRNA was greater than 100 in any one sample. The proportion of canonical to variant miRNAs was significantly different between normal retina and the two RB cell lines. This data was validated by comparison with the sequence data obtained from three sets of matched retina and RB on the Illumina HiSeq platform. The data on let-7i, was compared to published data in 293T cell line and found to be similar.

It has recently been shown that nibbler, a 3'-5'exoribonuclease, mediates 3'-end processing of miRNA that results in miRNA diversity and stability in Drosophila. We searched for the expression of human homologs of 3'-5'exonucleases in the retina and retinoblastoma. The expression of EXD2, a 3'-5'exonuclease was statistically different between normal retina, retinoblastoma and two cell lines - WERI and Y-79. The differential expression of EXD2 correlated with the relative abundance of the canonical/variant miRNA and with the degree of differentiation/aggressive behavior of the retinoblastoma tumors.

3506F

MALAT1 is deregulated and co-expressed with CREBBP in the long non-coding transcriptome of childhood Acute Lymphoblastics Leukemia. *R. Vidal[1], C. Richer[1], JF. Spinella[1], V. Saillour[1], M. Ouimet[1], S. Langlois[1], P. Cassart[1], J. Healy[1], E. Bareke[1], A. Droit[3], D. Sinnett[1,2].* 1) Centre de recherche, CHU Sainte-Justine, Montreal, QC, Canada; 2) Département de Pédiatrie, Faculté de Médecine, Université de Montréal, Montreal, QC, Canada; 3) Centre de recherche du CHUQ, Université Laval, Laval, QC, Canada.

Long non-coding RNAs (lncRNAs) are capable of influencing diverse cellular processes through various modes of action including regulation of gene expression. Consequently, lncRNAs are deregulated in a number of human diseases including cancers. Using RNA-seq technology, we describe the long non-coding transcriptome of childhood pre-B acute lymphoblastic leukemia (ALL) and identify, for the first time, lncRNAs that regulate downstream gene expression and that could be involved in driving leukemic transformation. We sequenced total RNA from 56 patients from the Quebec childhood ALL cohort and found a set of 352 deregulated lncRNAs significantly co-expressed with 1713 coding genes involved in key signaling pathways, including Wnt, PPAR and NF-kappa B, as well as genes involved in p53-regulated cell cycle control. Moreover, childhood ALL-specific lncRNAs also appear to play an important role in RNA processing mainly through the spliceosome and RNA degradation. Among the lncRNA identified in childhood ALL, MALAT1, known to control cell proliferation by regulating expression of a number of oncogenes, was found to be co-expressed (spearman correlation above 0.9) with the transcriptional coactivator and histone acetyltransferase CREB-binding protein (CREBBP), a gene known to be associated with bad prognosis in ALL. Our data suggest that transcriptional regulation by lncRNAs could be an important mechanism in promoting and/or maintaining leukemia development and in the future may assist in biomarker development and even be used in RNA-based therapies to ultimately improve diagnosis and prognosis of ALL in children.

3507W

Prediction of GWAS-identified risk loci in breast cancer and correlated SNPs through mapped epigenetic phenomena. *P.S. Rajagopal[1], Q. Li[2], P. Kraft[1], M. Freedman[2].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Medical Oncology, The Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA.

INTRODUCTION: The majority of loci identified from GWAS are located outside of known protein-coding regions. Projects such as ENCODE and GENCODE have begun epigenetic and non-coding RNA annotation of these regions. These databases have been used to associate GWAS findings en masse with putative function. However, this has been minimally explored for individual diseases, individual SNPs or predictive ability. Using breast cancer as an example, we correlated GWAS-identified risk loci with functional information from these databases to create a model that predicts a SNP's likely association with disease. **METHODS**: 76 index SNPs and 946 proxy SNPs (r^2=0.9) were selected based on the COGS breast cancer GWAS ($p \leq 10^{-7}$). These were matched to randomly selected non-GWAS SNPs by chromosome, minor allele frequency (MAF), r^2 and number of proxy SNPs. Information was collected on GRCh37.p12 position, MAF and presence(1)/absence(0) of a set of features: DNase I hypersensitivity, H3K4me3 histone modification and CTCF binding in MCF-7 breast cancer cells and human mammary epithelial cells (HMEC); RNA polymerase II binding in MCF-7; RegulomeDB score$\leq$4; interactions with promoters, enhancers and motifs from HaploReg; and presence in a transcript's coding region. We evaluated these elements for enrichment among GWAS-associated SNPs compared to the randomly selected SNPs. For each index SNP, binary counts were summed across all proxies such that values per SNP ranged between 0 and the total number of proxies for that SNP (sum). Variables were also created for any proxy having a trait (binary) and average across all proxies (average). A conditional logistic regression model was developed using forward selection ($p \leq 0.05$). **RESULTS**: Our final model included: RegulomeDB score $\leq$4 (binary) (OR=4.79, 95% CI:1.52-15.17), enhancers (sum) (OR=1.13, 95% CI:0.99-1.29), MCF-7 RNA polymerase II interaction (binary) (OR=9.02, 0.97-84.04) and SNP presence in transcripts (average) (OR=2.51, 95% CI=0.95-6.68). HMEC DNase I hypersensitivity (average) correlated closely with RegulomeDB score. We lacked enough information to evaluate individual RNA types, H3K4me3 or CTCF. **CONCLUSIONS**: We have demonstrated the feasibility of deriving a SNP-based predictor of association with disease based on genomic and epigenetic features enriched in known risk loci vs. random loci. This method synthesizes readily accessible information to bring GWAS findings to the post-GWAS era.